Sparse Nonnegative Solution of Underdetermined Linear Equations by Linear Programming

David L. Donoho and Jared Tanner

March 2005; Corrected April 8, 2005

Abstract

Consider an underdetermined system of linear equations $y = Ax$ with known $d \times n$ matrix A and known y. We seek the sparsest nonnegative solution, i.e. the nonnegative x with fewest nonzeros satisfying $y = Ax$. In general this problem is NP-hard. However, for many matrices A there is a threshold phenomenon: if the sparsest solution is sufficiently sparse, it can be found by linear programming.

In classical convex polytope theory, a polytope P is called *k-neighborly* if every set of k vertices of P span a face of P. Let a_j denote the j-th column of A, $1 \leq j \leq n$, let $a_0 = 0$ and let P denote the convex hull of the a_j. We say P is *outwardly k-neighborly* if every subset of k vertices *not including* 0 spans a face of P. We show that outward k-neighborliness is completely equivalent to the statement that, whenever $y = Ax$ has a nonnegative solution with at most k nonzeros, it is the nonnegative solution to $y = Ax$ having minimal sum.

Using this and classical results on polytope neighborliness we obtain two types of corollaries. First, because many $\lfloor d/2 \rfloor$-neighborly polytopes are known, there are many systems where the sparsest solution is available by convex optimization rather than combinatorial optimization — provided the answer has fewer nonzeros than half the number of equations. We mention examples involving incompletely-observed Fourier transforms and Laplace transforms.

Second, results on classical neighborliness of high-dimensional randomly-projected simplices imply that, if A is a typical uniformly-distributed random orthoprojector with $n = 2d$ and n large, the sparsest nonnegative solution to $y = Ax$ can be found by linear programming provided it has fewer nonzeros than 1/8 the number of equations.

We also consider a notion of weak neighborliness, in which the overwhelming majority of k-sets of a_j's not containing 0 span a face. This implies that most nonnegative vectors x with k nonzeros are uniquely determined by $y = Ax$. As a corollary of recent work counting faces of random simplices, it is known that most polytopes P generated by large n by $2n$ uniformly-distributed orthoprojectors A are weakly k-neighborly with $k \approx .558n$. We infer that for most n by $2n$ underdetermined systems having a sparse solution with fewer nonzeros than roughly half the number of equations, the sparsest solution can be found by linear programming.

Key Words and Phrases: Neighborly Polytopes. Cyclic Polytopes. Underdetermined Systems of Linear Equations. Linear Programming. Combinatorial Optimization. Convex Hull of Gaussian Samples. Positivity constraints in ill-posed problems.

Acknowledgements. DLD had partial support from NSF DMS 00-77261, and 01-40698 (FRG); and projects with NIH and ONR-MURI. DLD thanks MSRI and its 'neighborly' hospitality in the Winter 2005, while this was prepared, and the Clay Mathematics Institute for a Senior Scholar appointment. JT would like to acknowledge NSF fellowship DMS 04-03041.

1 Introduction

Consider an underdetermined system of linear equations $y = Ax$, where $y \in \mathbf{R}^d$, $x \in \mathbf{R}^n$, A is a $d \times n$ matrix, $d < n$ and y is considered known but x is unknown. In this paper only nonnegative solutions $x \geq 0$ are of interest. Enthusiasts of parsimony seek the *sparsest* solution – the one with fewest nonzeros. Formally, they consider

$$(NP) \qquad \min \|x\|_0 \text{ subject to } y = Ax, \quad x \geq 0.$$

Here the 0-'norm' $\|x\|_0$ counts the number of nonzeros. Because of the extreme non-convexity of the zero-'norm', (NP) is NP-hard in general. In this paper we consider the *convex* optimization problem

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

We will show that for many matrices A, whenever the solution to (NP) is sufficiently sparse, it is also the unique solution of (LP). As a general label, we call this phenomenon *NP/LP equivalence.*

We develop an understanding of this equivalence phenomenon using ideas from the theory of convex polytopes; the books of Grünbaum [18] and Ziegler [28] are useful starting points. Throughout the paper, we study a specific polytope P, definable in several equivalent ways. Let T^{n-1} denote the standard simplex in $\mathbf{R}^n$, i.e. the convex hull of the unit basis vectors e_i. Let T_0^n denote the solid simplex, i.e. the convex hull of T^{n-1} and the origin. We think of T^{n-1} as the *outward* part of T_0^n, i.e. the part one would see looking from 'outside'.

We focus attention in this paper on the convex polytope $P = AT_0^n \subset \mathbf{R}^d$. Equivalently, P is the convex hull of a certain pointset $\mathcal{A} \subset \mathbf{R}^d$, containing the columns of a_j, $j = 1, \dots, n$ of A, possibly together with the origin $a_0 = 0$; include the origin if it does not already belong to the convex hull of the $\{a_j\}_{j=1}^n$. For later use, set $N = \#\mathcal{A}$. Thus, $N = n$ if 0 belongs to the convex hull of the $\{a_j\}_{j=1}^n$, otherwise $N = n + 1$. Below we use the notation $T = T^{n-1}$ if $N = n$, and $T = T_0^n$, if $N = n + 1$. Then we may also write $P = AT$.

A general polytope Q is called *k-neighborly* if every set of k vertices spans a face of Q. Thus, all combinations of vertices generate faces. The standard simplex T^{n-1} is the prototypical neighborly object. The terminology and basic notions in neighborliness were developed by Gale [15, 16]; see also [18, 20, 28].

We modify this notion here, calling a polytope Q which contains 0 *outwardly k-neighborly* if every set of k vertices *not including the origin* 0 span a face. Roughly speaking, such a polytope behaves as a neighborly one except perhaps at any faces reaching the origin. Thus if Q is k-neighborly then it is also outwardly k-neighborly, but the notions are distinct. In addition outward k-neighborliness of AT_0^n implies k-neighborliness of AT^{n-1}, the 'outward part' of AT_0^n. Of course, when $0 \in AT^{n-1}$ neighborliness and outwardly neighborliness of $P = AT_0^n$ coincide. (Modification of neighborliness to exclude consideration of certain subsets of vertices has been useful previously; compare the notion of *central neighborliness* of centrosymmetric polytopes, where every k vertices not including an antipodal pair span a face; see the paper [7] for discussion and references.)

In Section 2 we connect neighborliness to the question of NP/LP equivalence.

