# New and Improved Conditions for Uniqueness of Sparsest Solutions of Underdetermined Linear Systems

Yun-Bin Zhao *

(First version in June 2012, Revised July 2013,
Published in *Applied Mathematics and Computation*, 224 (2013), pp.58-73.)

**Abstract.** The uniqueness of sparsest solutions of underdetermined linear systems plays a fundamental role in compressed sensing theory. Several new algebraic concepts, including the sub-mutual coherence, scaled mutual coherence, coherence rank, and sub-coherence rank, are introduced in this paper in order to develop new and improved sufficient conditions for the uniqueness of sparsest solutions. The coherence rank of a matrix with normalized columns is the maximum number of absolute entries in a row of its Gram matrix that are equal to the mutual coherence. The main result of this paper claims that when the coherence rank of a matrix is low, the mutual-coherence-based uniqueness conditions for the sparsest solution of a linear system can be improved. Furthermore, we prove that the Babel-function-based uniqueness can be also improved by the so-called sub-Babel function. Moreover, we show that the scaled-coherence-based uniqueness conditions can be developed, and that the right-hand-side vector $b$ of a linear system, the support overlap of solutions, and the range property of a transposed matrix can be also integrated into the criteria for the uniqueness of the sparsest solution of an underdetermined linear system.

**Key words.** Underdetermined linear system, sparsest solution, spark, coherence rank, mutual coherence, scaled mutual coherence, Babel function.

**AMS Subject Classifications:** 15A06, 94A12, 65K10, 15A29.

*School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom (y.zhao.2@bham.ac.uk). This work was partially supported by the Engineering and Physical Sciences Research Council (EPSRC) under the grant #EP/K00946X/1.

# 1 Introduction

Consider an underdetermined system of linear equations

$$Ax = b,$$

where $A$ is a given $m \times n$ matrix with $m < n$, and $b \in R^m$ is a given vector. Throughout this paper, we assume that $A$ has at least two rows, i.e., $m \geq 2$. Seeking for the sparsest solution of an underdetermined linear system has recently become an important and common request in many applications such as signal and image processing, compressed sensing, computer vision, statistical and financial model selections, and machine learning (see e.g., [5, 15, 26, 17] and the references therein). Let $\|x\|_0$ denote the cardinality, i.e., the number of nonzero components of the vector $x \in R^n$. Then finding a sparsest solution of a linear system is formulated as the so-called $\ell_0$-minimization problem

$$\min\{\|x\|_0 : \ Ax = b\},$$

which is known to be NP-hard [24, 1]. An intensive study of this problem has been carried out over the past few years (see e.g., [7, 11, 15, 26, 17]), and continues its growth in both theory and computational methods that stimulate further cross-disciplinary applications (see e.g., [6, 23, 25, 17]). However, the understanding of $\ell_0$-minimization problems, from theory to computational methods, remains very incomplete at the moment [5, 17]. For instance, the fundamental question of when an $\ell_0$-problem admits a unique solution has not yet addressed completely, and many existing uniqueness claims remain restrictive. The main purpose of this paper is to establish some new and improved sufficient conditions for a linear system to have a unique sparsest solution.

So far, sufficient criteria for the uniqueness of sparsest solutions have been developed by using such matrix properties as unique representation property [20], spark [12], mutual coherence [14], restricted isometry property (RIP)[8], null space property (NSP) [10, 30], exact recovery condition [27, 28], range space property (RSP) of $A^T$ [31, 32], and the verifiable conditions [22]. A crucial tool for the study of uniqueness is the spark, denoted by Spark($A$), which is the smallest number of columns of the matrix $A$ that are linearly dependent. The spark provides the guaranteed uniqueness of sparsest solutions, as shown by the result below.

**Theorem 1.1** ([12]). *If a linear system $Ax = b$ has a solution $x$ satisfying $\|x\|_0 < Spark(A)/2$, then $x$ is the unique sparsest solution of the system.*

The spark is difficult to compute. Any computable lower bound $0 < \phi(A) \leq \text{Spark}(A)$, however, produces a checkable sufficient condition for the uniqueness, such as $\|x\|_0 \leq \phi(A)/2$. The mutual coherence of a matrix (see the definition in section 2), denoted by $\mu(A)$, is such a property (e.g., [14, 16, 12, 21, 18, 27]) that yields a computable lower bound of the spark as follows

$$1 + \frac{1}{\mu(A)} \leq \text{Spark}(A), \tag{1}$$

which, together with Theorem 1.1, implies the following uniqueness claim.

**Theorem 1.2** ([14, 19, 16]). *If a linear system $Ax = b$ has a solution $x$ obeying*

$$\|x\|_0 < \left(1 + \frac{1}{\mu(A)}\right)/2, \tag{2}$$

*then $x$ is the unique sparsest solution of the system.*

The condition (2) is restrictive in many cases. In [27], the Babel function, denoted by $\mu_1(p)$, is introduced and shown to satisfy that $\text{Spark}(A) \geq \min\{p : \mu_1(p-1) \geq 1\} \geq 1 + 1/\mu$, yielding the following stronger uniqueness condition than (2).

**Theorem 1.3** ([27]). *If a linear system $Ax = b$ has a solution $x$ obeying*

$$\|x\|_0 < \frac{1}{2} \min\{p : \mu_1(p-1) \geq 1\}, \tag{3}$$

*then $x$ is the unique sparsest solution of the system.*

Theorems 1.2 and 1.3 are valid for general matrices. When $A = [\Phi \ \Psi]$ is a concatenation of two orthogonal matrices, Elad and Bruckstein [16] have shown that (2) can be improved to $\|x\|_0 < 1/\mu(A)$, and when $A$ consists of $J$ concatenated orthogonal bases, Gribonval and Nielsen [21] have shown that the uniqueness condition can be stated as $\|x\|_0 < \frac{1}{2}\left(1 + \frac{1}{J-1}\right)/\mu(A)$. For a general matrix $A$, however, it remains important, from a mathematical point of view, to address the question: *How can the bounds (2) and (3) be improved?* In this paper, we answer this question through the classic Brauder's Theorem. To this end, we introduce and use the sub-mutual coherence, which is the second largest inner product between two columns of a matrix with normalized columns, and the so-called coherence rank that turns out to be an important concept for the uniqueness of sparsest solutions. The sub-Babel function is also introduced in order to enhance the result of Theorem 1.3 above. One of our results in this paper claims that *for a general matrix $A$, when the coherence rank of $A$ is smaller than $1/\mu(A)$, the lower bound (1) of Spark(A), and thus the condition (2), can be improved.*

The spark of a matrix is invariant under nonsingular scalings, but the mutual coherence is not. Thus we introduce the concept of the scaled mutual coherence in section 4, which enables us to establish an optimal lower bound of the spark in certain sense. Note that the existing uniqueness conditions use matrix properties only, and the role of $b$ is completely overlooked. The sparsity of a solution, however, can also depend on the right-hand-side vector $b$ of a linear system. How to integrate $b$ into a uniqueness condition for sparsest solutions is worth addressing (as pointed out by Bruckstein et al. [5]). An instant application of the scaled mutual coherence yields a uniqueness condition that depends on the property of $A$ and $b$ altogether.

All the above-mentioned results are developed by identifying a lower bound for the spark of a matrix. Any improvement of the spark condition in Theorem 1.1 leads to a further enhancement of these results. Although it is hard to improve Theorem 1.1 in general, it is possible to do so in some situations. We show that the support overlap of solutions of a linear system is the information that can be used to achieve this goal (see section 5 for details). Finally, we introduce certain range properties of a matrix that can still guarantee the uniqueness of sparsest solutions to a linear system. Similar to the RIP [8, 7] and the NSP [10, 30], the range space property arises naturally from the analysis of the uniform recovery of sparse signals ([31, 32]).

This paper is organized as follows. We introduce several new concepts in section 2, and use them to develop improved uniqueness for sparsest solutions. The improvement of Babel-function-based uniqueness condition are given in section 3. The scaled-coherence-based uniqueness conditions and their applications are discussed in section 4. A further improvement of the spark condition via support overlap of solutions is demonstrated in section 5, and the range-property-based uniqueness is briefly introduced in section 6.

# 2 Improved conditions for uniqueness of sparsest solutions

Let $a_i, i = 1, ..., n$ be the columns of $A$. Recall that the mutual coherence of $A$ (see e.g., [14, 5]) is defined as

$$\mu(A) = \max_{i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \cdot \|a_j\|_2}.$$

So $\mu(A)$ is the maximum absolute value of the inner product between the normalized columns of $A$. The lower bound (1) plays a vital role in the development of the uniqueness theory and the performance guarantee of such algorithms as (orthogonal) matching pursuit, $\ell_1$-minimization, and iterative thresholding algorithms for the sparsest solution of linear systems (see e.g., [14, 16, 12, 18, 27, 28, 5, 15, 2, 3]). Any improvement of this lower bound may lead to an enhancement of many existing results in this field. In what follows, we develop an improved lower bound for spark (A) that leads to an improved sufficient conditions for a linear system to have a unique sparsest solution. Let us begin with a few concepts.

## 2.1 Sub-mutual coherence, coherence rank, and sub-coherence rank

Let us sort the different values of the inner product $|a_i^T a_j|/(\|a_i\|_2 \|a_j\|_2)$ in a descending order, and denote them by

$$\mu^{(1)}(A) > \mu^{(2)}(A) > \cdots > \mu^{(k)}(A).$$

Clearly, the largest one is $\mu^{(1)}(A) = \mu(A)$, the mutual coherence.

