Conditions for a Unique Non-negative Solution to an Underdetermined System

Meng Wang and Ao Tang
School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853

Abstract

This paper investigates conditions for an underdetermined linear system to have a unique nonnegative solution. A necessary condition is derived which requires the measurement matrix to have a row-span intersecting the positive orthant. For systems that satisfy this necessary condition, we provide equivalent characterizations for having a unique nonnegative solution. These conditions generalize existing ones to the cases where the measurement matrix may have different column sums. Focusing on binary measurement matrices especially ones that are adjacency matrices of expander graphs, we obtain an explicit threshold. Any nonnegative solution that is sparser than the threshold is the unique nonnegative solution. Compared with previous ones, this result is not only more general as it does not require constant degree condition, but also stronger as the threshold is larger even for cases with constant degree.

1. Introduction

Finding the sparest solution to a linear underdetermined system is in general a difficult problem. However, if the system satisfies certain conditions, then efficient recovery of the sparest solution is possible. Recently, there has been an explosion of research on this topic, see e.g., [1–5]. Formally, one wants to recover a n-dimensional signal x which is known apriori to be at most k-sparse from a m-dimensional ($m < n$) measurement $y = Ax$, where the m by n matrix A is referred to as the measurement matrix. [5] gives a sufficient condition known as Restricted Isometry Property (RIP) on A that guarantees the recovery of x via L_1 minimization which can be casted as a linear programming problem.

In many interesting cases, the vector x is known to be nonnegative. [6] gives a necessary and sufficient condition known as the outwardly neighborliness property of A for L_1 minimization to successfully recover a sparse non-negative solution. Moreover, recent studies [7–9] suggested that a sparse solution could be the unique non-negative solution there. This certainly leads to potentially better alternatives to L_1 minimization as in this case any optimization problem over this constraint set can recover the solution. To have a unique nonnegative solution, what is the requirement for the measurement matrix? How sparse that unique solution has to be? What is the relation between having a unique solution and successful recovery by L_1 minimization? Building on prior related literature, the first part of our paper (Section 3) discusses these questions.

Motivated by networking inference problems such as network tomography (see the example at the end of Section 3), we are particularly interested in systems where the measurement matrix is binary. There has not been many existing results on this type of systems except a few very recent papers [9–12]. A serious restriction in these papers is the requirement of the matrix to have constant column sum. In section 4 we make progress in this regard and our result allows different column sums.

We here summarize the main contribution of this paper. The paper focuses on characterizing the phenomenon that $\{x|Ax = Ax_0, x \geq 0, x_0 \geq 0\}$ is a singleton if x_0 is sparse enough. We demonstrate

- Different equivalent characterizations of the uniqueness property (Theorem 1),
- A necessary condition on matrix A such that the sparse solution is also the only solution (Theorem 2),
- Existence of a $(2m+1) \times n$ measurement matrix for any $n \geq 2m+2$ such that any m-sparse solution is the unique nonnegative solution (Theorem 3),
- Sparsity threshold for uniqueness for adjacency matrices of general expander graphs (Theorem 5).

2. Problem formulation and background

The goal of compressive sensing is to recover an n-dimensional vector x from a system of under-determined linear equations. $A^{m\times n} (m < n)$ is the measurement matrix, and $y = Ax$ is the m-dimensional measurement. In many applications, x is non-negative, which is our main focus here. In general, the task seems impossible as we have fewer measurements than variables. However, if x is known to be sparse, it can be recovered by solving the following problem,

$$\min \|x\|_0 \qquad \text{s.t.} Ax = y, x \geq 0, \tag{1}$$

where the L_0 norm $\|\cdot\|_0$ measures the number of nonzero entries of a given vector. Since (1) in general is NP-hard, people solve an alternative convex problem by replacing L_0 norm with L_1 norm where $\|x\|_1 = \sum_i |x_i|$. Note for a non-negative vector x, $\|x\|_1 = \mathbf{1}^T x$. [6] proves that if A is outwardly k-neighborly, then a k-sparse vector x can be recovered by solving the following L_1 minimization problem,

$$\min \mathbf{1}^T x \qquad \text{s.t.} Ax = y, x \geq 0. \tag{2}$$

In order to improve the recover ability, people also consider "weighted" L_1 minimization problem ([13] [9])

$$\min \beta^T x \qquad \text{s.t.} Ax = y, x \geq 0, \tag{3}$$

where the weight β is a positive vector.

In this paper we will show that for a certain class of matrices, if x is sufficiently sparse, not only can we recover x from (2) or (3), but also x is the only solution to $\{x|Ax = y, x \geq 0\}$. In other words, $\{x|Ax = y, x \geq 0\}$ is a singleton, and x can possibly be recovered by techniques other than L_1 minimization. Notice that if A has a column that is all 0, then the corresponding entry of x can never be recovered from y. Thus we assume A has no zero column throughout the paper.

3. Uniqueness with general measurement matrices

We focus on a class of matrices with a row-span intersecting the positive orthant as defined in [7].