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

- *The polytope P has N vertices and is outwardly k-neighborly,*
- *Whenever $y = Ax$ has a nonnegative solution x_0 having at most k nonzeros, x_0 is the unique nonnegative solution to $y = Ax$ and so the unique solution to (LP).*

Formalizing the notion of sparsity threshold of a matrix A, we see that LP/NP equivalence holds up to a certain *breakdown point*; namely, the largest value m such that every sparse vector with fewer than m nonzeros is the uniquely recovered by (LP). The highest value of k for which a polytope besides the d simplex can be k-neighborly is $\lfloor d/2 \rfloor$ [15, 16, 18]. The degree of outward neighborliness of AT_0^n is not better than the degree of neighborliness of AT^{n-1}. Hence if $n > d$, the equivalence breakdown point is not better than $\lfloor d/2 \rfloor + 1$.

1.1 Neighborly Polytopes

A polytope is called *neighborly* if it is k-neighborly for every $k = 1, \dots, \lfloor d/2 \rfloor$. Many families of neighborly polytopes are known. In Section 3, we use Theorem 1 and the structure of standard families of neighborly polytopes to give:

Corollary 1.1 *Let $d > 2$. For every $n > d$ there is a $d \times n$ matrix A such that NP/LP equivalence holds with breakdown point $\lfloor d/2 \rfloor + 1$.*

When we have a matrix A with this property, and a particular system of equations that must be solved, we can run (LP); if we find that the output has fewer nonzeros than half the number of equations, we infer that we have found the unique sparsest nonnegative solution.

For such matrices, *if it would be very valuable to solve (NP) – because the answer would be very sparse – we can solve it by convex optimization.* Conversely, *it is exactly in the cases where the answer to (NP) would not be very sparse that it might also be very expensive to compute!*

The standard examples of neighborly polytopes go back to Gale; some of these are reviewed in Section 3. They have interesting interpretations in terms of Fourier analysis and geometry of polynomials, and correspond to interesting matrices A. Section 3 shows how to apply them to get the above corollary and to get two results about inference in the presence badly incomplete data. The first concerns incomplete Fourier information:

Corollary 1.2 *Let $\mu^{(0)}$ be a nonnegative measure supported on some subset of the n known points $0 < t_1 < \cdots < t_n < 2\pi$. Let $\hat{\mu}_k$ denote the Fourier coefficient*

$$\hat{\mu}_k \equiv \sum_j \mu\{t_j\} \exp\{\sqrt{-1} \cdot k t_j\}.$$

Suppose that $y_k = \hat{\mu}_k^{(0)}$ is observed (without error) for $k = 1, \dots, m$, $2m < n$. If $\mu^{(0)}$ is supported on at most m points, the problem

$$\min \sum_j \mu\{t_j\} \quad \text{subject to } y_k = \hat{\mu}_k,\ k = 1, \dots, m; \qquad \mu\{t_j\} \geq 0, \quad j = 1, \dots n,$$

has $\mu^{(0)}$ as its unique solution.

Superficially, this problem seems improperly posed, since we have n unknowns – the mass of μ at each of the n points t_j – with only $2m < n$ data $\hat{\mu}_k$ to constrain them. Yet if the underlying object $\mu^{(0)}$ is sparsely supported, it is uniquely recoverable, in fact by convex optimization.

A parallel result can be given for partial Laplace transformation.

Corollary 1.3 *Let $\mu^{(0)}$ be a nonnegative measure supported on some subset of the n known points $-\infty < \tau_1 < \cdots < \tau_n < \infty$. Let $\tilde{\mu}_k$ denote the Laplace transform value*

$$\tilde{\mu}_k \equiv \sum_j \mu\{\tau_j\} \exp\{k \tau_j\}.$$

	$\delta = .1$	$\delta = .25$	$\delta = .5$	$\delta = .75$	$\delta = .9$
ρ_N	.060131	.087206	.133457	.198965	.266558
ρ_{VS}	.240841	.364970	.558121	.765796	.902596

Table 1: Phase transitions ρ_N and ρ_{VS} in strong and weak neighborliness

Suppose that $y_k = \tilde{\mu}_k^{(0)}$ *is observed (without error) for* $k = 1, \ldots, m$, $m < n$. *If* $\mu^{(0)}$ *is suported on at most* $m/2$ *points, the problem*

$$\min \sum_j \mu\{t_j\} \quad \text{subject to } y_k = \tilde{\mu}_k,\ k = 1, \ldots, m; \qquad \mu\{\tau_j\} \geq 0, \quad j = 1, \ldots n$$

has $\mu^{(0)}$ *as its unique solution.*

This problem again seems improperly posed, since we have n unknowns but only $m < n$ (real) data. Yet if $\mu^{(0)}$ is sparsely supported, it is uniquely recoverable, again by linear programming.

These corollaries are stated in a form that would be familiar to those in the signal-processing community, and in that form are not new. The first follows from a result in Donoho, Johnstone, Hoch and Stern [10, Theorem 3]; both follow from recent work by Jean-Jacques Fuchs [14] which we learned of after this paper was first submitted. In fact the basic idea behind both results was known to Carathéodory [4, 5] although not stated in this form.

These corollaries are proved here using the neighborliness of polytopes deriving from special curves in $\mathbf{R}^d$. This illustrates the point made by Theorem 1: results about neighborliness are *equivalent to* results about uniqueness. In some polytope literature confusion seems to have arisen, suggesting for example, that Carathéodory knew about neighborly polytopes, discounting Gale's contribution. In our view, Carathéodory pioneered uniqueness, Gale pioneered neighborliness, and others conflated the two, in effect anticipating the equivalence implied by Theorem 1, but apparently without making that equivalence explicit.

1.2 Random Polytopes

When introducing the neighborliness concept, Gale suggested that 'most' polytopes are neighborly [15]. Recently, the authors [11] studied neighborliness of random polytopes, considering high-dimensional cases $d_n = \lfloor \delta n \rfloor$, n large. They derived a function ρ_N such that polytopes P with n Gaussian-distributed vertices in $\mathbf{R}^d$ were roughly $\rho_N(d/n) \cdot d$-neighborly for large n. Thus, if $n = 2d$, they found $\rho_N(d/n) \approx .133$; compare Table 1. Applying their results gives

Corollary 1.4 *Fix* $\epsilon > 0$. *Let* $A_{d,n}$ *denote a random* $d \times n$ *matrix with columns drawn independently from a multivariate normal distribution on* $\mathbf{R}^d$ *with nonsingular covariance matrix. Suppose* d *and* n *are proportionally-related by* $d_n = \lfloor \delta n \rfloor$. *Then, with overwhelming probability for large* n, $A_{d,n}$ *offers the property of LP/NP equivalence up to breakdown point* $\geq (\rho_N(\delta) - \epsilon)d$.