**Definition 2.1.** *The sub-mutual coherence of $A$, $\mu^{(2)}(A)$, is the second largest absolute inner product between two normalized columns of $A$ :*

$$\mu^{(2)}(A) = \max_{i \neq j} \left\{ \frac{a_i^T a_j}{\|a_i\|_2 \cdot \|a_j\|_2} : \frac{a_i^T a_j}{\|a_i\|_2 \cdot \|a_j\|_2} < \mu(A) \right\}.$$

In order to introduce the next useful property of a matrix, let us consider the index set

$$S_i(A) := \left\{ j : j \neq i, \frac{a_i^T a_j}{\|a_i\|_2 \cdot \|a_j\|_2} = \mu(A) \right\}, \ i = 1, ..., n.$$

Without loss of generality, we assume that the columns of $A$ are normalized. It is easy to see that $S_i(A)$ counts the number of absolute entries equal to $\mu(A)$ in $i$th row of $G = A^T A$, the Gram matrix of $A$. Clearly, at least one of these sets is nonempty, since the largest absolute entry of $G$ is equal to $\mu(A)$. Denote the cardinality of $S_i(A)$ by $\alpha_i(A)$, i.e.,

$$\alpha_i(A) = |S_i(A)|, \ i = 1, ..., n.$$

Clearly, $0 \leq \alpha_i(A) \leq n - 1$. Let

$$\alpha(A) = \max_{1 \leq i \leq n} \alpha_i(A) = \max_{1 \leq i \leq n} |S_i(A)|, \tag{4}$$

which is a positive number. Let $i_0$ be an index such that

$$\alpha(A) = \alpha_{i_0}(A) = |S_{i_0}(A)|,$$

i.e., the $i_0$th row of $G$ has the maximal number of absolute entries equal to $\mu(A)$. Then we may define

$$\beta(A) = \max_{1\le i\le n,\ i\ne i_0} \alpha_i(A) = \max_{1\le i\le n,\ i\ne i_0} |S_i(A)|, \tag{5}$$

which is the second largest number among $\alpha_i(A), i = 1, ..., n$.

**Definition 2.2.** $\alpha(A)$, *given by (4), is called the coherence rank of* $A$, *and* $\beta(A)$, *given by (5), is called sub-coherence rank of* $A$.

For a given matrix $A$ with normalized columns, both $\alpha(A)$ and $\beta(A)$ can be easily obtained through its Gram matrix $G = A^T A$ or its absolute Gram matrix, denoted by abs$(G)$. By the definition of $\mu(A)$, there exists at least one off-diagonal absolute entry of $A$, say $|G_{ij}|$ (in $i$th row), which is equal to $\mu(A)$. By the symmetry of $G$, we also have $|G_{ji}| = \mu(A)$ (in $j$th row of $G$). Thus the symmetry of $G$ implies that $\beta(A) \ge 1$. So, for any matrix $A$, we have the relation

$$1 \le \beta(A) \le \alpha(A). \tag{6}$$

Geometrically, $\alpha(A)$ can be called the *Equiangle* of $A$ in the sense that it is the maximum number of columns of $A$ that have the same largest angle with respect to a column, say the $i_0$-column, of $A$.

**Remark 2.3.** When all columns of $A$ are generated by a single vector, then $\mu(A) = 1$ and $\alpha(A) = \beta(A) = n - 1$. When $A$ has at least two independent columns (not all columns are generated by a single vector), then $\alpha(A) < n - 1$. For the concatenation of two orthogonal bases $A = [\Psi\ \Phi]$, where $\Psi, \Phi$ are $m \times m$ orthogonal matrices, we see that $\alpha(A) \le n/2 = m$. As we have pointed out, all $\mu(A), \mu^{(2)}(A), \alpha(A)$ and $\beta(A)$ can be obtained straightaway from the Gram matrix of $A$. For example, when $A$ is given by

$$A = \begin{bmatrix} -0.9802 & 0.1 & 0.3521 & 0.9239 & 0.9239 & 0.7405 \\ -1.8282 & 0 & 1.0365 & 0.3827 & -0.3827 & -1.6821 \\ 0.3269 & 0 & 1.3563 & 0 & 0 & -0.2949 \end{bmatrix}, \tag{7}$$

then the Gram matrix of the normalized $A$ is given by

$$G = \begin{bmatrix} 1 & -0.4668 & -0.4908 & -0.7644 & -0.0981 & 0.5763 \\ -0.4668 & 1 & 0.2020 & 0.9239 & 0.9239 & 0.3978 \\ -0.4908 & 0.2020 & 1 & 0.4142 & -0.0409 & -0.5803 \\ -0.7644 & 0.9239 & 0.4142 & 1 & 0.7071 & 0.0217 \\ -0.0981 & 0.9239 & -0.0409 & 0.7071 & 1 & 0.7134 \\ 0.5763 & 0.3978 & -0.5803 & 0.0217 & 0.7134 & 1 \end{bmatrix},$$

from which we see that $\mu(A) = 0.9239 > \mu^{(2)}(A) = 0.7644$, and $\alpha(A) = 2 > \beta(A) = 1$.

## 2.2 Coherence-rank-based lower bounds for Spark(A)

Let us first recall the Brauer's theorem [4] (see also Theorem 2.3 in [29]), concerning the estimate of eigenvalues of a matrix. Let $\sigma(A) := \{\lambda : \lambda \text{ is an eigenvalue of } A\}$ be the spectrum of $A$.

**Theorem 2.4** (Brauer [4]). *Let $A = (a_{ij})$ be an $N \times N$ matrix with $N \geq 2$. Then, if $\lambda$ is an eigenvalue of $A$, there is a pair $(r, q)$ of positive integers with $r \neq q$ $(1 \leq r, q \leq N)$ such that*

$$|\lambda - a_{rr}| \cdot |\lambda - a_{qq}| \leq \Delta_r \Delta_q, \text{ where } \Delta_i := \sum_{j=1, j\neq i}^{N} |a_{ij}| \text{ for } 1 \leq i \leq N.$$

*Hence if $K_{ij}(A) = \{z : |z - a_{ii}| \cdot |z - a_{jj}| \leq \Delta_i \Delta_j\}$ for $i \neq j$, then $\sigma(A) \subseteq \bigcup_{i\neq j}^{N} K_{ij}(A)$.*

We make use of this classic theorem to prove the following result, which turns out to be an improved version of (1) when the coherence rank is low.

**Theorem 2.5.** *Let $A \in R^{m\times n}$ be a matrix with $m < n$, and let $\alpha(A)$ and $\beta(A)$ be defined by (4) and (5), respectively. Suppose that one of the following conditions holds: (i) $\alpha(A) < \frac{1}{\mu(A)}$; (ii) $\alpha(A) \leq \frac{1}{\mu(A)}$ and $\beta(A) < \alpha(A)$. Then $\mu^{(2)}(A) > 0$ and*

$$\mathit{Spark}(A) \geq 1 + \frac{2\left[1 - \alpha(A)\beta(A)\widetilde{\mu}(A)^2\right]}{\mu^{(2)}(A)\left\{\widetilde{\mu}(A)(\alpha(A) + \beta(A)) + \sqrt{\left[\widetilde{\mu}(A)(\alpha(A) - \beta(A))\right]^2 + 4}\right\}}, \tag{8}$$

*where $\widetilde{\mu}(A) := \mu(A) - \mu^{(2)}(A)$.*

*Proof.* Normalizing the columns of a matrix does not affect any of the Spark$(A)$, $\mu(A)$, $\mu^{(2)}(A)$, $\alpha(A)$ and $\beta(A)$. Thus, without loss of generality, we assume that all columns of $A$ have unit $\ell_2$-norms. Let $p = \text{Spark}(A)$. By the definition of spark, there exist $p$ columns of $A$ that are linearly dependent. Let $A_S$ be the submatrix consisting of these $p$ columns. Without lost of generality, we assume $A_S = (a_1, a_2, ..., a_p)$. Thus the $p \times p$ matrix $G_{SS} = A_S^T A_S$ is singular, since the columns of $A_S$ are linearly dependent. Note that all diagonal entries of $G_{SS}$ are equal to 1, and all off-diagonal absolute entries are less than or equal to $\mu(A)$. Under either condition (i) or (ii), we have $\alpha(A) \leq \frac{1}{\mu(A)}$. Hence it follows from (1) that

$$1 + \alpha(A) \leq 1 + \frac{1}{\mu(A)} \leq \text{Spark}(A).$$

So, $\alpha(A) \leq \text{Spark}(A) - 1 = p - 1$. Note that $G_{SS}$ is a $p \times p$ matrix. Thus in every row of $G_{SS}$, there exist at most $\alpha(A)$ absolute entries equal to $\mu(A)$, and the remaining $(p-1)-\alpha(A)$ absolute entries are less than or equal to $\mu^{(2)}(A)$. By the singularity of $G_{SS}$, $\lambda = 0$ is an eigenvalue of $G_{SS}$. Note that the entries of $G_{SS}$ are given by $G_{ij} = a_i^T a_j$ where $i, j = 1, ..., p$. Thus by Theorem 2.4, there exist two different rows, say $i$th and $j$th rows $(i \neq j)$, such that

$$|0 - G_{ii}| \cdot |0 - G_{jj}| \leq \Delta_i \Delta_j = \left(\sum_{k=1, k\neq i}^{p} |a_i^T a_k|\right)\left(\sum_{k=1, k\neq j}^{p} |a_j^T a_k|\right), \tag{9}$$

where $G_{ii} = G_{jj} = 1$ are two diagonal entries of $G_{SS}$. By the definition of $\alpha(A)$ and $\beta(A)$, one of these two rows contains at most $\alpha(A)$ entries with absolute values equal to $\mu(A)$, and the next row contains at most $\beta(A)$ entries with absolute values equal to $\mu(A)$. The remaining entries in these rows are less than or equal to $\mu^{(2)}(A)$. Therefore,

$$\begin{aligned}\left(\sum_{k=1, k\neq i}^{p} |a_i^T a_k|\right)\left(\sum_{k=1, k\neq j}^{p} |a_j^T a_k|\right) \quad \leq \quad & \left[\alpha(A)\mu(A) + (p - 1 - \alpha(A))\mu^{(2)}(A)\right] \cdot \Big[\beta(A)\mu(A) \\ & + (p - 1 - \beta(A))\mu^{(2)}(A)\Big].\end{aligned} \tag{10}$$