Definition 1 ([7]). *A has a row-span intersecting the positive orthant, denoted by $A \in \mathbf{M}^+$, if there exists a strictly positive vector β in the row space of A, i.e. $\exists h$ such that*

$$h^T A = \beta^T > 0. \tag{4}$$

We now state a simple observation regarding matrices in $\mathbf{M}^+$.

Lemma 1. *Let $a_i \in \mathbf{R}^m$ $(i = 1, 2, ..., n)$ be the i^{th} column of a matrix A, then $A \in \mathbf{M}^+$ if and only if $0 \notin \textbf{\textit{Conv}}(a_1, a_2, ..., a_n)$, where*

$$\textbf{\textit{Conv}}(a_1, a_2, ..., a_n) = \{y = \sum_{i=1}^{n} \lambda_i a_i | \sum_{i=1}^{n} \lambda_i = 1, \lambda \geq 0\} \tag{5}$$

Proof. If $A \in \mathbf{M}^+$, then there exists h such that $h^T A = \beta^T > 0$. Suppose we also have $0 \in \mathbf{Conv}(a_1, a_2, ..., a_n)$, then there exists $\lambda \geq 0$ such that $A\lambda = 0$ and $\mathbf{1}^T \lambda = 1$. Then $(h^T A)\lambda = \beta^T \lambda > 0$ as $\beta > 0$, $\lambda \geq 0$ and $\lambda \neq 0$. But $(h^T A)\lambda = h^T (A\lambda) = 0$ as $A\lambda = 0$. Contradiction! Therefore $0 \notin \mathbf{Conv}(a_1, a_2, ..., a_n)$.

Conversely, if $0 \notin \mathbf{Conv}(a_1, a_2, ..., a_n)$, there exists a separating hyperplane $\{x|h^T x + b = 0, h \neq 0\}$ that strictly separates 0 and $\mathbf{Conv}(a_1, a_2, ..., a_n)$. We assume WLOG that $h^T 0 + b < 0$ and $h^T x + b > 0$ for any point x in $\mathbf{Conv}(a_1, a_2, ..., a_n)$. Then $h^T a_i > -b > 0, \forall i$. Thus we conclude $h^T A > 0$.

□

As first discovered in [7], for a matrix A in $\mathbf{M}^+$, if a non-negative vector x_0 is sparse enough, then $\{x|Ax = Ax_0, x \geq 0\}$ admits x_0 as the unique nonnegative solution. We will state two necessary and sufficient conditions, one in high-dimensional geometry and one in null space property, to characterize this phenomenon. To this end, we need another definition.

Suppose $A^{m \times n} \in \mathbf{M}^+$, then there exist $h \in \mathbf{R}^m$ such that $h^T A = \beta^T > 0$. Define a polytope P as the convex hull of vectors $(a_i/\beta_i, i = 1, 2, ..., n)$, i.e.

$$\begin{aligned} P &\triangleq \mathbf{Conv}(\frac{a_1}{\beta_1}, \frac{a_2}{\beta_2}, ..., \frac{a_n}{\beta_n}) \\ &= \{y = \sum_{i=1}^{n} \frac{\lambda_i}{\beta_i} a_i \in \mathbf{R^m} | \sum_{i=1}^{n} \lambda_i = 1, \lambda \geq 0\} \end{aligned} \tag{6}$$

Definition 2 ([6]). *We say a polytope P is k-neighborly if every set of k vertices spans a face F of P. F is a face of P if there exists a vector α_F such that $\alpha_F^T x = c, \forall x \in F$, and $\alpha_F^T x < c, \forall x \notin F$ and $x \in P$.*

We present the following theorem in the same style as in [6] and [9].

Theorem 1. *If $A^{m \times n} \in \mathbf{M}^+$, i.e. there exists $\beta > 0$ in the space spanned by rows of A, then the following three properties of A are equivalent:*

- *The polytope P defined in (6) has n vertices and is k-neighborly.*
- *For any non-negative vector x_0 with a support size no greater than k, the set $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton.*
- *For any $w \neq 0$ in the null space of A, both the positive support and the negative support of w have a size at least $k + 1$.*

Proof. We first show that statement 2 and statement 3 are equivalent for any matrix A. [9] (Theorem 3.3) states this equivalence for matrices with constant column sum. However, their proof does not require A to have constant column sum. We reformulate the proof as follows.

Forward direction. Suppose statement 3 fails. WLOG we assume there exists w in the null space of A with the size of negative support less than $k + 1$. We write w as

$$w = \begin{bmatrix} w_s \\ w_{s^c} \end{bmatrix}, \tag{7}$$

where the index set $s \subset \{1, 2, ..., n\}$ has size $|s| \leq k$, $w_s < 0$, and $w_{s^c} \geq 0$. Define

$$x_1 = \begin{bmatrix} -w_s \\ 0 \end{bmatrix} \geq 0, \quad x_2 = \begin{bmatrix} 0 \\ w_{s^c} \end{bmatrix} \geq 0. \tag{8}$$

Clearly $x_1 \neq x_2$, and x_1 is a non-negative vector whose positive support has a size no greater than k. But $\{x|Ax =$

$Ax_1, x \geq 0\}$ also contains x_2, thus not a singleton. Then statement 2 is not true.