Line 1 of Table 1 gives results for different aspect ratios $\delta = d/n$ of the nonsquare matrix A. Thus if $n = 10d$, so the corresponding system is underdetermined by a factor 10, the typical matrix A with Gaussian columns offers LP/NP equivalence up to breakdown point exceeding $.06d$. For the typical A and for *every* problem instance y generated by a sparse vector x with nonzeros $\leq .06$ times the number of equations, (LP) delivers the sparsest solution.

1.3 Weak Neighborliness and Weak Equivalence

The notion of NP/LP equivalence developed in Theorem 1 demands, for a given A, equivalence at *all* problem instances (y, A) generated by *any* nonnegative sparse vector x_0 with at most k nonzeros. A weaker notion considers equivalence merely for *most* such problem instances. This idea is developed in Section 4 below, where it is shown that for matrices A where the corresponding pointset $\mathcal{A}$ is in general position NP/LP equivalence at a certain instance $y = Ax_0$ depends only on the support of x_0 and not the values of x_0 on its support. Hence, we define a measure on problem instances by simply counting the fraction of support sets of size k with a given property. We then meaningfully speak of a given A offering *NP/LP equivalence for most problem instances having nonnegative sparse solutions with most k nonzeros.*

We can also define two weaker notions of classical (resp. outward) neighborliness, saying that the polytope P is (k, ϵ)*-weakly neighborly* (resp. *weakly outwardly neighborly*) if, among all k-membered subsets of vertices (resp. among those not including 0), all except a fraction ϵ span $k-1$-faces of P. As it turns out, if the points $\mathcal{A}$ are in general position, weak neighborliness of P is the same thing as saying that $P = AT$ has at least $(1 - \epsilon)$ times as many $(k - 1)$-dimensional faces as T. Hence, the notion of weak-neighborliness is really about numbers of faces. We say that a face is *zerofree* if 0 does not occur as a vertex.

Theorem 2 *Let A be a $d \times n$ matrix, $d < n$ with pointset $\mathcal{A}$ in general position. For $1 \leq k \leq d - 1$, these two properties of A are equivalent*

- *The polytope $P = AT$ has at least $(1 - \epsilon)$ times as many zerofree $(k - 1)$-faces as T,*
- *Among all problem instances (y, A) generated by some nonnegative vector x_0 with at most k nonzeros, the solutions to (NP) and to (LP) are identical, except in a fraction $\leq \epsilon$ of instances.*

The authors [11], in recent work on high-dimensional random polytopes counted the faces of randomly-projected simplices. Building on work of Affentranger and Schneider [1] and especially Vershik and Sporyshev [27] they considered the case where d and n are large and proportional, and were able to get precise information about the phase transition between prevalence and scarcity of weak-neighborliness as k increases from 1 to $d - 1$. They studied a function ρ_{VS} (in honor of Vershik and Sporyshev who first implicitly characterized it) that maps out the phase transition in weak-neighborliness. Fix $\epsilon > 0$ and consider n large. Weak-neighborliness typically holds for $k < \rho_{VS}(d/n) \cdot d \cdot (1-\epsilon)$, while for $k > \rho_{VS}(d/n) \cdot d \cdot (1+\epsilon)$, weak neighborliness typically fails. They also showed that the same conclusions hold for weak outward neighborliness as for weak neighborliness. Numerical results are given in Table 1, in particular, the second line, where $\rho_{VS}(.1) \approx .24$. Informally, for *most* 10-fold underdetermined matrices A and *most* vectors with fewer nonzeros than 24% of the number of rows in A, the sparsest nonnegative solution can be found by (LP). In contrast, $\rho_N(.1) \approx .06$. Informally, if for a typical matrix A we insist that every instance of (NP) with a sufficiently sparse solution be solvable by (LP), then 'sufficiently sparse' must mean at most 6% d.

As a corollary, we obtain the following. Let $S_+(d, n, k)$ denote the collection of *all* systems of equations (y, A) having a nonnegative solution x_0 with at most k nonzeros. When A is a matrix with columns in general position, equivalence between (NP) and (LP) depends only on the support of x_0, as discussed in Lemma 4.2. Place a probability measure on $S_+(d, n, k)$ which makes the nullspace of A uniformly distributed among $n - d$ subspaces of $\mathbf{R}^n$ and which makes the support of the sparsest solution uniform on k-subsets of n objects. Using Table 1's entry showing $\rho_{VS}(1/2) > .558$, we have the following:

Corollary 1.5 *Consider the systems of equations* (y, A) *in* $S_+(n, 2n, .558n)$*. For* n *large, the overwhelming majority of such* (y, A) *pairs exhibit NP/LP equivalence.*

1.4 Contents

Section 2 proves Theorem 1, while Section 3 explains how Corollaries 1.1-1.2-1.3-1.4 follow from Theorem 1 and existing results in polytope theory. Section 4 studies weak neighborliness and justifies Corollary 1.5. Section 5 discusses (LP) in settings not neighborly in the usual sense, extensions to noisy data, and extensions to situations when nonnegativity is not enforced. Positivity is seen to be a powerful constraint.

2 Equivalence

2.1 Preliminaries

To begin, we relate (LP) to the polytope P. Note that the value of (LP) is a function of $y \in R^d$:

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

Note also that V is homogeneous: $V(ay) = aV(y)$, $a > 0$. We have defined the polytope $P = AT$ so that it is simply the 'unit ball' for V:

$$P = \{y : y \in AR^n_+ \text{ and } V(y) \leq 1\}.$$

To see this, write conv for the convex hull operation. The convexity and homogeneity of V guarantees that the right side is $\mathrm{conv}(\{0\} \cup \{a_j\}_{j=1}^n)$. We have defined P by cases; if $0 \in \mathrm{conv}(\{a_j\}_{j=1}^n)$, $P = AT^{n-1}$; otherwise $P = AT^n_0$. In each case $P = \mathrm{conv}(\{0\} \cup \{a_j\}_{j=1}^n)$.