Combining (9) and (10) leads to

$$\begin{aligned} 1 &\leq \left[\alpha(A)\mu(A) + (p-1-\alpha(A))\mu^{(2)}(A)\right] \cdot \left[\beta(A)\mu(A) + (p-1-\beta(A))\mu^{(2)}(A)\right] \\ &= \left[\alpha(A)\widetilde{\mu}(A) + (p-1)\mu^{(2)}(A)\right] \cdot \left[\beta(A)\widetilde{\mu}(A) + (p-1)\mu^{(2)}(A)\right], \end{aligned}$$

where $\widetilde{\mu}(A) := \mu(A) - \mu^{(2)}(A)$. By rearranging terms, the inequality above can be written as

$$\left[(p-1)\mu^{(2)}(A)\right]^2 + \left[(p-1)\mu^{(2)}(A)\right](\alpha(A) + \beta(A))\widetilde{\mu}(A) + \alpha(A)\beta(A)\widetilde{\mu}(A)^2 - 1 \geq 0. \tag{11}$$

We now show that $\mu^{(2)}(A) \neq 0$. In fact, if $\mu^{(2)}(A) = 0$, then (11) is reduced to $\alpha(A)\beta(A)\mu(A)^2 \geq 1$, which contradicts both conditions (i) and (ii). In fact, each of conditions (i) and (ii) implies that $\alpha(A)\beta(A)\mu(A)^2 < 1$. Thus $\mu^{(2)}(A)$ is positive. Note that the quadratic equation (in $t$)

$$t^2 + t(\alpha(A) + \beta(A))\widetilde{\mu}(A) + \alpha(A)\beta(A)\widetilde{\mu}(A)^2 - 1 = 0$$

has only one positive root. So it follows from (11) that

$$\begin{aligned} &(p-1)\mu^{(2)}(A) \\ &\geq \frac{-(\alpha(A)+\beta(A))\widetilde{\mu}(A) + \sqrt{[\widetilde{\mu}(A)(\alpha(A)+\beta(A))]^2 - 4(\alpha(A)\beta(A)\widetilde{\mu}(A)^2 - 1)}}{2} \\ &= \frac{-(\alpha(A)+\beta(A))\widetilde{\mu}(A) + \sqrt{[(\alpha(A)-\beta(A))\widetilde{\mu}(A)]^2 + 4}}{2} \\ &= \frac{2\left[1 - \alpha(A)\beta(A)\widetilde{\mu}(A)^2\right]}{(\alpha(A)+\beta(A))\widetilde{\mu}(A) + \sqrt{[(\alpha(A)-\beta(A))\widetilde{\mu}(A)]^2 + 4}}, \end{aligned}$$

which is exactly the relation (8). $\square$

The next proposition shows that the bound (8) is an improved lower bound for the spark under the condition of Theorem 2.5.

**Proposition 2.6.** *Let $\Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right)$ denote the right-hand side of the inequality (8). When $\alpha(A) < \frac{1}{\mu(A)}$, we have*

$$\Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right) \geq \left(1 + \frac{1}{\mu(A)}\right) + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \alpha(A)\mu(A)).$$

*When $\alpha(A) \leq \frac{1}{\mu(A)}$ and $\beta(A) < \alpha(A)$, we have*

$$\begin{aligned} \Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right) \;\geq\; & \left(1 + \frac{1}{\mu(A)}\right) + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \alpha(A)\mu(A)) \\ & + \frac{\alpha(A)\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1 + \alpha(A)\widetilde{\mu}(A))}, \end{aligned}$$

where $\widetilde{\mu}(A) = \mu(A) - \mu^{(2)}(A)$.

*Proof.* By using the fact $\sqrt{a^2 + b^2} \leq a + b$ for any $a, b \geq 0$, we have

$$\Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right) - 1$$

$$
\begin{aligned}
&= \frac{2\left(1-\alpha(A)\beta(A)\widetilde{\mu}(A)^2\right)}{\mu^{(2)}(A)\left\{\widetilde{\mu}(A)(\alpha(A)+\beta(A))+\sqrt{[\widetilde{\mu}(A)(\alpha(A)-\beta(A))]^2+4}\right\}} \\
&\geq \frac{2\left(1-\alpha(A)\beta(A)\widetilde{\mu}(A)^2\right)}{\mu^{(2)}(A)\left\{\widetilde{\mu}(A)(\alpha(A)+\beta(A))+[\widetilde{\mu}(A)(\alpha(A)-\beta(A))]+2\right\}} \\
&= \frac{1-\alpha(A)\beta(A)\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1+\alpha(A)\widetilde{\mu}(A))}.
\end{aligned} \tag{12}
$$

Case 1: $\alpha(A) < \frac{1}{\mu(A)}$. In this case, by (6), i.e., $\beta(A) \leq \alpha(A)$, it follows from (12) that

$$
\begin{aligned}
\Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right) - 1 &\geq \frac{1-\alpha(A)^2\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1+\alpha(A)\widetilde{\mu}(A))} \\
&= \frac{1-\alpha(A)\widetilde{\mu}(A)}{\mu^{(2)}(A)} \\
&= \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A)).
\end{aligned}
$$

Case 2: $\alpha(A) \leq \frac{1}{\mu(A)}$ and $\beta(A) < \alpha(A)$. In this case, since $\beta(A) \leq \alpha(A) - 1$, it follows again from (12) that

$$
\begin{aligned}
\Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right) - 1 &\geq \frac{1-\alpha(A)(\alpha(A)-1)\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1+\alpha(A)\widetilde{\mu}(A))} \\
&= \frac{1-\alpha(A)\widetilde{\mu}(A)}{\mu^{(2)}(A)} + \frac{\alpha(A)\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1+\alpha(A)\widetilde{\mu}(A))} \\
&= \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A)) \\
&\quad + \frac{\alpha(A)\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1+\alpha(A)\widetilde{\mu}(A))},
\end{aligned}
$$

as desired. □

Under the first case above, we see that $\left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A)) > 0$, and under the second case, we have

$$
\left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A)) + \frac{\alpha(A)\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1+\alpha(A)\widetilde{\mu}(A))} > 0.
$$

Thus, under the condition of Theorem 2.5, we have $\Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right) > 1 + \frac{1}{\mu(A)}$. Therefore, the lower bound of spark given by (8) does improve the bound (1) when the coherence rank, $\alpha(A)$, is small. Proposition 2.6 also indicates explicitly how much this improvement can be made at least.

If the Gram matrix $G$ of the normalized $A$ has two rows containing $\alpha(A)$ entries with absolute values equal to $\mu(A)$, then $\alpha(A) = \beta(A)$, in which case the lower bound (8) can be simplified to

$$
\Psi\left(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A)\right) = \left(1 + \frac{1}{\mu(A)}\right) + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A)).
$$

Note that $G$ has at most one absolute entry equal to $\mu(A)$ in its every row if and only if $\alpha(A) = \beta(A) = 1$. In this special case, the condition $\alpha(A) < 1/\mu(A)$ holds trivially when $\mu(A) < 1$. Thus, the next corollary follows immediately from Theorem 2.5.

**Corollary 2.7.** *Let $A \in R^{m\times n}$ be a matrix with $m < n$. If $\mu(A) < 1$ and $\alpha(A) = 1$, then $\mu^{(2)}(A) > 0$, and*

$$Spark(A) \geq 1 + \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \mu(A)).$$

Although Corollary 2.7 deals with a special case from a mathematical point of view, many matrices satisfy the property $\alpha(A) = 1$ together with $\mu(A) < 1$. Numerical experiments show that when a matrix is randomly generated, the coherence rank of the matrix is most likely equal to 1. In fact, the case $\alpha(A) \geq 2$ arises only when $A$ has at least two columns, each of which has the same angle to a column of the matrix, and such an angle is the largest one between a pair of columns of $A$. This phenomenon indicates that the coherence rank of a matrix is usually low in practice, typically $\alpha(A) = 1$.

### 2.3 Uniqueness via coherence and coherence rank

Consider the class of matrices

$$\begin{aligned}\mathcal{M} &= \left\{A \in R^{m\times n}: \text{ either } \alpha(A) \leq \frac{1}{\mu(A)} \text{ and } \beta(A) < \alpha(A), \text{ or } \alpha(A) < \frac{1}{\mu(A)}\right\} \\ &= \mathcal{M}_1 \cup \mathcal{M}_2, \end{aligned} \tag{13}$$

where

$$\mathcal{M}_1 = \left\{A \in R^{m\times n}: \ \alpha(A) < \frac{1}{\mu(A)}\right\}, \ \mathcal{M}_2 = \left\{A \in R^{m\times n}: \ \alpha(A) \leq \frac{1}{\mu(A)} \text{ and } \beta(A) < \alpha(A)\right\}.$$

We now state the main uniqueness claim of this section.

**Theorem 2.8.** *Let $A \in \mathcal{M}$, defined by (13). If the system $Ax = b$ has a solution $x$ obeying*

$$\|x\|_0 < \frac{1}{2}\left[1 + \frac{2\left(1 - \alpha(A)\beta(A)\widetilde{\mu}(A)^2\right)}{\mu^{(2)}(A)\left\{\widetilde{\mu}(A)(\alpha(A) + \beta(A)) + \sqrt{[\widetilde{\mu}(A)(\alpha(A) - \beta(A))]^2 + 4}\right\}}\right], \tag{14}$$

*where $\widetilde{\mu}(A) := \mu(A) - \mu^{(2)}(A)$, then $x$ is the unique sparsest solution to the linear system.*

This result follows instantly from Theorems 2.5 and 1.1. As shown by Proposition 2.6, condition (14) has improved the well-known condition (2) when $A$ is in class $\mathcal{M}$. This improvement is achieved by using the sub-mutual coherence $\mu^{(2)}(A)$ together with (sub-)coherence rank, instead of $\mu(A)$ only. Note that $\alpha(A), \beta(A), \mu(A)$ and $\mu^{(2)}(A)$ can be obtained straightforward from the Gram matrix $G = A^T A$. Thus the bound (14) can be easily computed.

By Theorem 2.5 and Proposition 2.6, we obtain the next result.