Converse direction. Suppose statement 2 is not true. Then there exists a non-negative k-sparse vector x_0 such that $\{x|Ax = Ax_0, x \geq 0\}$ is not a singleton, i.e. there exists $\tilde{x} \geq 0$ such that $A\tilde{x} = Ax_0$ and $\tilde{x} \neq x_0$. We assume WLOG that

$$x_0 = \begin{bmatrix} x_s \\ 0 \end{bmatrix} \geq 0, \quad \tilde{x} = \begin{bmatrix} \tilde{x}_s \\ \tilde{x}_{s^c} \end{bmatrix} \geq 0, \tag{9}$$

where $x_s > 0$ and $|s| \leq k$. Let $w = \tilde{x} - x_0$, then

$$w = \begin{bmatrix} \tilde{x}_s - x_s \\ \tilde{x}_{s^c} \end{bmatrix}. \tag{10}$$

Since $\tilde{x}_{s^c} \geq 0$, the negative support of w has a size no greater than k, thus statement 3 fails.

We now show that statement 1 and statement 2 are equivalent. Define $B = diag(\beta)$ and let $D = AB^{-1}$. Then there is a one-to-one correspondence $z = Bx$ between the two sets

$$\{x|Ax = y, x \geq 0\} \text{ and } \{z|Dz = y, z \geq 0\}. \tag{11}$$

For any non-negative k sparse x_0, $z_0 = Bx_0$ is also non-negative and k-sparse. $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton if and only if $Q = \{z|Dz = Dz_0 = Ax_0, z \geq 0\}$ is a singleton. The polytope P defined in (6) is the convex hull of the column vectors of D.

First we show statement 1 implies statement 2. Since P is k-neighborly, it is easy to check that a polytope $\gamma P = \gamma \mathbf{Conv}(\frac{a_1}{\beta_1}, \frac{a_2}{\beta_2}, ..., \frac{a_n}{\beta_n})$ is also k-neighborly for any $\gamma > 0$. Note that $h^T x = 1$ holds for any point $x \in P$, then P lies in an n-dimensional hyperplane $\{x|h^T x = 1\}$. Then for any $\gamma_1, \gamma_2 > 0$ and $\gamma_1 \neq \gamma_2$, the polytopes $\gamma_1 P$ and $\gamma_2 P$ belong to two disjoint hyperplanes $\{x|h^T x = \gamma_1\}$ and $\{x|h^T x = \gamma_2\}$ respectively, and hence $\gamma_1 P$ and $\gamma_2 P$ are disjoint.

For any point $z \in Q$, we have $x = B^{-1}z$ belongs to $\{x|Ax = Ax_0, x \geq 0\}$, thus

$$\mathbf{1}^T z = \mathbf{1}^T Bx = \beta^T x = h^T Ax = h^T(Ax) = h^T(Ax_0) = c, \tag{12}$$

where the constant $c = h^T Ax_0 > 0$. Then for any $z \in Q$, we have $Dz \in cP$. Conversely, since $\gamma_1 P$ and $\gamma_2 P$ are disjoint for any positive $\gamma_1 \neq \gamma_2$, then for any $z \geq 0$ such that $Dz \in cP$, we must have $\mathbf{1}^T z = c$. Since z_0 is k-sparse, Dz_0 belongs to a k-face $F \in cP$. Since cP is k-neighborly, Dz_0 has a unique representation as a convex combination of vertices of cP, i.e. there is a unique λ such that $Dz_0 = \sum_{i=1}^n \lambda_i \frac{a_i}{w_i}, \mathbf{1}^T \lambda = c, \lambda \geq 0$. But z_0 is already such a representation. Therefore Q is a singleton, which implies $\{x|Ax = Ax_0, x \geq 0\}$ is also a singleton.

Then we show statement 2 implies statement 1. Since $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton, then Q is a singleton. Then the L_1 minimization problem $\min_Q \mathbf{1}^T z$ can recover z_0. From the Theorem 1 of [6] we know that the polytope $P' = \mathbf{Conv}(0, \frac{a_1}{\beta_1}, \frac{a_2}{\beta_2}, ..., \frac{a_n}{\beta_n})$ has $n+1$ vertices and is outwardly k-neighborly. Then P, the outward part of P', has n vertices and is k-neighborly. □

Note that assuming that $A \in \mathbf{M}^+$, there could exist more than one pair of (h, β) such that $h^T A = \beta^T > 0$, and Theorem 1 holds for any such pair. Therefore different polytopes defined from different β's have the same neighborliness property, and we can just check the neighborliness property of any one of these polytopes. For a binary matrix A, one simple choice is that $h = \mathbf{1}$, and consequently $\beta_i = \mathbf{1}^T a_i, i = 1, 2, ..., n$.

The next theorem states that the singleton phenomena is a property only for matrices in $\mathbf{M}^+$.