We call *subconvex* combination a linear combination with nonnegative combinations summing to at most one. The previous paragraph can be reformulated so:

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

We adopt standard 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 $x \in P$, $x \neq c$. Thus $P = \mathrm{conv}(\mathrm{vert}(P))$. Vertices are just 0-dimensional faces, and a k-dimensional face of P is a k-dimensional set $F \subset P$ 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

$$\mathrm{vert}(AT) \subset A\,\mathrm{vert}(T), \tag{2.1}$$

which implies

$$\mathcal{F}_\ell(AT) \subset A\mathcal{F}_\ell(T), \qquad 0 \leq \ell < d; \tag{2.2}$$

and so the numbers of vertices obey

$$f_0(AT) \leq f_0(T). \tag{2.3}$$

2.2 Basic Insights

Theorem 1 involves two insights recorded here without proof. Similar lemmas were recently proven in [7]. The first explains the importance and convenience of having simplicial faces of P.

Lemma 2.2 (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.

The second insight: outward k-neighborliness can be thought of as saying that the low-dimensional zerofree faces of P are simply images under A of the faces of T^{n-1}, and hence simplices.

Lemma 2.3 (Alternate Form of Neighborliness). *Suppose the polytope $P = AT$ has N vertices and is outwardly k-neighborly. Then*

$$\forall \ell = 0, \ldots, k-1, \quad \forall\, F \in \mathcal{F}_\ell(T^{n-1}), \quad AF \in \mathcal{F}_\ell(AT). \tag{2.4}$$

Conversely, suppose that (2.4) holds; then $P = AT$ has N vertices and is outwardly k-neighborly.

2.3 Theorem 1, Forward Direction

We suppose that P is outwardly k-neighborly, that the nonnegative vector x_0 has at most k nonzeros, and show that the unique solution of (LP) is precisely x_0. We assume without loss of generality that the problem is scaled so that $1'x_0 = 1$; thus $x_0 \in T^{n-1}$.

Now since x_0 has at most k nonzeros, it belongs to a $k-1$-dimensional face F of the simplex: $F \in \mathcal{F}_{k-1}(T^{n-1})$. Hence y belongs to AF, which, by outward neighborliness and Lemma 2.3, is a $k-1$-dimensional face of P. Now, by Lemma 2.2, 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

$$1'x \leq 1.$$

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

2.4 Theorem 1, Converse Direction

By hypothesis, A has the property that, for every $y = Ax_0$ where x_0 has no more than k nonzeros, x_0 is the unique solution to the instance of (LP) generated by y. We will show that P has n vertices and is k-neighborly.

By considering the case $k = 1$ with every $x_i = e_i$, we learn that in each case the corresponding $y_i = Ax_i$ belongs to P and is uniquely representable among subconvex combinations of $(a_j)_{j=1}^n$ simply by a_i. This implies by Lemma 2.2 above that each y_i is a vertex of P, so P has at least n vertices. Now if $0 \notin \mathrm{conv}\{a_j\}_{j=1}^n$, 0 is also a vertex of P. Since by (2.3) the number of vertices of $P = AT$ is at most the number of vertices of T, we see that P has exactly N vertices.

Consider now $k > 1$, and a collection of k disjoint indices $i_1, \dots i_k$, $1 \le i_\ell \le n$. By hypothesis, for every x_0 of the form

$$x_0 = \sum_{\ell=1}^{k} \alpha_\ell e_{i_\ell},$$

with $\alpha_\ell \ge 0$ and $\sum_\ell \alpha_\ell = 1$, the corresponding problem (LP) 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 subconvex combination of columns of A, and that solution is provided by the corresponding x_0. All the x_0 under consideration populate a face F of T^{n-1}, determined by $i_1, \dots i_k$. By the converse part of Lemma 2.2, AF is a face in $\mathcal{F}_{k-1}(AT)$.

Combining the last two paragraphs with the converse part of Lemma 2.3, we conclude that P has N vertices and is outwardly k-neighborly. □

3 Corollaries

We first mention a standard fact about convex polytopes: [15], [18, Chapter 7].

Theorem 3.1 *For every $n > d > 1$ there are $\lfloor d/2 \rfloor$-neighborly polytopes in $\mathbf{R}^d$ with n vertices.*

Examples are provided by the *cyclic polytopes*, which come in two standard families:

- *'Moment Curve' Cyclic Polytopes.* Let $0 \le t_1 < \cdots < t_n < \infty$, and let the j-th column of the $d \times n$ matrix A be given by

  $$a_j = M(t_j), \quad j = 1, \dots, n,$$

  where $M : \mathbf{R}_+ \mapsto \mathbf{R}^d$ is the so-called moment curve

  $$M(t) = (t, t^2, \dots, t^d)^T.$$

  The polytope obtained from the convex hull of the $(a_j)_{j=1}^n$ is $\lfloor d/2 \rfloor$ neighborly; see Gale [16]. Note that A is a kind of non-square Vandermonde matrix.

- *'Trigonometric' Cyclic Polytopes.* Let $0 < t_1 < \cdots < t_n < 2\pi$, and, for $d = 2m$, let the j-th column of the $d \times n$ matrix A be given by $a_j = F(t_j)$ where $F : [0, 2\pi) \mapsto \mathbf{R}^d$ is the trigonometric moment curve

  $$F(t) = (\cos(t), \sin(t), \cos(2t), \sin(2t), \dots, \cos((d/2)t), \sin((d/2)t))^T.$$

  The polytope obtained from the convex hull of the $(a_j)_{j=1}^n$ is $\lfloor d/2 \rfloor$-neighborly, again see [16]. Note that A is a kind of non-square Fourier matrix.

Existing proofs of neighborliness of moment curve polytopes [16, 18], after a simple adaptation, give Corollary 1.1. Note that $t_1 = 0$ is an allowable value for which a polytope $\mathrm{conv}\{M(t_j)\}$ is $\lfloor d/2 \rfloor$-neighborly; since $M(0) = 0$, it follows that, for any specific choice for t_1, $P = \mathrm{conv}(\{0\} \cup \{M(t_j)\}_{j=1}^n)$ is $\lfloor d/2 \rfloor$ neighborly, as, when $t_1 \ne 0$, one could view P as $\mathrm{conv}\{M(t_j)\}_{j=0}^n$ with $t_0 = 0$. We conclude that every $P = \mathrm{conv}(\{0\} \cup \{M(t_j)\}_{j=1}^n)$ is outwardly-neighborly. Hence, defining the matrix $A = [M(t_1), \dots, M(t_n)]$ we get (LP)-(NP)-equivalence up to breakdown point $\lfloor d/2 \rfloor$. Corollary 1.1 follows.