**Theorem 2.9.** (i) *Let $A \in \mathcal{M}_1$, defined by (13). If the system $Ax = b$ has a solution $x$ obeying*

$$\|x\|_0 < \frac{1}{2}\left[1 + \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \alpha(A)\mu(A))\right], \tag{15}$$

*then $x$ is the unique sparsest solution of the linear system.*

(ii) *Let $A \in \mathcal{M}_2$, defined by (13). If the system $Ax = b$ has a solution $x$ obeying*

$$\|x\|_0 < \frac{1}{2}\left[1 + \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \alpha(A)\mu(A)) + \frac{\alpha(A)\widetilde{\mu}(A)^2}{\mu^{(2)}(A)(1 + \alpha(A)\widetilde{\mu}(A))}\right], \quad (16)$$

*then $x$ is the unique sparsest solution of the linear system.*

(iii) *Let $A$ be a matrix with $\mu(A) < 1$ and $\alpha(A) = 1$. Then the solution of $Ax = b$ satisfying*

$$\|x\|_0 < \frac{1}{2}\left[1 + \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \mu(A))\right] \quad (17)$$

*is the unique sparsest solution of the linear system.*

Result (iii) of the above theorem shows that for coherence-rank-1 matrices, the uniqueness criterion (2) can be always improved to (17). As we have pointed out, matrices (especially the randomly generated ones) are largely coherence-rank-1, unless the matrix is particularly designed.

**Example 2.10.** Consider a randomly generated $A$ below and the absolute Gram matrix of its column-normalized counterpart

$$A = \begin{bmatrix} 0.0010 & -0.7998 & -0.6002 & 0.0717 \\ 0.8001 & -0.3558 & 0.4798 & -0.1913 \\ 0.5999 & 0.4801 & -0.6398 & -0.6412 \end{bmatrix}, \ \mathrm{abs}(G) = \begin{bmatrix} 1 & 0.0025 & 0.0005 & 0.7989 \\ 0.0025 & 1 & 0.0022 & 0.4422 \\ 0.005 & 0.0022 & 1 & 0.4093 \\ 0.7989 & 0.4422 & 0.4093 & 1 \end{bmatrix}.$$

From $\mathrm{abs}(G)$, we see that $\alpha(A) = \beta(A) = 1$, $\mu(A) = 0.7989$, and $\mu^{(2)}(A) = 0.4422$. Note that $\mathrm{Spark}(A)/2 = 2$ for this example. The standard mutual bound (2) is $(1 + \frac{1}{\mu(A)})/2 = 1.1258$, which is improved to 1.2274 by (17).

# 3 Improvement of Babel-function-based uniqueness

Let $A \in R^{m\times n}$ be a matrix with normalized columns. Tropp [27] introduced the so-called Babel-function defined as

$$\mu_1(q) = \max_{\Lambda, |\Lambda| = q} \max_{j \notin \Lambda} \sum_{i\in\Lambda} |a_i^T a_j|$$

where $a_k, k = 1, ..., n$, are the columns of $A$, and $\Lambda$ is some subset of $\{1, ..., n\}$. By this function, the following lower bound for spark is obtained (see [27]):

$$\mathrm{Spark}(A) \geq \min_{1\leq q\leq n} \{q : \mu_1(q-1) \geq 1\}. \quad (18)$$

The Babel function can be equivalently defined/computed in terms of the Gram matrix $G = A^T A$. In fact, sorting every row of $\mathrm{abs}(G)$ in descending order yields the matrix $\widehat{G} = (\widehat{G}_{ij})$ with the first column equal to the vector of ones, consisting of the diagonal entries of $G$. Therefore, as pointed out in [15], the Babel function can be written as

$$\mu_1(q) = \max_{1\leq k\leq m} \sum_{j=2}^{q+1} |\widehat{G}_{kj}| = \sum_{j=2}^{q+1} |\widehat{G}_{k_0 j}|, \quad (19)$$

where $k_0$ denotes an index such that the above maximum is achieved. Since $\mu_1(q-1) \leq (q-1)\mu(A)$, it is evident that

$$\min_{1 \leq q \leq n} \{q : \mu_1(q-1) \geq 1\} \geq 1 + \frac{1}{\mu(A)}.$$

So the lower bond given by (18) is an enhanced version of (1). Some immediate questions arise: *Can we compare the lower bounds (18) and (8)? Can the lower bounds (18) and (8) be further improved?*

We first address the second question above, by showing that the Babel-function-based bound (18) can be further improved by using the so-called sub-Babel function. Again, Brauer's Theorem plays a fundamental role in deriving such an enhanced result. The sub-Babel function, denoted by $\mu_1^{(2)}(q)$, is defined as

$$\mu_1^{(2)}(q) = \max_{1 \leq k \leq m, k \neq k_0} \sum_{j=2}^{q+1} |\widehat{G}_{kj}|, \tag{20}$$

where $k_0$ is determined in (19). Clearly, we have

$$\mu_1^{(2)}(q) \leq \mu_1(q) \quad \text{for any } 1 \leq q \leq n-1. \tag{21}$$

We have the following improved version of (18).

**Theorem 3.1.** *For any matrix $A \in R^{m \times n}$, we have*

$$\mathit{Spark}(A) \geq \min_{1 \leq q \leq n} \left\{ q : \quad \mu_1(q-1) \cdot \mu_1^{(2)}(q-1) \geq 1 \right\}. \tag{22}$$

*Proof.* Let $p = \text{Spark}(A)$. Then there exist $p$ columns of $A$ that are linearly dependent. Without lost of generality, we assume $A_S = (a_1, a_2, ..., a_p)$ is the submatrix consisting of these $p$ columns. Since the columns of $A_S$ are linearly dependent and normalized, the $p \times p$ matrix $G_{SS} = A_S^T A_S$ is singular, and all diagonal entries of $G_{SS}$ are equal to 1. Thus by Theorem 2.4 (Brauer's Theorem), for any eigenvalue $\lambda$ of $G_{SS}$, there exist two different rows, say $i$th and $j$th rows ($i \neq j$), such that

$$|\lambda - G_{ii}| \cdot |\lambda - G_{jj}| \leq \Delta_i \Delta_j = \left( \sum_{k=1, k \neq i}^{p} |a_i^T a_k| \right) \left( \sum_{k=1, k \neq j}^{p} |a_j^T a_k| \right), \tag{23}$$

where $G_{ii} = G_{jj} = 1$ are two diagonal entries of $G_{SS}$. By the definition of Babel and sub-Babel functions, we see that

$$\max\{\Delta_i, \Delta_j\} \leq \mu_1(p-1), \quad \min\{\Delta_i, \Delta_j\} \leq \mu_1^{(2)}(p-1).$$

Thus it follows from (23) that

$$(\lambda - 1)^2 \leq \Delta_i \Delta_j = \max\{\Delta_i, \Delta_j\} \cdot \min\{\Delta_i, \Delta_j\} \leq \mu_1(p-1) \cdot \mu_1^{(2)}(p-1).$$

In particular, since $\lambda = 0$ is an eigenvalue of $G_{SS}$, we have

$$\mu_1(p-1) \cdot \mu_1^{(2)}(p-1) \geq 1. \tag{24}$$

So $p = \text{Spark}(A)$ implies that $p$ must satisfy (24). Therefore,

$$\text{Spark}(A) = p \geq \min_{1 \leq q \leq n} \left\{ q : \quad \mu_1(q-1) \cdot \mu_1^{(2)}(q-1) \geq 1 \right\},$$

as desired. $\quad \Box$.

The next proposition shows that the lower bound (22) is an improved version of (18).

**Proposition 3.2.** Denote by

$$q^* = \min_{1 \leq q \leq n} \left\{ q : \quad \mu_1(q-1) \cdot \mu_1^{(2)}(q-1) \geq 1 \right\}, \; \widehat{q} = \min_{1 \leq q \leq n} \left\{ q : \quad \mu_1(q-1) \geq 1 \right\}.$$

Then $q^* \geq \widehat{q}$. In particular, if $\mu_1^{(2)}(\widehat{q}-1) < \frac{1}{\mu_1(\widehat{q}-1)}$, then $q^* > \widehat{q}$.

*Proof.* By the definition of $q^*$, we see that $\mu_1(q^*-1) \cdot \mu_1^{(2)}(q^*-1) \geq 1$. This, together with (21), implies that $\mu_1(q^*-1) \geq 1$. Thus

$$q^* \geq \min_{1 \leq q \leq n} \{ q : \; \mu_1(q-1) \geq 1 \} = \widehat{q}.$$

We now further show that this inequality holds strictly when the value of the sub-Babel function are relatively small in the sense that $\mu_1^{(2)}(\widehat{q}-1) < \frac{1}{\mu_1(\widehat{q}-1)}$. In fact, under this condition, we have

$$\mu_1(\widehat{q}-1) \cdot \mu_1^{(2)}(\widehat{q}-1) < 1.$$

Note that both $\mu_1(q-1)$ and $\mu_1^{(2)}(q-1)$ are increasing functions in $q$. The inequality above shows that when $\mu_1(q-1) \cdot \mu_1^{(2)}(q-1) \geq 1$, we must have $q > \widehat{q}$. Therefore,

$$q^* = \min_{1 \leq i \leq n} \{ q : \; \mu_1(q-1) \cdot \mu_1^{(2)}(q-1) \geq 1 \} > \widehat{q},$$

which shows that (22) improves (18) for this case. $\quad \Box$.

The next proposition indicates that when the coherence rank of $A$ is relatively small, bound (22) is also an improved version of (8).

**Proposition 3.3.** Let $A \in R^{m \times n}$ be a given matrix. Let $q^*$ be defined as in Proposition 3.2. If $\alpha(A) < 1/\mu(A)$ and $\alpha(A) \leq q^* - 1$, then

$$q^* \geq 1 + \frac{2\left[1 - \alpha(A)\beta(A)\widetilde{\mu}(A)^2\right]}{\mu^{(2)}(A)\left\{ \widetilde{\mu}(A)(\alpha(A)+\beta(A)) + \sqrt{[\widetilde{\mu}(A)(\alpha(A)-\beta(A))]^2 + 4} \right\}},$$

where $\widetilde{\mu}(A) := \mu(A) - \mu^{(2)}(A)$.