Theorem 2. *For any matrix $A \notin \mathbf{M}^+$ and for any non-negative vector x_0, $\{x|Ax = Ax_0, x \geq 0\}$ is never a singleton.*

Proof. Since $A \notin \mathbf{M}^+$, then from Lemma 1 we know $0 \in \mathbf{Conv}(a_1, a_2, ..., a_n)$. Then there exists a vector $w \geq 0$ such that $Aw = 0$ and $\mathbf{1}^T w = 1$. Clearly $w \in \mathbf{Null}(A)$ and $w \neq 0$. Then for any $\gamma > 0$ we have $A(x_0 + \gamma w) = Ax_0 + \gamma Aw = Ax_0$, and $x_0 + \gamma w \geq 0$ provided $x_0 \geq 0$. Hence $x_0 + \gamma w \in \{x|Ax = Ax_0, x \geq 0\}$. □

Remark: Theorem 2 shows that a necessary condition for the singleton phenomenon to happen is that A belongs to $\mathbf{M}^+$. If $A^{m \times n}$ is a random matrix such that every entry is independently sampled from Gaussian distribution with zero mean, then the probability that 0 lies in the convex hull of the column vectors of A, or equivalently $\{x|Ax = Ax_0, x \geq 0\}$ is never a singleton for any $x_0 \geq 0$, is $1 - 2^{-n+1} \sum_{k=0}^{m-1} \binom{n-1}{k}$([14]), which goes to 1 asymptotically as n increases if $\lim_{n \to +\infty} \frac{m-1}{n-1} < \frac{1}{2}$.

We next show that for any positive integers m and n ($m < n$), there exists a matrix $A^{m \times n}$ in $\mathbf{M}^+$ such that the uniqueness property holds for all non-negative $\lfloor \frac{m-1}{2} \rfloor$-sparse vectors. We state the theorem as follows.

Theorem 3. *For any $n, p \in \mathbf{Z}^+$ such that $n \geq 2p + 2$, there exists a matrix $A^{(2p+1) \times n} \in \mathbf{M}^+$ such that $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton for any non-negative k-sparse signal x_0 if $k \leq p$.*

Proof. We borrow the idea from [15] in the proof of the existence of a m-neighborly polytope with n vertices in $2m$ space for any $n > m$.

First we state without proof an important result from [15]. For any positive integers q and j, let S_q be the $(q+1)$-dimensional unit sphere. Then there exist $2j + q$ different points $b_1, ..., b_{2j+q}$ that are uniformly distributed on S_q. By uniform distribution we mean that any open hemisphere $H(\alpha) = \{x|x \in S_q, \text{and } \alpha^T x > 0\}$ contains at least j points in $b_1, ..., b_{2j+q}$.

In our problem, let $n = 2j + q$, $j = p + 1$, then $q = n - (2p + 2)$. From the previous result, there exist points $b_1, ..., b_n \in \mathbf{R}^{q+1}$ that are uniformly distributed on S_q. In other words, for any $\lambda \neq 0$ in $\mathbf{R}^{q+1}$, $\lambda^T b_i > 0$ for at least $p + 1$ vectors among $b_i, (i = 1, ..., n)$. And similarly $-\lambda^T b_i > 0$ for at least $p + 1$ vectors among $b_i, (i = 1, ..., n)$.

Let

$$G = \left[\begin{array}{cccc} b_1, b_2, \ldots, b_n \end{array} \right]^T. \tag{13}$$

Let $\mathbf{Range}(G)$ be the subspace generated by the columns of G. For any $w \neq 0$ in $\mathbf{Range}(G)$, there exists some $\lambda \neq 0$ such that

$$w = G\lambda = \left[\begin{array}{cccc} \lambda^T b_1, \lambda^T b_2, \ldots, \lambda^T b_n \end{array} \right]^T. \tag{14}$$

Then w has at least $p+1$ negative terms and at least $p+1$ positive terms.

If $\mathbf{Range}(G)$ is the null space of some matrix A, then from Theorem 1 we know $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton for any non-negative p-sparse x_0. To construct such A, take the orthogonal complement of $\mathbf{Range}(G)$ in $\mathbf{R}^n$, denoted by $(\mathbf{Range}(G))^\perp$. Since $\mathbf{Range}(G)$ has dimension $q+1$, then $(\mathbf{Range}(G))^\perp$ has dimension $n-(q+1) = 2p+1$. Pick a basis $h_1, h_2, ..., h_{2p+1} \in \mathbf{R}^n$ for $(\mathbf{Range}(G))^\perp$, and define

$$A = \left[\begin{array}{cccc} h_1, h_2, \ldots, h_{2p+1} \end{array} \right]^T. \tag{15}$$

Then A is the desired matrix.

Clearly $0 \notin \mathbf{Conv}(A)$, since otherwise there exists $w \geq 0$ such that $Aw = 0$ and $w \neq 0$, contradicting the fact that for any $w \neq 0$ in the null space of A should have a negative support with size at least $p+1$. Therefore $A \in \mathbf{M}^+$ from Lemma 1.