Corollary 1.3 also follows from the outward-neighborliness of $P = \mathrm{conv}(\{0\} \cup \{M(t_j)\}_j)$. Let $y_k = \tilde{\mu}_k^{(0)}$. Represent $\mu^{(0)}$ by a vector x_0 with n entries, the j-th one representing $\mu^{(0)}\{\tau_j\}$. Define $t_j = \exp(\tau_j)$, $j = 1, \ldots, n$, and note that $y = Ax_0$ where A is the partial Vandermonde matrix associated with the moment curves above. Since the polytope associated to A is $\lfloor d/2 \rfloor$-outwardly-neighborly, if the measure $\mu^{(0)}$ is supported in no more than $\lfloor d/2 \rfloor$ points, it is uniquely recovered from data y by solving (LP).

To obtain Corollary 1.2, we apply the neighborliness of trigonometric cyclic polytopes in the Appendix, proving

Lemma 3.1 *The polytope conv$(\{0\} \cup \{F(t_j)\})$ is outwardly $\lfloor d/2 \rfloor$-neighborly.*

Applying this, we can obtain Corollary 1.2. Break the m observed complex data into real parts and imaginary parts, giving a vector y of length $d = 2m$. Since $\mu^{(0)}$ is a nonnegative measure supported at $0 < t_1 < \cdots < t_n < 2\pi$, represent it as a vector x_0 with j-entry $\mu^{(0)}\{t_j\}$. The data y are related to the vector x_0 through $y = Ax_0$ where A is the above partial Fourier matrix. The corresponding polytope is neighborly. Hence if the nonnegative vector x_0 has no more than $m = d/2$ nonzeros, it will be uniquely reconstructed (despite $n > d$) from the data y by (LP). (As stated earlier, Corollary 1.2 also follows from [10, Theorem 3]; in fact the underlying calculation in the proof of Theorem 3 in [10] can be seen to be the same as the 'usual' one in proving neighborliness of trigonometric cyclic polytopes, although at the time of [10] this connection was not known.) After this paper was originally submitted, the authors learned of work by Jean-Jacques Fuchs [14] also implying Corollaries 1.2-1.3.

A wide range of neighborly polytopes is known. A standard technique (already used in the two examples above) is to take n points on a curve $C : \mathbf{R} \mapsto \mathbf{R}^d$ [6, 22]. The curve must be a so-called *curve of order d*, meaning that each hyperplane of $\mathbf{R}^d$ intersects the curve in at most d points. This construction is of course intimately connected with the theory of Moment Spaces and with unicity of measures having specified moments [19]. Constructions based on oriented matroids and totally positive matrices have also been made by Sturmfels; see [25, 24]. In the context of this paper, we note that *if such a curve passes through the origin*, then of course $\mathrm{conv}(\{0\} \cup \{C(t_j)\})$ is neighborly, and so outwardly-neighborly as well. However, as Lemma 3.1 shows, outward neighborliness is possible even when such a curve does not pass through the origin,

Sturmfels has even shown that (for even d) in some sense curves of order d offer the 'only' example of neighborly polytopes (up to isomorphism). In short, it is known that polytopes offering full $\lfloor d/2 \rfloor$ neighborliness are special.

What is the generic situation? Gale [15] proposed that in some sense 'most' polytopes are neighborly. Goodman and Pollack [1] proposed a natural model of 'random polytope' in dimension d with n vertices. They suggested to take the standard simplex T^{n-1} and apply a uniformly-distributed random projection, getting the random polytope $P = AT^{n-1}$. Vershik and Sporyshev considered this question in the case where d and n increase to ∞ together in a proportional way. The companion paper [11] revisits the Vershik-Sporyshev model, asking about neighborliness of the resulting high-dimensional random polytopes. It proves:

Theorem 3 *Let $0 < \delta < 1$, 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 orthogonal projection. There is $\rho_N(\delta) > 0$ so that, for $\rho < \rho_N(\delta)$, with overwhelming probability for large n, $P = AT^{n-1}$ is $\lfloor \rho d \rfloor$-neighborly.*

Thus, typical Goodman-Pollack polytopes have neighborliness 'proportional to dimension'. (This result permits, but does not imply, that polytopes are not *fully* neighborly; i.e. the fact that $\rho_N < .5$ allows the possibility that k-neighborliness may not hold up to the upper limit $k = \lfloor d/2 \rfloor$. The lack of full neighborliness for $\delta < .42$ can be inferred from the lack of $d/2$-weak neighborliness described below.)

The Goodman-Pollack model is broader than at first appears. By a result of Baryshnikov and Vitale [2], P is affinely equivalent to the convex hull of a Gaussian random sample. We can conclude that

Corollary 3.1 *Let $A = A_{d,n}$ denote a random $d \times n$ matrix with columns a_j, $j = 1, \dots, n$ drawn independently from a multivariate normal distribution on $\mathbf{R}^d$ with nonsingular covariance. Suppose d and n are proportionally-related by $d_n = \lfloor \delta n \rfloor$. Let $\rho < \rho_N(\delta)$. Then, with overwhelming probability for large n, $P = conv\{a_j\}_{j=1}^n$ is $\lfloor \rho d \rfloor$-neighborly.*

The companion paper [11] implies that the preceding two results hold just as written also for $P = AT_0^n$, and $P = \mathrm{conv}(\{0\} \cup \{a_j\})$ respectively, when 'neighborly' is replaced by 'outwardly neighborly'. Corollary 1.4 follows.

4 Weak Neighborliness and Probabilistic Equivalence

4.1 Individual Equivalence and General Position

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

For general A the task of describing such points may be very complicated; we adopt a simplifying assumption. Recall the definition of $\mathcal{A}$: Let $a_0 = 0$ and, if $0 \notin \mathrm{conv}\{a_j\}_{j=1}^n$, let $\mathcal{A} = \{a_j\}_0^n$. Otherwise let $\mathcal{A} = \{a_j\}_1^n$. We say that *that $\mathcal{A}$ is in general position* in $\mathbf{R}^d$ if no k-plane of $\mathbf{R}^d$ contains more than $k+1$ a_j's (i.e. viewing the a_j as points of $\mathbf{R}^d$). Under this assumption, the face structure of P is very easy to describe. A remark in [20, Page 81] (compare also [7]) proves

Lemma 4.1 *Suppose that $\mathcal{A}$ is in general position. Then for $k \le d-1$, the k-dimensional faces of $P = conv(\mathcal{A})$ are all simplicial.*

Recalling Lemma 2.2, it follows that, when $\mathcal{A}$ is in general position, whenever y belongs to a k-dimensional face of P with $k \le d-1$, there is a corresponding unique solution of (LP). This remains true for every y in that same face of P, and the unique solution involves a convex combination of the vertices of that same face. The vertices are identified with members of $\mathcal{A}$. Those members are identified either with the origin or with certain canonical unit basis vectors of $\mathbf{R}^n$. Hence, the collection of such convex combinations of vertices is in one-one correspondence with points in a specific k-face of T. Moreover, by the uniqueness in Lemma 2.2, a k-face of T can arise in this way in association with only one k-face of P. Hence for $k \le d-1$, we have a bijection between k-faces of P, and a subset $\mathcal{S}_k$ of the k-faces of T. We think of $\mathcal{S}_k$ as the subset of k-faces of T destined to survive as faces under the projection $T \mapsto AT$ onto $\mathbf{R}^d$.

