*Bull. Aust. Math. Soc.* **96** (2017), 353–360
doi:10.1017/S000497271700034X

# TWO NEW LOWER BOUNDS FOR THE SPARK OF A MATRIX

HAIFENG LIU✉, JIHUA ZHU and JIGEN PENG

(Received 18 February 2017; accepted 7 March 2017; first published online 8 June 2017)

**Abstract**

The $l_0$-minimisation problem has attracted much attention in recent years with the development of compressive sensing. The spark of a matrix is an important measure that can determine whether a given sparse vector is the solution of an $l_0$-minimisation problem. However, its calculation involves a combinatorial search over all possible subsets of columns of the matrix, which is an NP-hard problem. We use Weyl's theorem to give two new lower bounds for the spark of a matrix. One is based on the mutual coherence and the other on the coherence function. Numerical examples are given to show that the new bounds can be significantly better than existing ones.

2010 *Mathematics subject classification*: primary 05A20; secondary 90C27.

*Keywords and phrases*: lower bound, spark, mutual coherence, coherence function, compressive sensing.

## 1. Introduction

Compressive sensing (CS) is an innovative idea in the fields of signal processing and information theory (see, for example, [1, 2, 7, 9, 12]). Compressive sensing suggests that one can recover certain signals from far fewer samples than traditional methods by means of the optimisation problem:

$$\min \|x\|_0 \quad \text{such that } Ax = b, \tag{1.1}$$

where $\|x\|_0$ represents the number of nonzero entries of a vector, $A \in C^{m\times n}$ $(m < n)$ is a full row rank matrix, $b \in C^m$ is a nonzero vector and $x \in C^n$ is the goal to be calculated; (1.1) is called the $l_0$-minimisation problem [6]. It is an NP-hard problem. Its solution requires a combinatorial search through all possible solutions, which is not feasible. Fortunately, the problem can be computationally tractable for some coefficient matrices with special properties. One such property involves the spark of a matrix [3], defined as follows.

Definition 1.1. Given a matrix $A \in C^{m\times n}$, the spark of $A$, denoted by spark($A$), is the smallest number of columns from $A$ that are linearly dependent.

H. F. Liu, J. H. Zhu and J. G. Peng are supported by the Natural Science Foundation of China (Grant Nos. 11401463, 61573273 and 11131006) and the Fundamental Research Funds for the Central Universities.

Recall that rank($A$) is the largest number of columns from $A$ that are linearly independent. While there is an apparent resemblance between the definitions of rank and spark, they are essentially different. For a given matrix $A$, computing rank($A$) is tractable, but computing spark($A$) involves a combinatorial search over all possible subsets of columns of $A$, which is an NP-hard problem [10].

From the definition, provided $A$ has no zero column, $2 \leq \text{spark}(A) \leq \text{rank}(A) + 1$. (The coefficient matrix in compressive sensing always meets the condition imposed here [3, 5].) The following theorem from [3] establishes the importance of spark($A$) by showing that it can be used to determine whether a sparse vector is the solution of an $l_0$-minimisation problem.

Theorem 1.2. *If a linear system $Ax = b$ has a solution $x$ satisfying $\|x\|_0 < \frac{1}{2}\text{spark}(A)$, then this solution is necessarily the sparsest one, that is, the solution of problem (1.1).*

The spark of a matrix has applications in other fields. For example, in coding theory, it can be used to calculate the minimum distance of a code. In psychometrics, it is termed the Kruskal rank and is employed in the context of studying the uniqueness of tensor decomposition [5].

From Theorem 1.2, it can be observed that the larger spark($A$) is, the bigger the signal space among which compressive sensing can guarantee an exact recovery. Thus, good lower bounds for the spark of a matrix are significant. In this paper, we give two new lower bounds for spark($A$). The new lower bounds are presented in Section 3 after reviewing some preliminaries. Numerical examples to show the performance of the new bounds are given in Section 4.

## 2. Preliminaries

In this section, we review two existing lower bounds for the spark of a matrix. Throughout, $[n]$ represents the set $\{1, 2, \ldots, n\}$, card($S$) denotes the cardinality of the set $S$, $a_k$ indicates the $k$th column of the matrix $A$ and $\lambda_i(A)$ is its $i$th largest eigenvalue.