□

Conversely, for a given a measurement matrix $A^{m\times n}$, we can reverse the steps in the proof of Theorem 3 to find the threshold of sparsity such that the singleton property holds. To be specific, we choose a basis of its null space $\mathbf{Null}(A)$, say $s_1, s_2, ..., s_{n-m} \in \mathbf{R}^n$. Let $G = \left[\begin{array}{cccc} s_1 & s_2 & \ldots & s_{n-m} \end{array} \right]$, then take the row vectors of G and normalize them to unit norm. Let $b_1, b_2, ..., b_n \in \mathbf{R}^{n-m}$ denote the vectors after normalization. Since for any w in $\mathbf{Null}(A)$, there exists a λ in $\mathbf{R}^{n-m}$ such that $w = G\lambda$, then w_i has the same sign as $\lambda^T b_i$ for all $i = 1, ..., n$. Let K be the minimum number of nodes among $b_1, b_2, ..., b_n$ that are covered by a open hemisphere, i.e. $K = \min_\alpha(\sum_{i=1}^n \mathbf{1}_{\{b_i \in H(\alpha)\}})$, where $\mathbf{1}_{\{b_i \in H(\alpha)\}}$ is 1 if b_i belongs to $H(\alpha)$, and 0 otherwise. Then $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton for any non-negative x_0 with a support size less than K. However, it is in general hard to compute such K since that requires searching over all the open hemispheres.

The following proposition states that Theorem 3 is the "best" we can hope for in some sense.

Proposition 1. *Let $A^{m\times n}$ with its columns $a_i \in \mathbf{R^n}, i = 1, 2, ..., n$ in general position such that $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton for any non-negative signal x_0 that is at most p-sparse. If $m \leq 2p$, then $n \leq m$.*

Proof. First from Theorem 2 we know that A belongs to $\mathbf{M}^+$, i.e. there exists h such that $h^T A = \beta^T > 0$. We will prove our claim by contradiction. Suppose we have $n \geq m+1$, pick the first $m+1$ columns of A, i.e. $a_1, a_2, ..., a_{m+1}$. Then the equations

$$\sum_{i=1}^{m+1} \lambda_i \frac{a_i}{\beta_i} = 0 \tag{16}$$

have m equations and $m+1$ variables $\lambda_1, \lambda_2, ..., \lambda_{m+1}$, and have a non-zero solution. Taking the inner product of both sides of (16) with h, we have

$$\sum_{i=1}^{m+1} \lambda_i = 0. \tag{17}$$

Since $a_i \in \mathbf{R^n}, i = 1, 2, ..., n$ are in general position, none of λ_i is zero. Since A is in $\mathbf{M}^+$, from Lemma 1 we know $0 \notin \mathbf{Conv}(A)$, thus λ should have both positive and negative terms. Collecting positive and negative terms of λ separatively, we can rewrite (16) as follows,

$$\sum_{i\in I_p} \lambda_i \frac{a_i}{\beta_i} = \sum_{i\in I_n} \lambda_i \frac{a_i}{\beta_i}, \tag{18}$$

where I_p is the set of indices of positive terms and I_n is the set of indices of negative terms. We also have $\sum_{i\in I_p} \lambda_i = \sum_{i\in I_n} \lambda_i \triangleq r > 0$ from (17).

Since $|I_p| + |I_n| = m+1 \leq 2p+1$, we assume WLOG that $|I_p| \leq p$. From Theorem 1 we know is p-neighborly, i.e. for any index set I with size p, there exists α such that $\alpha^T \frac{a_i}{\beta_i} = c$ for any i in I, and $\alpha^T \frac{a_i}{\beta_i} < c$ for all i not in I. We consider specifically an index set I which contains I_p and its corresponding vector α. Taking the inner product of both sides of (18) with α, we would get rc on the left and some value strictly greater than rc on the right, giving a contradiction. □

As mentioned earlier, L_1 minimization is a widely-used technique to recover a sparse signal from its low-dimensional measurements. Interestingly, if A is in $\mathbf{M}^+$, which means the row space of A contains some positive vector β, then the success of weighted L_1 minimization using β as the weight is equivalent to the singleton property. To see this, we first state the following theorem.

Theorem 4. *Given a matrix A and any vector h, $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton if and only if x_0 is the unique solution to the following linear program,*

$$\min(h^T A)x \qquad s.t. Ax = Ax_0, x \geq 0. \tag{19}$$

Proof. For any point x in $\{x|Ax = Ax_0, x \geq 0\}$, $h^T Ax = h^T(Ax) = h^T(Ax_0)$ is a constant. Thus x is a solution to (19). Therefore $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton if and only if x_0 is the unique solution to (19). □

Remark: If $\{x|Ax = Ax_0, x \geq 0\}$ is not a singleton, then (19) has infinite number of solutions. That is because if x_1 and x_2 are two different solutions of (19), then $\lambda x_1 + (1-\lambda)x_2$ is also a solution for any $\lambda \in [0, 1]$.