The k-faces of T are in bijection with the supports of the vectors belonging to those faces. Since two vectors x_0 and x_1 with unit sum and with common support belong to the same face of T, and since each face as a whole survives or does not survive projection, we conclude:

Lemma 4.2 *Suppose that $\mathcal{A}$ is in general position and that x_0 has at most $d-1$ nonzeros. The property of individual equivalence depends only on the support of x_0; if x_0 and x_1 have nonzeros in the same positions, then they are either both points of individual equivalence or neither points of individual equivalence.*

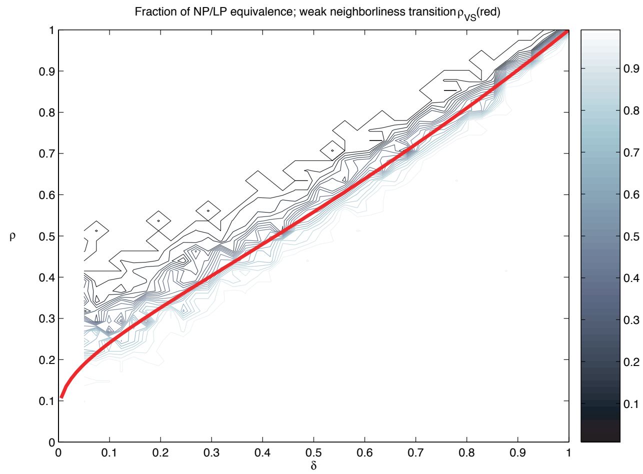

Figure 1: Empirical verification of the NP/LP equivalence phase transition as a function of δ with $d_n = \lfloor \delta n \rfloor$ and sparsity $k = \lfloor \rho d_n \rfloor$ in the case of $n = 200$. The fraction of successes in (LP) recovering (NP) in gray scale, and the calculated weak neighborliness transition curve $\rho_{VS}(\delta)$ overlaid in red. Note that weak neighborliness exceeds $d/2$ for $\delta > .425$; see Subsection 4.3.

There are of course $\binom{n}{k}$ supports of size k. This gives us a natural way to measure 'typicality' of individual equivalence.

Definition 1 *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 individual equivalence if individual equivalence holds for a fraction $\geq (1-\epsilon)$ of all supports of size k.*

A practical computer experiment can be conducted to approximate ϵ for a given A and k. One randomly generates a sparse vector x_0 with randomly-chosen support and arbitrary positive values on the support. One forms $y = Ax_0$, and solves (LP). Then one checks whether the solution of (LP) is again x_0. $\epsilon(A, k)$ can be estimated by the fraction of computer experiments where failure occurs. Experiments of this kind reveal that for A a typical random $d \times 2d$ orthoprojector, individual equivalence is typical for $k < .558d$. See Figure 1, which shows that the experimental outcomes track well the prediction ρ_{VS}.

4.2 Individual Equivalence and Face Numbers

We are now in a position to prove Theorem 2 using the above lemmas. For a polytope Q possibly containing 0 as a vertex, $\tilde{f}_k(Q)$ denote the number of *zerofree* k-faces, i.e. the number of faces of Q not having 0 as a vertex. Restating Theorem 2 in the terminology of this section we have:

Theorem 4 *Let $\mathcal{A}$ be in general position. These statements are equivalent for $k < d$:*

- *The zerofree face numbers of AT and T agree within a factor $1 - \epsilon$:*

$$(1-\epsilon)\tilde{f}_{k-1}(T) \leq \tilde{f}_{k-1}(AT) \leq \tilde{f}_{k-1}(T).$$

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

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

$$\frac{\#(\text{supports giving equivalence})}{\#(\text{supports of size } k)} = \frac{\tilde{f}_{k-1}(AT)}{\tilde{f}_{k-1}(T)}.$$

□

Of course, counting faces of polytopes is an old story. This result points to a perhaps surprising *probabilistic* interpretation. Suppose the points in $\mathcal{A}$ are in general position. We randomly choose a nonnegative vector x with $k < d$ nonzeros in such a way that all arrangements of the nonzeros are equally likely; the distribution of the amplitudes of the nonzeros can be arbitrary. We then generate $y = Ax$. If the quotient polytope P has 99% as many $(k-1)$-faces as T, then there is a 99% chance that x is both the sparsest nonnegative representation of y and also the unique nonnegative representation of y. This is a quite simple and, it seems, surprising outcome from mere face counting.

4.3 Interpreting Table 1

The authors' paper [11] derives numerical information about the Vershik-Sporyshev phase transition $\rho_{VS}(\delta) > 0$, i.e. the transition so that for $\rho < \rho_{VS}(\delta)$, the $\lfloor \rho d \rfloor$-dimensional face numbers of AT^{n-1} are the same as those of T to within a factor $(1 + o_P(1))$, while for $\rho > \rho_{VS}(\delta)$ they differ by more than a factor $(1 + o_P(1))$. They show that the same conclusion holds for the zerofree face numbers of AT_0^n.