*Proof.* Since $\alpha(A) \leq q^* - 1$, by the definition of $\alpha(A)$ and $\beta(A)$, it follows from (19) and (20) that

$$\mu_1(q^*-1) \leq \alpha(A)\mu(A) + (q^*-1-\alpha(A))\mu^{(2)}(A),$$
$$\mu_1^{(2)}(q^*-1) \leq \beta(A)\mu(A) + (q^*-1-\beta(A))\mu^{(2)}(A).$$

These relations, together with the definition of $q^*$, imply that

$$\begin{aligned} 1 &\leq \mu_1(q^*-1) \cdot \mu_1^{(2)}(q^*-1) \\ &\leq \left[\alpha(A)\mu(A) + (q^*-1-\alpha(A))\mu^{(2)}(A)\right] \cdot \left[\beta(A)\mu(A) + (q^*-1-\beta(A))\mu^{(2)}(A)\right]. \end{aligned}$$

Thus we obtain the same inequality as (11) with $p$ replaced by $q^*$. Following (11), and repeating the same proof therein, we deduce that

$$q^* \geq 1 + \frac{2\left[1 - \alpha(A)\beta(A)\widetilde{\mu}(A)^2\right]}{\mu^{(2)}(A)\left\{\widetilde{\mu}(A)(\alpha(A)+\beta(A)) + \sqrt{[\widetilde{\mu}(A)(\alpha(A)-\beta(A))]^2+4}\right\}},$$

where $\widetilde{\mu}(A) := \mu(A) - \mu^{(2)}(A)$. $\square$.

It is also worth briefly comparing the Babel-function-based bound (18) and those developed in section 2 of this paper. At a first glance, it seems that (18) is more sophisticated than those developed in section 2. However, two types of bounds are mutually independent in the sense that one cannot definitely dominate the other in general. For example, when $\alpha(A) \leq \widehat{q} - 1$ and $\alpha(A) < 1/\mu(A)$ where $\widehat{q}$ is defined in Proposition 3.2, we have

$$1 \leq \mu_1(\widehat{q}-1) = \max_{1\leq k\leq m} \sum_{j=2}^{\widehat{q}} |\widehat{G}_{kj}| \leq \alpha(A)\mu(A) + (\widehat{q} - 1 - \alpha(A))\mu^{(2)}(A).$$

Thus,

$$\widehat{q} \geq 1 + \frac{1-\alpha(A)\widetilde{\mu}(A)}{\mu^{(2)}(A)} = \left(1 + \frac{1}{\mu(A)}\right) + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \alpha(A)\mu(A)).$$

In this case, the Babel-function-based bound (18) is tighter than bound (15). However, when $\widehat{q} - 1 < \alpha(A)$, the relationship between the bounds (15) and (18) can be complicated. The bound (15) and the one in Theorem 2.8 might be tighter than (18). Indeed, let us assume that $\widehat{q} - 1 < \alpha(A) \leq p - 1$ where $p = \text{Spark}(A)$, and $\alpha(A) < 1/\mu(A)$. Then (15) indicates that

$$p = \left\lceil 1 + \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A))\right\rceil + t^*$$

for some integer $t^* \geq 0$. This can be written as

$$\widehat{q} = \left\lceil 1 + \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A))\right\rceil + t^* - (p - \widehat{q}).$$

If $t^* < p - \widehat{q}$, then the above inequality implies that

$$\widehat{q} \leq 1 + \frac{1}{\mu(A)} + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1-\alpha(A)\mu(A)).$$

By Proposition 2.6, the right-hand side of the above is dominated by $\Psi(\alpha(A), \beta(A), \mu(A), \mu^{(2)}(A))$. Therefore, as a lower bound of spark, (8) is tighter than (18) in this special case.

# 4 Scaled mutual coherence

The mutual coherence is an important concept for the development of the uniqueness of sparsest solutions, and it is also crucial for the performance guarantee and stability analysis for many sparsity-seeking algorithms, such as basis pursuit, orthogonal matching pursuit, and thresholding algorithms (see e.g., [9, 16, 12, 18, 27, 28, 13, 5, 15, 17]). So it is worth considering how this concept can be further enhanced in order to possibly provide an improved bound for the spark. In this section, we introduce the scaled mutual coherence, which may lead to an optimal coherence-based estimate of the spark in certain sense. In theory, the improved results established in previous sections can be either extended or further improved by choosing a suitable scaling matrix.

### 4.1 Uniqueness via the scaled mutual coherence

Note that Spark(A), where $A \in R^{m\times n}$ with $m < n$, is invariant under a nonsingular linear transformation in the sense that

$$\mathrm{Spark}(A) = \mathrm{Spark}(WA)$$

for any nonsingular matrix $W \in R^{m\times m}$. However, the mutual coherence $\mu(A)$ is not. That is,

$$\mu(A) \neq \mu(WA)$$

in general (see Examples 4.3 and 4.4 in this section). Thus the improved conditions (14) -(17) still have a room for a further improvement by using a suitable nonsingular scaling $W$. Motivated by this observation, we consider the weighted inner product between every pair of columns of a matrix, and define

$$\mu_W(A) = \max_{i\neq j} \frac{|(Wa_i)^T Wa_j|}{\|Wa_i\|_2 \cdot \|Wa_j\|_2} = \mu(WA).$$

Similarly, we define

$$\mu_W^{(2)}(A) = \max_{i\neq j} \left\{ \frac{|(Wa_i)^T Wa_j|}{\|Wa_i\|_2 \cdot \|Wa_j\|_2} : \frac{|(Wa_i)^T Wa_j|}{\|Wa_i\|_2 \cdot \|Wa_j\|_2} < \mu_W(A) \right\} = \mu^{(2)}(WA).$$

In this paper, $\mu_W(A)$ and $\mu_W^{(2)}(A)$ are referred to be as the scaled mutual coherence and the scaled sub-mutual coherence, respectively. It makes sense to introduce the next definition.

**Definition 4.1.** *Let*

$$\mu_*(A) := \min_W \{\mu_W(A) : \ W \in R^{m\times m} \ \textit{is nonsingular}\}.$$

*$\mu_*(A)$ is called the optimal scaled mutual coherence (OSMC) of $A$.*

By definition, we have $\mu_*(A) \leq \mu_W(A)$ for any nonsingular $W \in R^{m\times m}$ and any $A \in R^{m\times n}$. In particular, by setting $W = I$ (the identity matrix), we see that $\mu_*(A) \leq \mu(A)$ for any $A$. As shown by the next result, the OSMC provides a theoretical lower bound for the spark that is better than any other scaled-mutual-coherence-based bound.

**Theorem 4.2.** *For any $m \times n$ ($m < n$) matrix $A$ with nonzero columns, we have $\mu_*(A) > 0$, and*

$$1 + \frac{1}{\mu_W(A)} \leq 1 + \frac{1}{\mu_*(A)} \leq \textit{Spark}(A)$$

*for any nonsingular matrix $W \in R^{m\times m}$. Hence if the system $Ax = b$ has a solution satisfying*

$$\|x\|_0 \leq \left(1 + \frac{1}{\mu_*(A)}\right)/2,$$

*or more restrictively, if there is a nonsingular matrix $W$ such that $\|x\|_0 < (1 + 1/\mu_W(A))/2$, then $x$ is the unique sparsest solution to the linear system.*

*Proof.* Let $W$ be an arbitrary nonsingular matrix. We consider the scaled matrix $WA$. Let $D = \mathrm{diag}(1/\|Wa_1\|_2, ..., 1/\|Wa_n\|_2)$ where $a_i, i = 1, ..., n$ are the columns of $A$. Then $WAD$ is a matrix with normalized columns. Clearly, this normalization does not change the spark (and hence, spark$(WAD)$ = spark$(WA)$ = spark$(A)$.) We also note that $WAD(D^{-1}x) = Wb$ and $Ax = b$ have the same sparsity of solutions. So without loss of generality, we assume that all columns of $WA$ have unit $\ell_2$-norms. Let $p = \mathrm{Spark}(A)$. By definition, there exist $p$ columns of $A$ that are linearly dependent. Let $A_S$ consist of these $p$ columns. Then the matrix

$$G_{SS}^{(W)} := (WA_S)^T(WA_S) = A_S^T W^T W A_S$$

is a $p \times p$ singular matrix due to the linear dependence of columns of $A_S$. Since $WA$ is normalized, all diagonal entries of $G_{SS}^{(W)}$ are equal to 1, and off-diagonal entries are less than or equal to $\mu_W(A)$. By the singularity of $G_{SS}^{(W)}$, this matrix has a zero eigenvalue. Thus by Gerschgorin's theorem, there exists a row of the matrix, say the $i$th row, such that

$$1 \leq \sum_{j \neq i} |(G_{SS}^{(W)})_{ij}| \leq (p-1)\mu_W(A),$$

which implied that

$$\mu_W(A) \geq 1/(p-1) > 0.$$

Note that the inequality above holds for any nonsingular matrix $W \in R^{m\times m}$. Taking the minimum value of the left-hand side yields $\mu_*(A) \geq 1/(p-1) > 0$, and thus

$$p(= \mathrm{Spark}(A)) \geq 1 + \frac{1}{\mu_*(A)}.$$

The right-hand side of the inequality above is greater than or equal $1 + 1/\mu_W(A)$ for any nonsingular $W$, since $\mu_W(A) \leq \mu_*(A)$. The uniqueness of sparsest solutions of the linear system $Ax = b$ follows immediately from Theorem 1.1. $\square$.

With a scaling matrix $W$, we denote the scaled coherence rank and scaled sub-coherence rank by $\alpha_W(A) = \alpha(WA)$ and $\beta_W(A) = \beta(WA)$, respectively. Let us define a class of matrices as follows:

$$\begin{aligned} \widetilde{\mathcal{M}} &= \Big\{ A \in R^{m\times n} : \text{there is a nonsingular } W \in R^{m\times m} \text{ such that} \\ &\quad \text{either } \alpha_W(A) \leq \frac{1}{\mu_W(A)} \text{ and } \beta_W(A) < \alpha_W(A), \text{ or } \alpha_W(A) < \frac{1}{\mu_W(A)} \Big\}. \end{aligned} \tag{25}$$

By applying the same proof of Theorem 2.5 to the scaled matrix $WA$, the lower bound of spark, together with uniqueness conditions for sparsest solutions in section 2, can be stated in terms of $\mu_W(A), \mu_W^{(2)}(A), \alpha_W(A)$ and $\beta_W(A)$. We omit the statement of the counterparts of Theorems 2.8 and 2.9 for the scaled coherence and the scaled coherence rank.