Corollary 1. *For matrix $A \in \mathbf{M}^+$ with $h^T A = \beta^T > 0$, (3) with weight β admits a unique non-negative solution x_0 if and only if $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton.*

To see possible application of Corollary 1, we briefly introduce network inference problems here. Network inference problems are problems to extract individual parameters based on aggregate measurements in networks. For example, how to calculate loss rate/delay of each link based

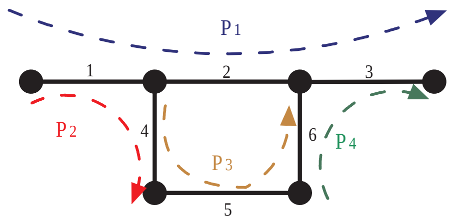


Figure 1. A network with six links and four paths

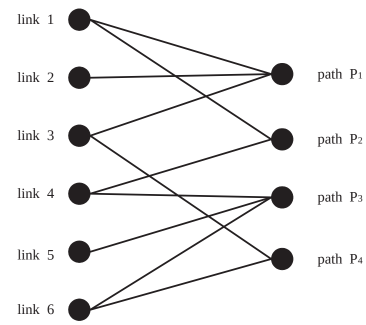


Figure 2. The bipartite graph corresponding to (20)

on end-to-end loss rate/delay measurements along certain paths? Another example is from traffic volume at each link and routing information, how to determine traffic between each source and destination pair. There has been active research in this area including a wide spectrum of approaches ranging from theoretical reasoning to empirical measurements. See [16–18] as a sample list.

These network inference problems are intrinsically "inverse problems" and underdetermined. One in general needs to add other conditions to make the problem mathematically solvable [19, 20]. However, if the target object is known to be sparse already[1], then the solution may be the unique non-negative solution.

In network inference problems, the measurement matrix A is typically a binary routing matrix with rows and columns indexed by the paths and links of a network. A_{ij} is 1 if link j belongs to path i, and 0 otherwise. Let's say we want to recover link queueing delays and use the vector x to denote them and it is known to be sparse. We hope to locate these bottle-neck links and quantify their delays via path delay measurements y. The delay of a path is the sum of delays of links it passes through. Take the network in Fig. 1 as an example. It contains six links and four measurement paths, and the routing matrix A is:

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{20}$$

From Corollary 1, one particular instance in network inference where the success of L_1 minimization and the singleton property of $\{x|Ax = Ax_0, x \geq 0\}$ is equivalent is that there exist a subset of paths such that they are disjoint from each other, and their union cover all the links. For example, consider the network in Fig. 1. Path P_1 and path P_3 are disjoint, and for any link in the network, it belongs to either P_1 or P_3. Mathematically, we have $[\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T A = [\begin{array}{cccc} 1 & 1 & 1 & 1 \end{array}]^T$, where A is the routing matrix in (20). Thus from Corollary 1, a non-negative vector x_0 is the unique solution to L_1 minimization if and only if $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton.

[1]This assumption can be valid in many cases. For example, if most links are not congested, then there is no congestion loss or queueing delay.

4. Uniqueness with expander measurement matrices

As mentioned earlier, in many problems such as ones on network inference, A is a binary matrix which is in $\mathbf{M}^+$. In this section, we specialize to this case.

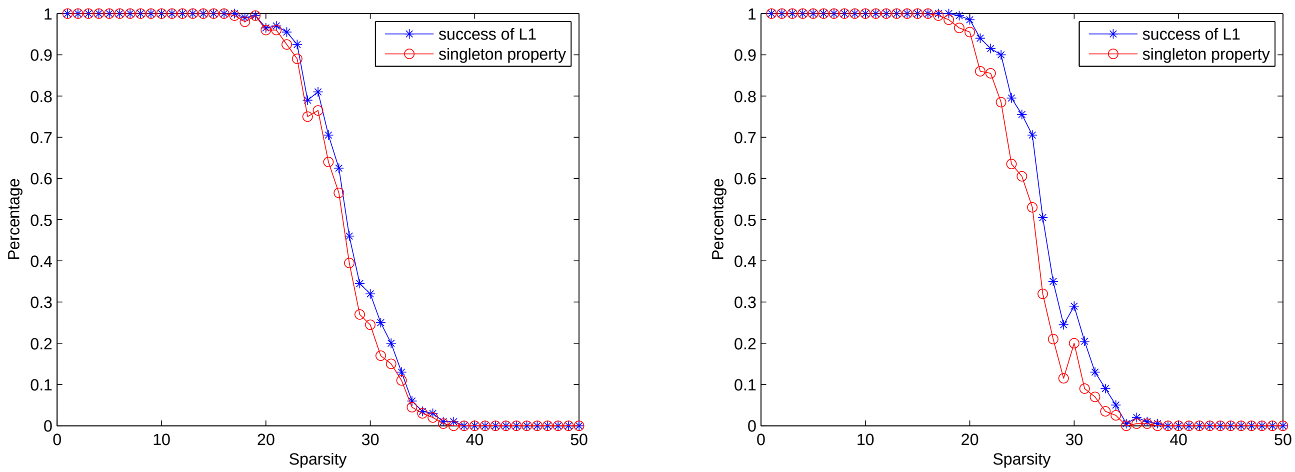
Theorem 1 gives two equivalent characterization of the singleton property of a matrix in $\mathbf{M}^+$. But as discussed after Theorem 3, it is hard to find the sparsity threshold of a given matrix A such that $\{x|Ax = Ax_0, x \geq 0\}$ is a singleton for all non-negative x_0 with a positive support below the threshold. However, we can give a simple sufficient condition of the singleton property if A is the adjacency matrix of an expander graph. An adjacency matrix of a graph is a binary matrix with $A_{ij} = 1$ if node i and node j are connected and zero otherwise. [9, 10, 12] studied related problems using expander graph with constant left degree. We instead employ a general definition of expander which does not require constant left degree.