**2.1. Lower bound for the spark of a matrix based on mutual coherence.** The mutual coherence of a matrix [4] gives a way of measuring the dependence between its columns.

Definition 2.1. The mutual coherence, $\mu(A)$, of a matrix $A \in C^{m\times n}$ is the largest absolute normalised inner product between different columns from $A$, that is,

$$\mu(A) = \max_{1\leq i\leq n} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}.$$

Clearly, the mutual coherence of a matrix $A$ satisfies $\mu(A) \leq 1$, by the Cauchy–Schwarz inequality. For the identity matrix, the mutual coherence is zero because different columns are orthogonal. For a nonzero $m \times n$ matrix with more columns than rows, that is, $n > m$, $\mu(A)$ is strictly positive. The next theorem gives a lower bound for the spark of a matrix in terms of the mutual coherence.

Theorem 2.2 [3, 5]. *For a given matrix $A \in C^{m\times n}$,*

$$\mathrm{spark}(A) \geq 1 + \frac{1}{\mu(A)}.$$

**2.2. Lower bound for the spark of a matrix based on the coherence function.** The coherence function of a matrix [6] can be viewed as a generalisation of the mutual coherence.

Definition 2.3. Let $A \in C^{m\times n}$ be a matrix with $l_2$-normalised columns, $a_1, a_2, \ldots, a_n$, that is, $\|a_j\|_2 = 1$ $(j = 1, 2, \ldots, n)$. Given $p > 0$, the $l_p$-coherence function $\mu_p$ of the matrix $A$ is defined for $s = 1, 2, \ldots, n-1$ by

$$\mu_p(s) = \max_{i\in[n]} \max\bigg\{\bigg(\sum_{j\in S} |(a_i, a_j)|^p\bigg)^{1/p} \,\bigg|\, S \subset [n], \mathrm{card}(S) = s, i \notin S\bigg\}.$$

It can be seen that for any $p > 0$, $\mu_p(1) = \mu(A)$. At first glance, computing the coherence function for large $s$ is exponential and prohibitive, but this is not true [5]. Calculating the Gram matrix $G = A^T A$, taking the absolute value of each entry and sorting each row in descending order yields a nonnegative matrix $H = (h_{ij})_{n\times n}$. The first entry of every row of $H$ is 1, corresponding to the main diagonal entry of $G$. Then

$$\mu_p(s) = \max_{1\leq i\leq n} \bigg(\sum_{j=2}^{s+1} h_{ij}^p\bigg)^{1/p}.$$

Thus, the computational cost of calculating the $l_p$-coherence function is mainly spent in obtaining the Gram matrix and sorting every row of a nonnegative matrix of order $n$. The $l_1$-coherence function is also called the Babel function. Tropp [11] used it to give sufficient conditions under which both orthogonal matching pursuit and basis pursuit can converge to the solution of an $l_0$-minimisation problem. The next theorem gives a lower bound for spark$(A)$ in terms of the $l_1$-coherence function.

Theorem 2.4 [5]. *If the matrix $A \in C^{m\times n}$ has $l_2$-normalised columns $a_1, a_2, \ldots, a_n$, that is, $\|a_j\|_2 = 1$ $(j = 1, 2, \ldots, n)$, then*

$$\mathrm{spark}(A) \geq \min_{1\leq p\leq n} \{p \mid \mu_1(p-1) \geq 1\}.$$

Remark 2.5. If $A \in C^{m\times n}$ has $l_2$-normalised columns, then $\mu_1(p-1) \leq (p-1)\mu(A)$ for $p = 2, \ldots, n$. Thus,

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

Consequently, the lower bound for the spark of a matrix in terms of the $l_1$-coherence function is better than the lower bound involving the mutual coherence.

