Sparse nonnegative solution of underdetermined linear equations by linear programming

David L. Donoho* and Jared Tanner

Department of Statistics, Stanford University, Stanford, CA 94305-4065

Consider an underdetermined system of linear equations $y = Ax$ with known y and $d \times n$ matrix A. We seek the nonnegative x with the 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. We explain this by the theory of convex polytopes. Let a_j denote the jth column of A, $1 \le j \le n$, let $a_0 = 0$ and P denote the convex hull of the a_j. We say the polytope 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 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. We also consider weak neighborliness, where the overwhelming majority of k-sets of a_js not containing 0 span a face of P. This implies that most nonnegative vectors x with k nonzeros are uniquely recoverable from $y = Ax$ by linear programming. Numerous corollaries follow by invoking neighborliness results. For example, for most large n by $2n$ underdetermined systems having a solution with fewer nonzeros than roughly half the number of equations, the sparsest solution can be found by linear programming.

neighborly polytopes | cyclic polytopes | combinatorial optimization | convex hull of Gaussian samples | positivity constraints in ill-posed problems

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 article only nonnegative solutions $x \ge 0$ are of interest. Enthusiasts of parsimony seek the sparsest solution, the one with fewest nonzeros. Formally, they consider the optimization problem

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

Here the 0-norm $\|x\|_0$ counts the number of nonzeros. Because of the extreme nonconvexity of the zero-norm, (NP) is NP-hard in general. In this article, we consider the linear program

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

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 by using ideas from the theory of convex polytopes; the books of Grünbaum (1) and Ziegler (2) are useful starting points. Throughout the article, 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^n_0 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^n_0, i.e., the part one would see looking from "outside."

We focus attention in this article on the convex polytope $P = AT^n_0 \subset \mathbf{R}^d$. P also has a representation as the convex hull of a certain point set $\mathcal{A} \subset \mathbf{R}^d$ we refer to frequently. Specifically, let

$\mathcal{A}$ consist of the columns $a_j, j = 1, \ldots, n$ of A, possibly together with 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 $T = T^n_0$, if $N = n + 1$. Then also $P = AT$.

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 (3, 4); see also refs. 1, 2, and 5.

We modify this notion here, calling a polytope Q that contains 0 outwardly k-neighborly if all sets 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^n_0 implies neighborliness of AT^{n-1}, the outward part of AT^n_0. Of course, when $0 \in AT^{n-1}$ neighborliness and outwardly neighborliness of AT^n_0 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 ref. 6 for discussion and references.]

In Section 2 we connect outward 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 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 uniquely recovered by (LP). The highest value of k for which a polytope besides the simplex can be k-neighborly is $\lfloor d/2 \rfloor$ (1, 3, 4). Hence if $n > d$, equivalence breakdown must occur as soon as the number of nonzeros $k = \lfloor d/2 \rfloor + 1$.

1.1. Neighborly Polytopes. A polytope is called neighborly if it is k-neighborly for every $k = 1, \ldots, \lfloor d/2 \rfloor$. Many families of neighborly polytopes are known. In Section 3, we use *Theorem 1* and the existence of neighborly polytopes to give the following.

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 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!

Examples of neighborly polytopes go back to Gale (3, 4) and Motzkin (7); 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 of 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 $k = 1, \ldots, m, 2m < n$. If $\mu^{(0)}$ is supported on at most m points, the problem

$$\min \sum_j \mu\{t_j\} \quad \text{subject to } y_k = \hat{\mu}_k, \, k = 1, \ldots, m;$$

$$\mu\{t_j\} \geq 0, \, j = 1, \ldots, n,$$

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. This corollary was previously known to us; it follows from a result in ref. 8. It also follows from recent work by Fuchs (9).

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*

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

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

$$\min \sum_j \mu\{t_j\} \quad \text{subject to } y_k = \tilde{\mu}_k, \, k = 1, \ldots, m;$$

$$\mu\{\tau_j\} \geq 0, \, j = 1, \ldots, n$$

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 proved here by using the neighborliness of polytopes based on certain partial Fourier and partial Vandermonde matrices, respectively. They also follow from recent work by Fuchs (9), who gave a direct proof of uniqueness. We find the neighborliness connection is instructive; it makes available a whole range of similar examples, provides knowledge about the atypicality of such examples, and builds a bridge to a body of distinguished literature, going back as far as Carathéodory (10, 11).