Every $m \times n$ binary matrix A is the adjacency matrix of an unbalanced bipartite graph with n left nodes and m right nodes. There is an edge between right node i and left node j if and only if $A_{ij} = 1$. Let d_j denote the degree of left node j, and let d_l and d_u be the minimum and maximum of left degrees. Define $\rho = d_l/d_u$, then $0 < \rho \leq 1$. For example, the bipartite graph corresponding to the routing matrix in (20) is shown in Fig. 2. Here $d_l = 1$, $d_u = 2$, and $\rho = 0.5$.

Definition 3 ([21]). *A bipartite graph with n left nodes and m right nodes is an (α, δ) expander if for any set S of left nodes of size at most αn, the size of the set of its neighbors $\Gamma(S)$ satisfies $|\Gamma(S)| \geq \delta|E(S)|$, where $E(S)$ is the set of edges connected to nodes in S, and $\Gamma(S)$ is the set of right nodes connected to S.*

Our next main result is stated in the following theorem regarding the singleton property of an adjacency matrix of a general expander.

Theorem 5. *For an adjacency matrix A of an (α, δ) expander with left degrees in the range $[d_l, d_u]$, if $\delta\rho > \frac{\sqrt{5}-1}{2} \approx 0.618$, then for any non-negative k-sparse vector x_0 with $k \leq \frac{\alpha n}{1+\delta\rho}$, $\{Ax = Ax_0, x \geq 0\}$ is a singleton.*



(a) positive matrix (b) binary matrix

Figure 3. Comparison of L_1 recovery and singleton property

Proof. From Theorem 1, we need to prove that for any $w \neq 0$ such that $Aw = 0$, we have $|w_-| \geq \frac{\alpha n}{1+\delta\rho} + 1$ and $|w_+| \geq \frac{\alpha n}{1+\delta\rho} + 1$, where w_- and w_+ are negative support and positive support of w respectively.

We will prove by contradiction. Suppose WLOG that there exists $w \neq 0$ in $\mathbf{Null}(A)$ such that $|w_-| = k \leq \frac{\alpha n}{1+\delta\rho}$, then the set $E(w_-)$ of edges connected to nodes in w_- satisfies

$$d_l k \leq |E(w_-)| \leq d_u k. \tag{21}$$

Then the set $\Gamma(w_-)$ of neighbors of w_- satisfies

$$d_u k \geq |E(w_-)| \geq |\Gamma(w_-)| \geq \delta|E(w_-)| \geq \delta d_l k, \tag{22}$$

where the second to last equality comes from the expander property.

Notice that $\Gamma(w_-) = \Gamma(w_+) = \Gamma(w_- \cup w_+)$, otherwise $Aw = 0$ does not hold, then

$$|w_+| \geq \frac{|\Gamma(w_+)|}{d_u} = \frac{|\Gamma(w_-)|}{d_u} \geq \frac{\delta d_l k}{d_u} = \delta\rho k. \tag{23}$$

Now consider the set $w_- \cup w_+$, we have $|w_- \cup w_+| \geq (1+\delta\rho)k$. Pick an arbitrary subset $\tilde{w} \in w_- \cup w_+$ such that $|\tilde{w}| = (1+\delta\rho)k \leq \alpha n$. From expander property, we have

$$|\Gamma(\tilde{w})| \geq \delta|E(\tilde{w})| \geq \delta d_l|\tilde{w}| = \delta\rho(1+\delta\rho)d_u k > d_u k. \tag{24}$$

The last inequality holds since $\delta\rho(1+\delta\rho) > 1$ provided $\delta\rho > \frac{\sqrt{5}-1}{2}$. But $|\Gamma(\tilde{w})| \leq |\Gamma(w_- \cup w_+)| = |\Gamma(w_-)| \leq d_u k$. A contradiction arises, which completes the proof. □

Corollary 2. *For an adjacency matrix A of an (α, δ) expander with constant left degree d, if $\delta > \frac{\sqrt{5}-1}{2}$, then for any non-negative k-sparse vector x_0 with $k \leq \frac{\alpha n}{1+\delta}$, $\{Ax = Ax_0, x \geq 0\}$ is a singleton.*