The next example shows that $\mathcal{M} \subset \widetilde{\mathcal{M}}$, i.e., $\widetilde{\mathcal{M}}$ is strictly larger than $\mathcal{M}$. This shows that by a suitable scaling, the results in section 2 may be further improved. In fact, when $\alpha(A) \leq \frac{1}{\mu(A)}$ does not hold (in which case $A \notin \mathcal{M}$), the scaled matrix $WA$ may satisfy the condition $\alpha_W(A) < \frac{1}{\mu_W(A)}$ (so $WA \in \mathcal{M}$, and hence $A \in \widetilde{\mathcal{M}}$), as shown by the next example.

**Example 4.3.** Consider the matrix (7) given in Remark 2.3. For this matrix, $\mu(A) = 0.9239$, $\mu^{(2)}(A) = 0.7644$, $\alpha(A) = 2$, $\beta(A) = 1$, and the bound (2) is 1.0498. Note that this matrix does not belong to $\mathcal{M}$, since $\alpha(A) \not\leq 1/\mu(A)$. So (14)-(17) cannot apply to this matrix. Now, we randomly generate a scaling matrix as follows

$$W = \begin{bmatrix} -0.9415 & -0.5320 & -0.4838 \\ -0.1623 & 1.6821 & -0.7120 \\ -0.1461 & -0.8757 & -1.1742 \end{bmatrix}.$$

It is easy to verify that $\mu_W(A) = 0.8954$, $\mu_W^{(2)}(A) = 0.8302$, and $\alpha_W(A) = \beta_W(A) = 1$. In fact, after this scaling, the absolute Gram matrix of the normalized $WA$ is given by

$$\text{abs}(G^{(W)}) = \begin{bmatrix} 1.0000 & 0.3561 & 0.7138 & 0.8302 & 0.3978 & 0.8954 \\ 0.3561 & 1.0000 & 0.5753 & 0.8130 & 0.7126 & 0.0973 \\ 0.7138 & 0.5753 & 1.0000 & 0.8227 & 0.0177 & 0.4874 \\ 0.8302 & 0.8130 & 0.8227 & 1.0000 & 0.1707 & 0.4969 \\ 0.3978 & 0.7126 & 0.0177 & 0.1707 & 1.0000 & 0.7634 \\ 0.8954 & 0.0973 & 0.4874 & 0.4969 & 0.7634 & 1.0000 \end{bmatrix}.$$

Thus by this scaling, the original coherence rank $\alpha(A) = 2$ is down to $\alpha_W(A) = 1$. Note that the scaled bound $(1 + \frac{1}{\mu_W(A)})/2$ in Theorem 4.2 is 1.0584, improving the original unscaled bound (2). This example shows that while $A \notin \mathcal{M}$, we have $WA \in \mathcal{M}$, and hence $A \in \widetilde{\mathcal{M}}$.

From simulations, we observe that when the coherence rank of a matrix is high in the sense that $\alpha(A) \geq 2$, it is quite sensitive to a scaling $W$, which may immediately reduce $\alpha(A)$ to $\alpha_W(A) = 1$, as shown by the above example. When the coherence rank $\alpha(A) = 1$, it is insensitive to a scaling $W$, and it is highly likely that $\alpha_W(A)$ remains 1.

**Example 4.4.** Consider $A$ and the absolute Gram matrix abs$(G)$ of its normalized counterpart

$$A = \begin{bmatrix} 0.0010 & -0.7998 & -0.6002 & 1.4290 \\ 0.8001 & -0.3558 & 0.4798 & 1.2393 \\ 0.5999 & 0.4801 & -0.6398 & -0.6849 \end{bmatrix}, \ \text{abs}(G) = \begin{bmatrix} 1 & 0.0025 & 0.0005 & 0.2894 \\ 0.0025 & 1 & 0.0022 & 0.9523 \\ 0.005 & 0.0022 & 1 & 0.0870 \\ 0.2894 & 0.9523 & 0.0870 & 1 \end{bmatrix}.$$

For this example, $\alpha(A) = \beta(A) = 1$, $\mu(A) = 0.9523$, and $\mu^{(2)}(A) = 0.2894$. The standard bound (2) is $(1 + \frac{1}{\mu(A)})/2 = 1.025$, which is improved to 1.0824 by (17). We now use the scaling matrix

$$W = \begin{bmatrix} -0.2078 & 0.9393 & 0.1905 \\ -0.9381 & 0.5715 & 0.3268 \\ 0.6702 & 0.2228 & 0.7662 \end{bmatrix},$$

which is a randomly generated nonsingular matrix. This scaling matrix yields $\mu_W(A) = 0.8343$, $\mu_A^{(2)}(A) = 0.7272$, and $\alpha_W(A) = \beta_W(A) = 1$. The original bound (2) can be further improved by bound $(1 + \frac{1}{\mu_W(A)})/2 = 1.0993$, and the unscaled bound (17) can be improved to 1.1139 by using the scaled mutual coherence.

Note that if the OSMC is attainable, i.e., there exists a nonsingular $W^*$ such that $\mu_*(A) = \mu_{W^*}(A)$. Then Theorem 4.3 holds for the OSMC. However, the optimal scaling $W^*$ is difficult to obtain in general. Also, for a given linear system, which scaling matrix should be used in order to improve the uniqueness claims is not obvious in advance. However, the scaled coherence can be viewed as a unified method for developing other coherence-type conditions for the uniqueness of sparsest solutions. It is worth mentioning that the Babel function can be also generalized to the weighted case, and related uniqueness claims can be made as well.

### 4.2 Application

Note that the existing uniqueness claims for sparsest solutions of linear systems are general and hold true uniformly for all $b$. These claims are made largely by using the property of $A$ only, and the role of $b$, which is solution-dependent, has been overlooked. Clearly, the property of the sparsest solution is usually dependent on $A$ and $b$. So it is interesting to incorporate the information $b$ into a uniqueness criterion for sparsest solutions. The scaled mutual coherence can be used to achieve this goal. Indeed, let $\phi$ be a mapping from $R^m$ to $R^m_{++}$ (the positive orthant of $R^m$). Denote by $\Phi_u = \text{diag}(\phi(u))$, a nonsingular diagonal matrix with diagonal entries $\phi_i(u) > 0, i = 1, ..., m$. Setting $u = b$, we see that the system $Ax = b$ is equivalent to

$$(\Phi_b A)x = \Phi_b b. \tag{26}$$

For instance, we let $\phi(u)$ be separable, i.e., $\phi(u) = (\phi_1(u_1), ..., \phi_n(x_n))^T$, and we define

$$\phi_i(t) = \begin{cases} 1/t & \text{if } t \neq 0 \\ 1 & \text{otherwise.} \end{cases} \tag{27}$$

By this choice, we have $\Phi_b b = \text{diag}(\phi(b))b = |\text{sign}(b)|$. Note that $\text{Spark}(A) = \text{Spark}(\Phi_b A)$, and the sparsity of solutions of the scaled system (26) is exactly the same as that of $Ax = b$. However, as we have seen before, a scaling matrix may change the mutual coherence, and a suitable scaling may improve the mutual-coherence-based uniqueness claims for sparsest solutions of a linear system. Through a scaling matrix dependent on $b$, the contribution of $b$ to the uniqueness of sparsest solutions can be demonstrated by the next two corollaries.

**Corollary 4.5.** *If the system $Ax = b$, where $A \in R^{m \times n}$ with $m < n$, has a solution satisfying $\|x\|_0 < \left(1 + \frac{1}{\mu(\Phi_b A)}\right)/2$, then $x$ is the unique sparsest solution to the linear system.*

Applying to the scaled system (26), this corollary follows from Theorems 4.2 and 1.1 straightaway, and this result can be improved when the scaled coherence rank $\alpha(\Phi_b A)$ is relatively small, as indicated by the next result.

**Corollary 4.6.** *Let $A$ be an $m \times n$ matrix with $m < n$.*
(i) *Suppose that either $\alpha(\Phi_b A) \leq \frac{1}{\mu(\Phi_b A)}$ and $\beta(\Phi_b A) < \alpha(\Phi_b A)$ or $\alpha(\Phi_b(A)) < \frac{1}{\mu(\Phi_b A)}$. If the system $Ax = b$ has a solution $x$ obeying*

$$\|x\|_0 < \frac{1}{2}\left[1 + \frac{2\left(1 - \alpha(\Phi_b A)\beta(\Phi_b A)\widetilde{\mu}(\Phi_b A)^2\right)}{\mu^{(2)}(\Phi_b A)\left\{\widetilde{\mu}(\Phi_b A)(\alpha(\Phi_b A) + \beta(\Phi_b A)) + \sqrt{[\widetilde{\mu}(\Phi_b A)(\alpha(\Phi_b A) - \beta(\Phi_b A))]^2 + 4}\right\}}\right],$$

*where $\widetilde{\mu}(\Phi_b A) := \mu(\Phi_b A) - \mu^{(2)}(\Phi_b A)$, then $x$ is the unique sparsest solution to the linear system. In particular, the same conclusion holds if $x$ obeys*

$$\|x\|_0 < \frac{1}{2}\left[1 + \frac{1}{\mu(\Phi_b A)} + \left(\frac{1}{\mu^{(2)}(\Phi_b A)} - \frac{1}{\mu(\Phi_b A)}\right)(1 - \alpha(\Phi_b A)\mu(\Phi_b A))\right],$$

(ii) *If $\phi$ is chosen such that $\mu(\Phi_b A) < 1$ and $\alpha(\Phi_b A) = 1$, then the solution $x$ of $Ax = b$ satisfying*

$$\|x\|_0 < \frac{1}{2}\left[1 + \frac{1}{\mu(\Phi_b A)} + \left(\frac{1}{\mu^{(2)}(\Phi_b A)} - \frac{1}{\mu(\Phi_b A)}\right)(1 - \mu(\Phi_b A))\right] \tag{28}$$

*is the unique sparsest solution of the linear system.*

The next example shows that when $b$ is involved, the uniqueness claim for sparsest solutions can be improved in some situations.