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

Phase transition	$\delta = 0.1$	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 0.9$
ρ_N	0.060131	0.087206	0.133457	0.198965	0.266558
ρ_{VS}	0.240841	0.364970	0.558121	0.765796	0.902596

1.2. Random Polytopes. When introducing the neighborliness concept, Gale (3) suggested that "most" polytopes are neighborly. Recently, we (12) studied neighborliness of random polytopes, considering high-dimensional cases $d_n = \lfloor \delta n \rfloor$, *n* large. We 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 $n = 2d$, we found $\rho_N(d/n) \approx 0.133$; compare Table 1. Applying our results gives the following.

Corollary 1.4. *Fix $\varepsilon > 0$. 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 $\geq (\rho_N(\delta) - \varepsilon)d$.*

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 of 10, the typical matrix *A* with Gaussian columns offers LP/NP equivalence up to a breakdown point exceeding $0.06d$. For the typical *A* and for every problem instance *y* generated by a sparse vector *x* with nonzeros ≤ 0.06 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, where it is shown that for matrices *A* where the corresponding point set $\mathcal{A}$ is in general position NP/LP equivalence at a certain instance $y = Ax_0$ depends only on the support of x_0 and not on the values of x_0 for 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 the most *k* nonzeros.

We can also define two weaker notions of classical [respectively (resp.) outward] neighborliness, saying that the polytope *P* is (k, ε)-weakly neighborly (resp. weakly outwardly neighborly) if, among *k*-membered subsets of vertices (resp. 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 $(1 - \varepsilon)$ 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 point set $\mathcal{A}$ in general position. For $1 \leq k \leq d - 1$, the following two properties of A are equivalent.*

- *The polytope $P = AT$ has at least $(1 - \varepsilon)$ 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 (LP) are identical, except in a fraction $\leq \varepsilon$ of instances.*

In recent work on high-dimensional random polytopes (12), we counted the faces of randomly projected simplices. Building on work of Affentranger and Schneider (13) and especially Vershik and Sporyshev (14) we 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 $d-1$. We 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 $\varepsilon > 0$ and consider n large. Weak neighborliness typically holds for $k < \rho_{VS}(d/n) \cdot d \cdot (1 - \varepsilon)$, whereas for $k > \rho_{VS}(d/n) \cdot d \cdot (1 + \varepsilon)$, weak neighborliness typically fails. We 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 $\rho_{VS}(0.1) \approx 0.24$. 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, $\rho_N(0.1) \approx 0.06$. 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 makes the support of the sparsest solution uniform on k-subsets of n objects. Using Table 1's entry showing $\rho_{VS}(1/2) > 0.558$, we have the following.

Corollary 1.5. *Consider the systems of equations (y, A) in $S_+(n, 2n, 0.558n)$. For n large, the overwhelming majority of such (y, A) pairs exhibit NP/LP equivalence.*

1.4. Contents. Section 2 proves *Theorem 1*, and Section 3 explains how *Corollaries 1.1, 1.2, 1.3,* and *1.4* follow from *Theorem 1* and existing results in polytope theory. Section 4 studies weak neighborliness and justifies *Corollary 1.5*. Section 5, which is published as *Supporting Text* in the supporting information on the PNAS web site, 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 conv($\{0\} \cup \{a_j\}_{j=1}^n$). We have defined P by cases; if $0 \in$ conv($\{a_j\}_{j=1}^n$), $P = AT^{n-1}$; otherwise, $P = AT_0^n$. In each case $P =$ 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 as follows.