Theorem 5 together with Corollary 2 is an extension to existing results. Theorem 3.5 of [9] shows that for an (α, δ) expander with constant left degree d, if $d\delta > 1$, then there exists a matrix $\tilde{A}$ (a perturbation of A) such that $\{\tilde{A}x = \tilde{A}x_0, x \geq 0\}$ is a singleton for nonnegative x_0 with sparsity up to $\delta\alpha n$. Our result instead can directly quantify the sparsity threshold needed for uniqueness for the original measurement matrix A. [12] discussed the success of L_1 recovery of a general vector x for expanders with constant left degree. If we apply Theorem 1 of [12] to cases where x is known to be non-negative, the result can be interpreted as that $\{Ax = Ax_0, x \geq 0\}$ is a singleton for any nonnegative x_0 with a sparsity up to $\frac{1}{2}\alpha n$ if $\delta > \frac{5}{6} \approx 0.833$. Corollary 2 implies that if $\delta > \frac{\sqrt{5}-1}{2} \approx 0.618$, the singleton property holds up to a sparsity of $\frac{1}{1+\delta}\alpha n$, which is larger than $\frac{1}{2}\alpha n$ for all $\delta < 1$.

5. Simulation

In this section, we generate a random $m \times n$ matrix A with $n = 2m = 100$ and empirically study the uniqueness property and the success of L_1 minimization for non-negative vectors with different sparsity. For a sparsity k, we select a support set S with size $|S| = k$ uniformly at random, and sample a non-negative vector x_0 on S with independent and identically distributed entries uniformly on the unit interval. Then we check whether $\{Ax = Ax_0, x \geq 0\}$ is singleton or not by solving (19) . For each instance, we also check whether L_1 minimization can recover x_0 from Ax_0 or not. Under a given sparsity k, we repeat the above experiment 200 times.

The results are shown in Fig. 3. In Fig. 3(a), A is a positive matrix with each entry sampled uniformly from the unit interval. In Fig. 3(b), A is a random binary matrix, and the sum of each column ranges from 2 to 6. We can see that if x_0 is sparse enough, it is the only solution to the constraint set. More interestingly, the thresholds where the singleton property breaks down and where the fully recovery of L_1

minimization breaks down are quite close.

6. Conclusion

This paper studies the phenomenon that $\{Ax = Ax_0, x \geq 0\}$ is a singleton if x_0 is sparse enough. We prove that this is a special property for matrices with a row span intersecting the positive orthant and give two necessary and sufficient conditions characterizing it. We show the existence of a $(2p+1) \times n$ matrix for any p and n satisfying $n \geq 2p+2$ such that its singleton property holds up to the sparsity of p. For the adjacency matrix of a general expander, we prove the singleton property holds for all k-sparse non-negative vectors where k is proportional to n.

There are several possible directions one can go along to further this study. The most intriguing one is to obtain uniqueness property threshold for a given measurement matrix. Another interesting question is to investigate whether the success of L_1 in this case is largely due to the unique solution as hinted by Fig. 3.

Acknowledgement

The authors would like to thank Dr. Weiyu Xu for helpful discussions.

References

[1] E. Candes, "Compressive sampling," in *Proc. International Congress of Mathematics*, 2006, pp. 1433–1452.

[2] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, July 2007.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[4] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[5] ——, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," in *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 27, 2005, pp. 9446–9451.

[7] A. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813–4820, Nov. 2008.

[8] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Preprint*, 2008.

[9] M. A. Khajehnejad, A. G. Dimakis, W. Xu, and B. Hassibi, "Sparse recovery of positive signals with minimal expansion," *Preprint*, 2009.

[10] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," in *Information Theory Workshop, 2007. ITW '07. IEEE*, Sept. 2007, pp. 414–419.

[11] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss., "Combining geometry and combinatorics: a unified approach to sparse signal recovery," *Preprint*, 2008.

[12] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices," *MIT-CSAIL Technical Report*, 2008.

[13] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *IEEE Trans. Inf. Theory*, vol. 14, no. 5-6, pp. 877–905, Dec. 2008.

[14] J. Wendel, "A problem in geometric probability," *Mathematica Scandinavica*, no. 11, pp. 109–111, 1962.

[15] D. Gale, "Neighboring vertices on a convex polyhedron," in *Linear Inequalities and Related Systems, Annals of Mathematics Studies*, no. 38. Princetion, NJ: Princeton Univ. Press, 1956, pp. 255–263.

[16] A. Coates, A. Hero III, R. Nowak, and B. Yu, "Internet tomography," *IEEE Signal Process. Mag.*, vol. 19, no. 3, pp. 47–65, May 2002.

[17] H. Nguyen and P. Thiran, "Using end-to-end data to infer lossy links in sensor networks," in *Proc. IEEE Infocom*, Apr. 2006, pp. 1–12.

[18] N. Duffied, "Network tomography of binary network performance characteristics," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5373–5388, Dec. 2006.

[19] Y. Zhang, M. Roughan, C. Lund, and D. Donoho, "An information-theoretic approach to traffic matrix estimation," in *Proc. ACM SIGCOMM*, Aug. 2003, pp. 301–312.

[20] H. X. Nguyen and P. Thiran, "Network loss inference with second order statistics of end-to-end flows," in *Proc. ACM SIGCOMM*, 2007, pp. 227–240.

[21] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Analysis of low density codes and improved designs using irregular graphs," in *Proc. ACM STOC*, 1998, pp. 249–258.