**Example 4.7.** Consider the system $Ax = b$ where $A$ is a $3 \times 5$ matrix given by

$$A = \begin{bmatrix} 1 & -3 & -6 & 4 & -3 \\ 2 & 3 & -2 & -2 & 3 \\ 3 & -2 & 1 & 0 & 4 \end{bmatrix}, \text{abs}(G) = \begin{bmatrix} 1 & 0.1709 & 0.2922 & 0 & 0.6875 \\ 0.1709 & 1 & 0.3330 & 0.8581 & 0.3656 \\ 0.2922 & 0.3330 & 1 & 0.6984 & 0.4285 \\ 0 & 0.8581 & 0.6984 & 1 & 0.6903 \\ 0.6875 & 0.3656 & 0.4285 & 0.6903 & 1 \end{bmatrix},$$

where $\text{abs}(G)$ is the absolute Gram matrix of the normalized $A$. From $\text{abs}(G)$, we see that $\mu(A) = 0.8581$, $\mu^{(2)}(A) = 0.6984$, and $\alpha(A) = \beta(A) = 1$. Thus the standard bound (2) is 1.0827, which is improved to 1.1016 by (17). In order to see which $b$ can further improve these bounds, let us randomly generate a vector $b$, for instance, $b = (3.6159, -3.5189, 2.6954)^T$. Let $\phi$ be given by (27). Then the absolute Gram matrix of the scaled matrix $\Phi_b A$ with normalized columns is given by

$$\text{abs}(G(\Phi_b A)) = \begin{bmatrix} 1.0000 & 0.3180 & 0.1608 & 0.0107 & 0.7833 \\ 0.3180 & 1.0000 & 0.2454 & 0.8042 & 0.1178 \\ 0.1608 & 0.2454 & 1.0000 & 0.6784 & 0.4231 \\ 0.0107 & 0.8042 & 0.6784 & 1.0000 & 0.5928 \\ 0.7833 & 0.1178 & 0.4231 & 0.5928 & 1.0000 \end{bmatrix},$$

from which we see that after this $b$-involved scaling, the coherence has changed to $\mu(\Phi_b A) = 0.8042$ and $\mu^{(2)}(\Phi_b A) = 0.7833$, and the coherence rank remains unchanged. The scaled bound $(1 + \frac{1}{\mu(\Phi_b A)})/2 = 1.1217$ and the scaled bound (28) equal to 1.1250 both improve the unscaled bound (2) and (17).

## 5 A further improvement via support overlap

Many uniqueness conditions for sparsest solutions of a linear system were derived from Theorem 1.1 by using the lower bound of Spark$(A)$. In this section, we point out that Theorem 1.1 itself might be improved in some situations by the support overlap of solutions of a linear system, leading to an enhanced spark-type uniqueness condition. We use Supp$(x)$ to denote the support of $x$, i.e., Supp$(x) = \{i : x_i \neq 0\}$.

**Definition 5.1.** *The support overlap $S^*$ of the solution of $Ax = b$ is the index set*

$$S^* = \bigcap_{x \in \mathcal{Y}} \mathit{Supp}(x),$$

*where $\mathcal{Y} = \{x : Ax = b\}$, the solution set of the linear system.*

Clearly, $S^*$ might be empty if there is no common index for the support of solutions. However, when some columns of $A$ are crucial, and they must be used for the representation of $b$, the support overlap $S^*$ is nonempty for these cases.

**Theorem 5.2.** *Let $S^*$ be the support overlap of the solution of the system $Ax = b$. If the system has a solution $x$ satisfying*

$$\|x\|_0 < \frac{1}{2}(|S^*| + Spark(A)), \tag{29}$$

*then $x$ is the unique sparsest solution of the linear system.*

*Proof.* Let $x$ be a solution of the system $Ax = b$ satisfying (29). We now prove that it is the unique sparsest solution of the linear system. We assume the contrary that the linear system has a solution $y \neq x$ with $\|y\|_0 \leq \|x\|_0$. Since $A(y - x) = 0$, which implies that the columns $a_i, i \in \text{Supp}(y - x)$ of $A$ are linearly dependent, we have

$$\|y - x\|_0 = |\text{Supp}(y - x)| \geq \text{Spark}(A). \tag{30}$$

Note that for any $u, v \in R^n$, the value of $\|\text{diag}(u)v\|_0$ is the number of $i$'s such that $u_i v_i \neq 0$. So it is easy to see that

$$S^* = \min \left\{ \|\text{diag}(x)u\|_0 : x, u \in \mathcal{Y} \right\}.$$

Thus, for any $u, v \in R^n$, we have

$$\|u - v\|_0 \leq \|u\|_0 + \|v\|_0 - \|\text{diag}(u)v\|_0,$$

and hence

$$\begin{aligned} \|y - x\|_0 &\leq \|y\|_0 + \|x\|_0 - \|\text{diag}(x)y\|_0 \\ &\leq 2\|x\|_0 - \|\text{diag}(x)y\|_0 \\ &\leq 2\|x\|_0 - |S^*|, \end{aligned} \tag{31}$$

where the first inequality follows from $\|y\|_0 \leq \|x\|_0$ and the second inequality follows from the fact $\|\text{diag}(x)y\|_0 \geq |S^*|$ for any $x, y \in \mathcal{Y}$. It follows from (30) and (31) that $2\|x\|_0 - |S^*| \geq \text{Spark}(A)$, which contradicts with (29). Thus $x$ is the unique sparsest solution of the linear system. □

As a result, all previous mutual-coherence-type uniqueness criteria for sparsest solutions of a linear system can be further improved when the value of $|S^*|$ or its lower bound is available. Taking Theorem 2.9 (iii) as an example, we have the following result.

**Corollary 5.3.** *Let $A \in R^{m\times n}$, where $m < n$, be a matrix with $\mu(A) < 1$ and $\alpha(A) = 1$. Suppose that $|S^*| \geq \gamma^*$ where $\gamma^*$ is known. Then if the system $Ax = b$ has a solution $x$ satisfying*

$$\|x\|_0 < \frac{1}{2}\left[\gamma^* + \left(1 + \frac{1}{\mu(A)}\right) + \left(\frac{1}{\mu^{(2)}(A)} - \frac{1}{\mu(A)}\right)(1 - \mu(A))\right], \tag{32}$$

*$x$ is the unique sparsest solution of the linear system.*

When the support overlap $S^*$ is nonempty, we have $|S^*| \geq 1$. All the aforementioned mutual coherence type bounds for uniqueness of sparsest solutions can be further improved by at least 0.5. Such an improvement can be crucial, as shown by the next example.

**Example 5.4.** Consider the system $Ax = b$ where

$$A = \begin{bmatrix} -1 & 0 & -4 & 2 & 4 \\ 0 & -1 & -1 & 1 & 2 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

Clearly, the last two columns are linearly dependent. So Spark$(A) = 2$, and Theorem 1.1 cannot confirm the uniqueness of any sparsest solution. However, note that the third column of $A$ is vital and must be used to represent $b$. This means that $x_3 \neq 0$ for any solution of the linear system. So, $|S^*| \geq 1 = \gamma^*$. Note that the solution $x^* = (0, 0, 1/2, 0, 0)^T$ satisfies that

$$\|x\|_0 = 1 < 1.5 = (\gamma^* + \mathrm{Spark}(A))/2 \leq (|S^*| + \mathrm{Spark}(A))/2.$$

By Theorem 5.2, $x^*$ is the unique sparsest solution of the linear system. This example shows that by incorporating the support overlap $S^*$, the result of Theorem 1.1 can be remarkably improved when $S^* \neq \emptyset$.

## 6 Uniqueness via range property of $A^T$

The exact recovery of all $k$-sparse vectors in $R^n$ by a single matrix $A$ is called the uniform recovery. To uniformly recover sparse vectors, some matrix properties should be imposed on $A$. The restricted isometry property (RIP) [8] and null space property [10, 30] are two well-known conditions for the uniform recovery. Recently, the so-called range space property (RSP) of order $k$ was proposed in [31, 32], which can also characterize the uniform recovery. All uniform recovering conditions imply that the linear system $Ax = y := Ax^0$ has a unique sparsest solution. In fact, these conditions have more capability than just ensuring the uniqueness of sparsest solutions of a linear system. For instance, they also guarantee that a linear system has a unique least $\ell_1$-norm solution, leading to the strong equivalence between $\ell_0$- and $\ell_1$-minimization problems, which is fundamental for the development of compressed sensing theory. In this section, we briefly discuss and develop certain more relaxed range properties of $A^T$ that guarantee the uniqueness of sparsest solutions. Our first range property is defined as follows, which was first introduced in [33] for a convergence analysis of reweighted $\ell_1$-methods for the sparse solution of a linear system.

**Definition 6.1** (Range Property (I)). *Let $A$ be a full-rank $m \times n$ matrix with $m < n$. Let $B$ be an $(n - m) \times n$ matrix consisting of the basis of the null space of $A$. $B^T$ is said to satisfy a range space property (RSP) of order $k$ with a constant $\rho > 0$ if*

$$\|\xi_{\overline{J}}\|_1 \leq \rho \|\xi_J\|_1$$

*for all $\xi \in \mathcal{R}(B^T)$, the range space of $B^T$, where $J \subseteq \{1, ..., n\}$ with $|J| = k$ is the indices of $k$ smallest absolute components of $\xi$, and $\overline{J} = \{1, ..., n\} \backslash J$.*

Based on the above definition, we have the next result.