Lemma 2.1. *Consider the problem of representing $y \in \mathbf{R}^d$ as a subconvex combination of the columns $(a_1, \ldots, a_n)$. 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 ref. 1 for more details. In discussing the (closed, convex) polytope P, we commonly refer to its vertices $v \in$ 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 $P \setminus \{v\}$, i.e., a value c so that $\lambda_v(v) = c$ and $\lambda_v(x) < c$ for $x \in P$, $x \neq c$. Thus $P =$ conv(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 vert(F) $\subset$ vert(P); thus if F is a face, $F =$ conv(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 of 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

$$\text{vert}(AT) \subset A \text{ vert}(T), \qquad \textbf{[2.1]}$$

which implies

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

and so the numbers of vertices obey

$$f_0(AT) \leq f_0(T). \qquad \textbf{[2.3]}$$

2.2. Basic Insights. *Theorem 1* involves two insights recorded here without proof. Similar lemmas were recently proven in ref. 6. 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*

(i) *x has a unique representation as a convex combination of vertices of P.*
(ii) *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 i and ii for every $x \in F$. Then F is a k-simplex and a k-face of P.

The second insight is that 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*

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

$$\textbf{[2.4]}$$

Conversely, suppose that Eq. 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 2.2*,

Donoho and Tanner

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 outwardly 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* that each y_i is a vertex of P, so P has at least n vertices. Now if/ε conv$\{a_j\}_{j=1}^n$, 0 is also a vertex of P. Since by Eq. **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 $i_1, \ldots, i_k$, $1 \leq i_\ell \leq n$. By hypothesis, for every x_0 of the form

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

with $\alpha_\ell \geq 0$ 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 of the x_0 under consideration populate a face F of T^{n-1}, determined by $i_1, \ldots, i_k$. 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 (ref. 3 and see chapter 7 in ref. 1).

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 $0 \leq t_1 < \cdots < t_n < \infty$, and let the jth column of the $d \times n$ matrix A be given by

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

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

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

The polytope obtained from the convex hull of the $(a_j)_{j=1}^n$ is $\lfloor d/2 \rfloor$ neighborly; see Gale (4). Note that A is a kind of nonsquare Vandermonde matrix.

- Trigonometric cyclic polytopes: Let $0 < t_1 < \cdots < t_n < 2\pi$, and, for $d = 2m$, let the jth 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

$$F(t) = (\cos(t), \sin(t), \cos(2t), \sin(2t), \ldots,$$

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

The polytope obtained from the convex hull of the $(a_j)_{j=1}^n$ is $\lfloor d/2 \rfloor$-neighborly, again see ref. 4. Note that A is a kind of nonsquare Fourier matrix.

Existing proofs of neighborliness of moment curve polytopes (1, 4), after a simple adaptation, give *Corollary 1.1*. Given a sequence (t_j) with $t_1 = 0$ the polytope conv$\{M(t_j)\}$ is $\lfloor d/2 \rfloor$-neighborly; since $M(0) = 0$, it follows that, for any sequence of (t_j), $P = \text{conv}(\{0\} \cup \{M(t_j)\}_{j=1}^n)$ is $\lfloor d/2 \rfloor$ neighborly. Hence $P = \text{conv}(\{0\} \cup \{M(t_j)\}_{j=1}^n)$ is outwardly neighborly. Hence, defining the matrix $A = [M(t_1), \ldots, M(t_n)]$, we get (LP)-(NP)-equivalence up to breakdown point $\lfloor d/2 \rfloor + 1$. *Corollary 1.1* follows.

Corollary 1.3 also follows from the outward-neighborliness of $P = \text{conv}(\{0\} \cup \{M(t_j)\})$. Let $y_k = \bar{\mu}_k^{(0)}$. Represent $\mu^{(0)}$ by a vector x_0 with n entries, the jth 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 first adapt the proof of the neighborliness of trigonometriccyclic polytopes to find that every polytope conv$(\{0\} \cup \{F(t_j)\})$ is outwardly $\lfloor d/2 \rfloor$-neighborly. Details are given in the appendix of ref. 15.

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 outwardly 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 theorem 3 in ref. 8; in fact, the underlying calculation in the proof of theorem 3 in ref. 8 can be seen to be the same as the "usual" one in proving neighborliness of trigonometric cyclic polytopes, although at the time of ref. 8 this connection was not known.) After this article was originally submitted, we learned about work by Jean-Jacques Fuchs (9) 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 $C: \mathbf{R} \mapsto \mathbf{R}^d$ (7, 16). 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 (17). Constructions based on oriented matroids and totally positive matrices have also been made by Sturmfels (18, 19). In the context of this article, we note that if such a curve passes through the origin, then, of course, conv$(\{0\} \cup \{C(t_j)\})$ is neighborly, and so outwardly neighborly as well. However, as the trigonometric moment curve shows, outward neighborliness is possible even when such a curve does not pass through the origin.