Obviously $\rho_N(\delta) \leq \rho_{VS}(\delta)$. Fixing some small $\epsilon > 0$, we have with overwhelming probability for large d that

$$\begin{array}{lll} P = AT & \text{is} & (\tilde{\rho}_N \cdot d)\text{-outwardly -neighborly, and} \\ & & (\tilde{\rho}_{VS} \cdot d, \epsilon)\text{-weakly-outwardly -neighborly;} \end{array}$$

here $\tilde{\rho}_N \equiv \rho_N(\delta) - \epsilon$, and $\tilde{\rho}_{VS} \equiv \rho_{VS}(\delta) - \epsilon$ obey

$$0 < \tilde{\rho}_N \approx \rho_N(\delta) < \tilde{\rho}_{VS} \approx \rho_{VS}(\delta).$$

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

- ρ_N, the smaller, is still fairly large, perhaps surprisingly so. While it tends to zero as $\delta \to 0$, it does so only at a rate $O(1/\log(1/\delta))$; and for moderate δ it is on the other of .1.
- ρ_{VS} is substantially larger than ρ_N. The fact that it 'crosses the line' $\rho = 1/2$ for δ near .425 is noteworthy; this means that while a polytope can only be $\lfloor d/2 \rfloor$ neighborly, it can be $> d/2$ weakly neighborly! In fact we know $\rho_{VS}(\delta) \to 1$ as $\delta \to 1$ [27, 11]. For $\epsilon > 0$ and δ sufficiently close to 1, for sufficiently large d, typical weak neighborliness can exceed $d(1-\epsilon)$! This is an important difference between neighborliness and weak neighborliness, and is the source of Corollary 1.5.

5 Discussion

5.1 When $\mathcal{A}$ is not in General Position

There are interesting problems where $\mathcal{A}$ is not in general position, but (LP) saves the day. Here is an example, based on the dictionary of dyadic intervals.

Let $n = 2^J$ be dyadic (i.e. let J be integral), and define the dyadic subintervals of $\{0, \dots, n-1\}$ recursively, with $\chi_{0,0} = \{0, \dots, n-1\}$, $\chi_{1,0}$ the left half of $\chi_{0,0}$, and $\chi_{1,1}$ the right half of $\chi_{0,0}$. In general, we view $\chi_{j,k}$ as a parent to be split into two equal children - $\chi_{j+1,2k}$, the left half, and $\chi_{j+1,2k+1}$ the right half. This defines a family $0 \le j \le J$, $0 \le k < 2^j$.

We can use these dyadic intervals to cover arbitrary nondyadic intervals $I_{ab} = \{a \le t < b\}$, here a and b are integers with $0 \le a < b \le n-1$. There are of course many ways that a nondyadic interval can be covered by dyadic ones.

A very attractive and historically important approach is provided by Whitney covering. For a given interval I, consider the *maximal* dyadic intervals $\chi_{j,k} \subset I$. Here by maximal, we mean dyadic intervals which are subsets of I but whose parents are not subsets of I. Clearly the collection of such maximal dyadic intervals covers I. The reader will want to check the following:

Lemma 5.1 *The Whitney covering of I is the sparsest decomposition into dyadic intervals.*

In effect, this is the reason for having the Whitney decomposition! In particular, the Whitney covering has at most $2J = O(\log(n))$ terms, compared to the 'worst' representation, which can have as many as $O(n)$ terms.

Consider now the matrix A whose columns are the $2n-1$ different indicators of dyadic intervals. In detail, let $\tilde{a}_{jk}$ be a normalized indicator vector for χ_{jk}, i.e. an n-vector which is 0 for t outside $\chi_{j,k}$ and $2^{-(J-j)/2}$ for t inside $\chi_{j,k}$. Note that $\|\tilde{a}_{jk}\|_2 = 1$. Now A is the matrix whose columns are an enumeration of the $\tilde{a}_{j,k}$; it is n by $2n-1$. Note that the columns of A are not in general position, since the parent-child set decomposition $\chi_{j,k} = \chi_{j+1,2k} \cup \chi_{j+1,2k+1}$ implies the parent-child linear dependency

$$\tilde{a}_{j,k} = (\tilde{a}_{j+1,2k} + \tilde{a}_{j+1,2k+1})/\sqrt{2}.$$

Hence even sparse representations using A can never be unique. Nevertheless, we will see that (LP) is an advantageous approach.

Consider vectors $y = y_{ab}$ which are indicators of not-necessarily dyadic intervals I_{ab}. Any such interval has a representation $y = Ax$ in terms of nonnegative superpositions of dyadic intervals. These range from sparse to very nonsparse. At one extreme, y can be represented in terms of superpositions of $(b-a)$ singletons; at the other extreme it can be represented by using at most $2J$ well-chosen dyadic intervals .

Lemma 5.2 *The solution $y_{ab} = Ax$ obtained from (LP) has nonzeros corresponding to the maximal dyadic intervals $\chi_{j,k} \subset I_{ab}$. Thus (LP) delivers the Whitney decomposition of $I_{a,b}$.*

Hence (LP) gives the sparsest representation available by dyadic intervals; it has at most $2J = O(\log(n))$ terms, compared to the 'pixel' representation, which has as many as $O(n)$ terms.

Note that here (a) the columns of A are not in general position and (b) the representation of y is not unique, but (LP) and (NP) are still equivalent.

5.2 Noise tolerance

In general, 'real' data contain noise. The (LP) viewpoint can accommodate this naturally, by relaxing exact equality to an approximation:

$$(LP_\epsilon) \qquad \min 1'x \text{ subject to } \|y - Ax\|_2 \le \epsilon, \quad x \ge 0.$$

This is again a convex optimization problem. It was studied in the case of a partial Fourier matrix in [10, Theorem 3].

There are natural 'noise cognizant' variants of Corollaries 1.2 and 1.3 from the introduction. In fact for more general A one can show that the solution $x_{1,\epsilon}$ of (LP_ϵ) gives a stable recovery of x_0. Formally, the result is:

Theorem 5 *Let A be a matrix giving an outwardly k-neighborly polytope P. Suppose x_0 has no more than k nonzeros. There is $C = C(A,k) < \infty$ so that $\|y - Ax_0\|_2 \leq \epsilon$ implies*

$$\|x_{1,\epsilon} - x_0\|_2 \leq C\epsilon, \quad \epsilon > 0.$$

This can be inferred from the above results by 'soft' means. Such a result has been given previously in the partial Fourier case (extending Corollary 1.2) some time ago in [10, Theorem 3]. That reference provides a figure illustrating a geometric interpretation of this result, giving an immediate 'visual' proof.

5.3 Removing the Nonnegativity Constraint

The articles[8], [7], discussed the case where x can take positive or negative values. The analogs of (NP) and (LP) become

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

and

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

In both problems the condition $x \geq 0$ is absent. Hence the sum in (LP) has been replaced by the ℓ^1 norm in (P_1). The problem (P_0) (which extends (NP)) is NP-hard in general; at the same time (P_1) can be posed as a standard linear program.