**Theorem 6.2.** *Let $A \in R^{m \times n}$ and $B \in R^{(n-m) \times n}$ be full-rank matrices satisfying $AB^T = 0$, where $m < n$. Suppose that $B^T$ has a RSP of order $(n - k)$. Then the solution $x$ of the system $Ax = b$ obeying $\|x\|_0 \leq k/2$ is the unique sparsest solution of the linear system.*

*Proof.* First, under the condition of the theorem, we have the following statement (see e.g., Proposition 3.6 in [33]): *$B^T$ has the RSP of order $(n - k)$ with a constant $\rho > 0$ if and only if $A$ has the NSP of order $k$ with the same constant $\tau = \rho$.* Therefore, by the definition of NSP of order $k$, we have $\|\eta_\Lambda\|_1 \leq \tau \|\eta_{\overline{\Lambda}}\|_1$ for all $\eta \in \mathcal{N}(A)$ and all $\Lambda \subseteq \{1, 2, ..., n\}$ with $|\Lambda| \leq k$, where $\overline{\Lambda} = \{i : i \notin \Lambda\}$. This implies that the solution $x$ with $\|x\|_0 \leq k/2$ must be unique. In fact, we note that two $(k/2)$-sparse solutions $x$ and $y$ satisfy $A(x-y)=0$, i.e., $x-y \in \mathcal{N}(A)$. Let $\Lambda = \mathrm{Supp}(x-y)$. Since $x-y$ is at most $k$-sparse, we have $|\Lambda| \leq k$. By the NSP of order $k$, we have

$$\|x-y\|_1 = \|(x-y)_\Lambda\|_1 \leq \tau \|(x-y)_{\overline{\Lambda}}\|_1 = 0,$$

which implies that $x=y$. Thus the $(k/2)$-sparse solution is the uniqueness sparsest solution of the linear system. $\Box$

The above theorem impose range property on the basis of the null space of $A$, instead of on $A$ itself. We now impose a range property on $A$ directly.

**Definition 6.3** (Range Property (II)). *There exists an integer $k$ such that for any disjoint subsets $\Lambda_1, \Lambda_2$ of $\{1, ..., n\}$ with $|\Lambda_1| + |\Lambda_2| = k$ and $|\Lambda_2| \leq 1$, the range space $\mathcal{R}(A^T)$ contains a vector $\eta$ satisfying $\eta_i = 1$ for all $i \in \Lambda_1$, $\eta_i = -1$ for all $i \in \Lambda_2$, and $|\eta_i| < 1$ for $i \notin \Lambda_1 \bigcup \Lambda_2$.*

The above definition is a relaxed version of the range property introduced in [31]. Under the above range property (II), we can prove the following result.

**Theorem 6.4.** *Suppose that $A \in R^{m \times n}$ with $m < n$ satisfies the range property (II). Then if the system $Ax = b$ has a solution satisfying $\|x\|_0 \leq k/2$, $x$ is the unique sparsest solution of the linear system.*

*Proof.* Under the range property (II), we first prove that any $k$ columns of $A$ are linearly independent. In fact, let $\Lambda = \{\gamma_1, ..., \gamma_k\}$ be an arbitrary subset of $\{1, ..., n\}$ with $|\Lambda| = k$. We now prove that the columns of $A_\Lambda$ are linearly independent. It is sufficient to show that $z_\Lambda = 0$ is the only solution to the system $A_\Lambda z_\Lambda = 0$. In fact, let us assume $A_\Lambda z_\Lambda = 0$. Then $z = (z_\Lambda, z_{\overline{\Lambda}} = 0) \in R^n$ is in $\mathcal{N}(A)$. Consider the disjoint sets $\Lambda_1 = \Lambda$, and $\Lambda_2 = \emptyset$. By the range property (II), there exists a vector $\eta \in \mathcal{R}(A^T)$ with $\eta_i = 1$ for all $i \in \Lambda_1 = \Lambda$. By the orthogonality of $\mathcal{N}(A)$ and $\mathcal{R}(A^T)$, we have

$$0 = z^T \eta = z_\Lambda^T \eta_\Lambda + z_{\overline{\Lambda}}^T \eta_{\overline{\Lambda}} = z_\Lambda^T \eta_\Lambda,$$

which is nothing but

$$z_{\gamma_1} + z_{\gamma_2} + \cdots + z_{\gamma_k} = 0. \tag{33}$$

Now we consider an arbitrary pair of disjoint sets:

$$\Lambda_1 = \Lambda \backslash \{\gamma_i\}, \quad \Lambda_2 = \{\gamma_i\},$$

which satisfy that $|\Lambda_1| + |\Lambda_2| = k$ and $|\Lambda_2| \leq 1$. By the range property (II), there exists an $\eta \in \mathcal{R}(A^T)$ with $\eta_{\gamma_j} = 1$ for every $j \neq i$ and $\eta_{\gamma_i} = -1$. Again, it follows from $z^T \eta = 0$ that

$$(z_{\gamma_1} + \cdots + z_{\gamma_{i-1}} + z_{\gamma_{i+1}} \cdots + z_{\gamma_k}) - z_{\gamma_i} = 0,$$

which holds for every $i$ with $1 \leq i \leq k$. Combining these relations and (33) implies that $z_{\gamma_i} = 0$ for all $i = 1, ..., k$, i.e., $z_\Lambda = 0$. So any $k$ columns of $A$ are linearly independent. This implies that $k < \mathrm{Spark}(A)$. The desired result follows immediately from Theorem 1.1. $\Box$.

## 7 Conclusions

Through such concepts as sub-mutual coherence, scaled mutual coherence, coherence rank, and sub-Babel function, we have developed several new and improved sufficient conditions for a linear system to have a unique sparsest solution. The key result established in this paper claims that when the coherence rank of a matrix is low, the mutual-coherence-based lower bound for the spark of a matrix can be improved. We have also demonstrated that the scaled mutual coherence, which yields a unified uniqueness claim, may further improve the unscaled coherence-based uniqueness conditions if a suitable scaling matrix is used. The scaled mutual coherence makes it possible to integrate the right-hand-side vector $b$ of a linear system into a uniqueness criterion for the sparsest solution of a linear system. Moreover, the support overlap of solutions and certain range property of a matrix also play an important role in the uniqueness of sparsest solutions of linear systems.

## References


[1] E. Amaldi and V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, *Theoret. Comput. Sci.*, 209(1998), 237-260.

[2] T. Blumensath and M. E. Davies, Iterative Thresholding for Sparse Approximations, *Journal of Fourier Analysis and Applications*, 14 (2008), pp. 629-654.

[3] T. Blumensath and M. E. Davies, Iterative Hard Thresholding for Compressed Sensing, *Applied and Computational Harmonic Analysis,* 27 (2009), pp. 265-274.

[4] A. Brauer, Limits for the characteristic roots of a matrix, *Duke Math. J.*, 13 (1946), 387-395.

[5] A. Bruckstein, D. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, 51 (2009), 34-81.

[6] A. Budillon, A. Evangelista, and G. Schirinzi, Three-dimensional SAR focusing from multipass signals using compressive sampling, *IEEE Trans. Geoscience and Remote Sensing,* 49(2011), 488-499.

[7] E. Candès, Compressive sampling, *International Congress of Mathematicians*, Vol. III, 2006, 1433-1452.

[8] E.Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), 4203-4215.

[9] S.S. Chen, D.L. Donoho, and M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, 20 (1998), 33-61.

[10] A. Cohen, W. Dahmen and R. Devore, Compressed sensing and best $k$-term aproximation, *J. Amer. Math. Soc.*, 22 (2009), 211-231.

[11] D. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, 52 (2006), 1289-1306.

[12] D. Donoho and M. Elad, Optimality sparse representation in general (non-orthogonal) dictionaries via $\ell_1$ minimization, *Proc. Natl. Acad. Sci.*, 100 (2003), 2197-2202.

[13] D. Donoho, M. Elad and V. Temlyakov, Stable recovery of sparse overcomplete representation in the presence of noise, *IEEE Trans. Inform. Theory*, 52 (2006), 6-18.

[14] D. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inform. Theory*, 47 (2001), 2845-2862.

[15] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, 2010.

[16] M. Elad and A. Bruckstein, A generalized uncertainty priciple and sparse representation in pairs of bases, *IEEE Trans. Inform. Theory*, 48 (2002), 2558-2567.

[17] Y.C. Eldar and G. Kutyniok, *Compressive Sensing: Theory and Applciations*, Cambridge University Press, 2012.

[18] A. Feuer and A. Nemirovski, On sparse representation in pairs of bases, *IEEE Trans. Inform. Theory*, 49 (2003), 1579-1581.

[19] J.J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Trans. Inform. Theory,* 50 (2004), pp.1341-1344.

[20] I. Gorodnitsky and B. Rao, Sparse signal reconstruction from limited data using FOCUSS: A reweighted minimum norm algorithm, *IEEE Trans. Signal Proc.*, 45(1997), 600-616.

[21] R. Gribonval and M. Nielsen, Sparse decompositions in unions of bases, *IEEE Trans. Inform. Theory,* 49 (2003), 3320-3325.

[22] A. Juditsky, F. Kilinc Karzan, and A. Nemirovski, Verifiable conditions of $\ell_1$-recovery for sparse signals with sign restrictions, *Math. Program.*, Ser. B, 127 (2011), 89-122.

[23] A. Majumdar and R.K. Ward, An algorithm for sparse MRI reconstruction by Schatten $p$-norm minimization, *Magnetic Resonance Imaging*, 29 (2011), 408-417.

[24] B. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.*, 24 (1995), 227-234.

[25] A. Shabani, R. Kosut, M. Sohseni, H. Rabitz, A. Broome, M. Almeida, A. Fedrizzi and A. White, Efficient measurement of quantum dynamics via compressive sensing, *Pysical Review Letters*, 106, 100401 (2011).

[26] J.-L. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, 2010.

[27] J. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory*, 50 (2004), 2231-2242.

[28] J.Tropp, Just relax: convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inform. Theory*, 52 (2006), 1030-1051.

[29] R.S. Varga, *Matrix Iterative Analysis,* Second Edition, Springer-Verlag, Berlin, Heidelberg, 2000.

[30] Y. Zhang, Theory of compressive sensing via $\ell_1$-mimimization: A non-RIP analysis and extensions, *JORC*, 1 (2013), 79-105.

[31] Y.B. Zhao, RSP-based analysis for sparsest and least $\ell_1$-norm solutions to underdetermined linear systems, Technical Report, University of Birmingham, 2012.

[32] Y.B. Zhao, Equivalence and strong equivalence between sparsest and least $\ell_1$-norm nonnegative solutions to linear systems and their applications, Technical Report, University of Birmingham, 2012.

[33] Y.B. Zhao and D. Li, Reweighted $\ell_1$-minimization for sparse solutions to underdetermined linear systems, *SIAM J. Optim.*, 22 (2012), 1065-1088.