Sturmfels (18, 19) has 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 (3) proposed that in some sense most polytopes are neighborly. Goodman and Pollack proposed a natural model of random polytope in dimension d with n vertices (see ref. 13). They suggested taking the standard simplex T^{n-1} and apply a uniformly distributed random projection, getting the random polytope AT^{n-1}. Vershik and Sporyshev

(14) considered this question in the case where d and n increase to ∞ together in a proportional way. In ref. 12 we revisit the Vershik-Sporyshev model, asking about neighborliness of the resulting high-dimensional random polytopes. It proves the following.

Theorem 3.2. *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* $d \times n$ *orthogonal projection. There is* $\rho_N(\delta) > 0$ *so that, for* $\rho < \rho_N(\delta)$, *with overwhelming probability for large* n, 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 $\rho_N < 0.5$ 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 $\delta < 0.42$ can be inferred from the lack of $d/2$-weak neighborliness described below.)

The Goodman-Pollack model is broader than it first appears. By a result of Baryshnikov and Vitale (20), P is affinely equivalent to the convex hull of a Gaussian random sample. We can conclude the following.

Corollary 3.1. *Let* $A = A_{d,n}$ *denote a random* $d \times n$ *matrix with columns* a_j, $j = 1, \ldots, n$ *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, $\text{conv}\{a_j\}_{j=1}^n$ *is* $\lfloor \rho d \rfloor$-*neighborly.*

Ref. 12 implies that the preceding two results hold just as written also for $P = AT_0^n$, and $\text{conv}(\{0\} \cup \{a_j\})$, respectively. *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) that 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 $0 \notin \text{conv}\{a_j\}_{j=1}^n$, let $\mathcal{A} = \{a_j\}_0^n$. Otherwise let $\mathcal{A} = \{a_j\}_1^n$. We say 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_js (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 ref. 20 (compare ref. 6) proves the following.

Lemma 4.1. *Suppose that* $\mathcal{A}$ *is in general position. Then for* $k \leq d - 1$, *the* k-*dimensional faces of* $P = \text{conv}(\mathcal{A})$ *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 $k \leq d - 1$, 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 certain canonical unit basis vectors of $\mathbf{R}^n$. Hence, the collection of such convex combinations of vertices is in one-to-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 $k \leq d - 1$, 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 common support belong to the same face of T, and since each face as a whole survives or does not survive projection, we conclude the following.

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.*

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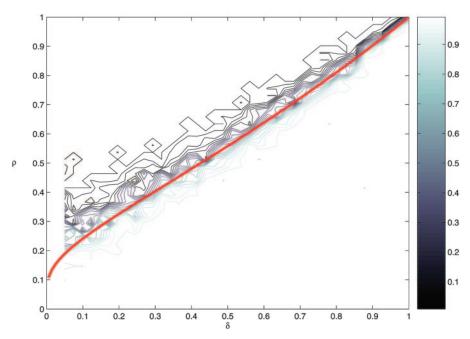

Fig. 1. Empirical verification of the NP/LP equivalence phase transition as a function of δ with $d_n = \lfloor \delta n \rfloor$ and sparsity $k = \lfloor \rho d_n \rfloor$ in the case of $n = 200$; weak neighborliness transition ρ_{VS}(red). The fraction of successes in (LP) recovering (NP) is in grayscale, and the calculated weak neighborliness transition curve $\rho_{VS}(\delta)$ is overlaid in red. Note that weak neighborliness exceeds $d/2$ for $\delta > 0.425$; see subsection 4.3.

Definition: Given a $d \times n$ matrix A, we say that a fraction $\geq (1 - \varepsilon)$ of all vectors x with k nonzeros are points of individual equivalence if individual equivalence holds for a fraction $\geq (1 - \varepsilon)$ 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. $\varepsilon(A, k)$ 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 $k < 0.558d$. See Fig. 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* by 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 the following.

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

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

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

- *A fraction $\geq (1 - \varepsilon)$ 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 Eq. **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.