In this setting, there is an analog of Theorem 1. Consider the polytope P obtained as a convex hull of the $2n$ signed columns of A, $(\pm_j a_j)$. This polytope is centrally symmetric (i.e. invariant under reflection $y \mapsto -y$ and is called centrally-k-neighborly if every set of k columns not containing an antipodal pair spans a face. It turns out that if P has $2n$ vertices and is centrally k-neighborly, then, whenever (P_0) has a solution with at most k nonzeros, (P_1) has the same unique solution. So given a problem instance (y, A), we can solve the linear program (P_1) and if we obtain a solution with at most k nonzeros, we know that we have solved (P_0). There is very active research on such problems; for a sampling of different perspectives, see [3, 9, 12, 13, 17, 26, 23] and related work.

There are substantial quantitative differences between the breakdown points when nonnegativity constraints are present and when they are not. While many $\lfloor d/2 \rfloor$-neighborly polytopes are known, a result of McMullen and Shephard [21] assures us that the most we can hope for is a $\lfloor (d+1)/3 \rfloor$-centrally-neighborly polytope, and it is not known that this bound is attainable for large n and d. No general construction of maximally-centrally-neighborly polytopes is known. Hence there is no nice analog of Corollary 1.1, and positivity allows the neighborliness phase transition to jump from at most $\approx d/3$ to $\approx d/2$. Positivity is quite valuable.

The article [8] considered random polytopes obtained essentially from matrices A with normally-distributed columns. With phase transition bounds $\rho_W^{\pm}$ and $\rho_N^{\pm}$ defined for central-neighborliness analogously to ρ_{VS} and ρ_N, numerical results are shown in Table 2. While clearly $\rho_W^{\pm} \leq \rho_{VS}$ and $\rho_N^{\pm} \leq \rho_N$, the degree of quantitative difference is striking. Recalling Corollary 1.5, for most large n by $2n$ systems of equations in *nonnegative* unknowns having a 51% sparse solution, the solution of (LP) is the unique sparsest solution. A comparable statement for *general* unknowns would consider n by $\sqrt{2}n$ systems with 51% sparse solutions; this allows many fewer unknowns than in the case of nonnegative x. Again positivity is quantitatively quite valuable.

Table 2: Phase Transitions in the case where x_0 is sparse but may have both signs, using methods from [8].

	$\delta = .1$	$\delta = .25$	$\delta = .5$	$\delta = .75$	$\delta = .9$
$\rho_N^{\pm}$	.048802	.065440	.089416	.117096	.140416
$\rho_W^{\pm}$	.188327	.266437	.384803	.532781	.677258

Appendix

The argument for Lemma 3.1, constructs a cyclic polytope Q so that $P \subset Q$. It uses two lemmas. The first is strategic, the second tactical.

Lemma A.1 *Suppose that $0 \notin conv\{a_j\}$. Suppose that there exist $b_1, ..., b_L$ distinct from 0 so that*

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

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

Proof. By hypothesis, Q is k-neighborly, and has the a_j as vertices. Hence each $F \in \mathcal{F}_{k-1}(AT^{n-1})$ is a face of Q. But $P \subset Q$, and P has the a_j among its vertices. Each such face F is then also a $(k-1)$-face of P. $\Box$

Lemma A.2 *Let M be a perfect square and let $III \in R^M$ denote the 'comb' sequence which is one at integer multiples of $\sqrt{M}$. The discrete Fourier transform of III is again III.*

Proof. This is a well-known property of the 'Shah' function or 'Dirac Comb'. $\Box$

Now turn to Lemma 3.1. Let $\omega_k = 2\pi(k-1)/(m+1)^2$, $k = 0, \dots, (m+1)^2$. WLOG suppose that the ω_k are disjoint from the t_j's and consider

$$b_l = F(\omega_{(m+1)l}), \quad l = 0, ..., m.$$

Applying Lemma A.2, we have

$$0 = \sum_{\ell=0}^{m} b_l;$$

indeed the entries of the sum correspond to values of the Fourier transform of III at frequencies $0 < k < m + 1$, at which it is zero. Hence,

$$0 \in \text{conv}\{b_l\}.$$

Consider now

$$Q = \text{conv}(\{a_j\} \cup \{b_l\}).$$

This is a cyclic polytope in $\mathbf{R}^d$ with $n + m + 1$ vertices, and is $\lfloor d/2 \rfloor$-neighborly. Apply now Lemma A.1 to conclude that $P = \text{conv}(\{a_j\} \cup \{0\}$ is outwardly k-neighborly.

$\Box$

References

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

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

[3] 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.

[4] Constantin Carathéodory. Ueber den variabilitaetsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehman. *Math. Ann.*, 64:95–115, 1907.

[5] Constantin Carathéodory. Ueber den variabilitaetsbereich der fourierschen konstanten von positiven harmonischen funktionen. *Rend. Circ. Mat. Palermo*, 32:193–217, 1911.

[6] Douglas Derry. Convex hulls of simple space curves. *Canad. J. Math.*, 8:383–388, 1956.

[7] David L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. 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, 2005.

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

[10] David L. Donoho, Iain M. Johnstone, Jeffrey C. Hoch, and Alan S. Stern. Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B*, 54(1):41–81, 1992. With discussion and a reply by the authors.

[11] David L. Donoho and Jared Tanner. Neighborliness of randomly-projected simplices in high dimensions. Technical report, Department of Statistics, Stanford University, 2005.

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

[14] Jean-Jacques Fuchs. Sparsity and uniqueness for some specific underdetermined systems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, March 2005, Philadelphia, PA.*, 2005.

[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] S. Karlin and L.S. Shapley. *Geometry of Moment Spaces*, volume 21 of *Memoirs AMS*. American Mathematical Society, 1953.

[20] P. McMullen and G. C. Shephard. *Convex polytopes and the upper bound conjecture*. Cambridge University Press, London, 1971. Prepared in collaboration with J. E. Reeve and A. A. Ball, London Mathematical Society Lecture Note Series, 3.

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

[22] T. S Motzkin. Comonotone curves and polyhedra, abstract 111. *Bull. Am. Math. Soc.*, 63:35, 1957.

[23] Mark Rudelson and Roman Vershynin. Geometric approach to error-correcting codes and reconstruction of signals. Technical report, Department of Mathematics, University of California, Davis, 2005.

[24] Bernd Sturmfels. Neighborly polytopes and oriented matroids. *European J. Combin.*, 9(6):537–546, 1988.

[25] Bernd Sturmfels. Totally positive matrices and cyclic polytopes. In *Proceedings of the Victoria Conference on Combinatorial Matrix Analysis (Victoria, BC, 1987)*, volume 107, pages 275–281, 1988.

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

[27] 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.

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