Our work in ref. 12 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 $(1 + o_P(1))$, whereas for $\rho > \rho_{VS}(\delta)$ they differ by more than a factor $(1 + o_P(1))$. We 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 $\varepsilon > 0$, we have with overwhelming probability for large d that

$$P = AT \text{ is } (\tilde{\rho}_N \cdot d)\text{-outwardly neighborly, and}$$

$$(\tilde{\rho}_{VS} \cdot d, \ \varepsilon)\text{-weakly outwardly neighborly;}$$

here $\tilde{\rho}_N \equiv \rho_N(\delta) - \varepsilon$, and $\tilde{\rho}_{VS} \equiv \rho_{VS}(\delta) - \varepsilon$ 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 0.1.
- ρ_{VS} is substantially larger than ρ_N. The fact that it "crosses the line" $\rho = 1/2$ for δ near 0.425 is noteworthy; this means that whereas 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 $\delta \to 1$ (12, 14). For $\varepsilon > 0$ and δ sufficiently close to 1, for sufficiently large d, typical weak neighborliness can exceed $d(1 - \varepsilon)$! This is an important difference between neighborliness and weak neighborliness and is the source of *Corollary 1.5*.

For a discussion of further implications of these results and relationships to other work, see *Supporting Text* and also ref. 15.

1. Grünbaum, B. (2003) *Convex Polytopes*, Graduate Texts in Mathematics (Springer, New York), 2nd Ed., Vol. 221.
2. Ziegler, G. M. (1995) *Lectures on Polytopes*, Graduate Texts in Mathematics (Springer, New York), Vol. 152.
3. Gale, D. (1956) in *Linear Inequalities and Related Systems*, Annals of Mathematics Studies (Princeton Univ. Press, Princeton), No. 38, pp. 255–263.
4. Gale, D. (1963) *Proc. Sympos. Pure Math.* **VII**, 225–232.
5. McMullen, P. & Shephard, G. C. (1971) *Convex Polytopes and the Upper Bound Conjecture* (Cambridge Univ. Press, Cambridge, U.K.).
6. Donoho, D. L. (2005) *Neighborly Polytopes and the Sparse Solution of Underdetermined Systems of Linear Equations* (Department of Statistics, Stanford Univ., Stanford, CA), Technical Report 2005-04.
7. Motzkin, T. S. (1957) *Bull. Am. Math. Soc.* **63**, 35.
8. Donoho, D. L., Johnstone, I. M., Hoch, J. C. & Stern, A. S. (1992) *J. R. Stat. Soc. Ser. B* **54**, 41–81.
9. Fuchs, J.-J. (2005) in *Proceedings of the ICASSP, March 2005, Philadelphia, PA*, ed. Bystrom, M. (IEEE Signal Processing Society, Piscataway, NJ), pp. 729–732.
10. Carathéodory, C. (1907) *Math. Ann.* **64**, 95–115.
11. Carathéodory, C. (1911) *Rend. Circ. Mat. Palermo* **32**, 193–217.
12. Donoho, D. L. & Tanner, J. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 9452–9457.
13. Affentranger, F. & Schneider, R. (1992) *Discrete Comput. Geom.* **7**, 219–226.
14. Vershik, A. M. & Sporyshev, P. V. (1992) *Selecta Math. Soviet.* **11**, 181–201.
15. Donoho, D. L. & Tanner, J. (2005) *Sparse Nonnegative Solution of Underdetermined Linear Equations by Linear Programming* (Department of Statistics, Stanford Univ., Stanford, CA), Technical Report 2005-06.
16. Derry, D. (1956) *Canad. J. Math.* **8**, 383–388.
17. Karlin, S. & Shapley, L. S. (1953) *Geometry of Moment Spaces*, Memoirs of AMS (Am. Math. Soc., Providence, RI), Vol. 21.
18. Sturmfels, B. (1988) *Linear Algebra Appl.* **107**, 275–281.
19. Sturmfels, B. (1988) *Eur. J. Combin.* **9**, 537–546.
20. Baryshnikov, Y. M. & Vitale, R. A. (1994) *Discrete Comput. Geom.* **11**, 141–147.

APPLIED MATHEMATICS