## 3. Two new lower bounds for the spark of a matrix

### 3.1. A new lower bound for the spark of a matrix based on mutual coherence.

THEOREM 3.1. *For a given matrix* $A \in C^{m\times n}$,

$$\operatorname{spark}(A) \geq 1 + \frac{1}{\mu(A)^2}. \tag{3.1}$$

REMARK 3.2. Since $\mu(A) \leq 1$ for an arbitrary matrix $A$,

$$1 + \frac{1}{\mu(A)^2} \geq 1 + \frac{1}{\mu(A)}. \tag{3.2}$$

By (3.2), Theorem 3.1 gives a better lower bound for the spark than Theorem 2.2.

Our proof of Theorem 3.1 uses Weyl's theorem (Lemma 3.3) in matrix analysis.

LEMMA 3.3 [8, Corollary 4.3.15]. *Assume that* $M, N \in C^{n\times n}$ *are two Hermitian matrices. Let* $\lambda_1(M) \geq \lambda_2(M) \geq \cdots \geq \lambda_n(M)$ *and* $\lambda_1(N) \geq \lambda_2(N) \geq \cdots \geq \lambda_n(N)$ *be their eigenvalues arranged in nonincreasing order. Then*

$$\lambda_i(M) + \lambda_n(N) \leq \lambda_i(M+N) \leq \lambda_i(M) + \lambda_1(N).$$

LEMMA 3.4. *Let* $B = (b_{ij}) \in C^{n\times n}$ *be a Hermitian matrix. Then every disc*

$$D_k = \Bigg\{ z \in R \mid |z - b_{kk}| \leq \Bigg( \sum_{\substack{j=1 \\ j\neq k}}^{n} |b_{ij}|^2 \Bigg)^{1/2} \Bigg\}$$

*contains at least one eigenvalue of* $B$ *for* $k = 1, 2, \ldots, n$.

PROOF. Construct a Hermitian matrix $C = (c_{ij}) \in C^{n\times n}$ as follows. The entries of $C$ in the $k$th row and column are zeros except for the diagonal entry and the rest of $C$ is the same as the matrix $B$, that is,

$$C = \begin{pmatrix} b_{11} & \cdots & 0 & \cdots & b_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & b_{kk} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n1} & \cdots & 0 & \cdots & b_{nn} \end{pmatrix}.$$

Then $b_{kk}$ is an eigenvalue of $C$. Let $M = C, N = B - C$. By Lemma 3.3,

$$\lambda_n(B - C) \leq \lambda_i(B) - \lambda_i(C) \leq \lambda_1(B - C).$$

Thus, $B$ has at least one eigenvalue $\lambda(B)$ satisfying

$$\lambda_n(B - C) \leq \lambda(B) - b_{kk} \leq \lambda_1(B - C). \tag{3.3}$$

https://doi.org/10.1017/S000497271700034X Published online by Cambridge University Press

It can be calculated that

$$|\lambda I-(B-C)|=\begin{vmatrix}\lambda & \cdots & -b_{1k} & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ -b_{k1} & \cdots & \lambda & \cdots & -b_{kn}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & \cdots & -b_{nk} & \cdots & \lambda\end{vmatrix}$$
$$=\lambda^{n-2}\Big(\lambda^2-\sum_{\substack{j=1\\ j\neq k}}^{n} b_{ji}b_{ij}\Big)=\lambda^{n-2}\Big(\lambda^2-\sum_{\substack{j=1\\ j\neq k}}^{n}|b_{ij}|^2\Big).$$

Thus, the eigenvalues of $B-C$ are

$$\Big(\sum_{\substack{j=1\\ j\neq k}}^{n}|b_{ij}|^2\Big)^{1/2},\quad \underbrace{0,\ldots,0}_{n-2\text{ times}}\quad\text{and}\quad -\Big(\sum_{\substack{j=1\\ j\neq k}}^{n}|b_{ij}|^2\Big)^{1/2}.$$

From (3.3),

$$-\Big(\sum_{\substack{j=1\\ j\neq k}}^{n}|b_{ij}|^2\Big)^{1/2}\leq\lambda(B)-b_{kk}\leq\Big(\sum_{\substack{j=1\\ j\neq k}}^{n}|b_{ij}|^2\Big)^{1/2}.\qquad\square$$

Lemma 3.4 is similar to the Gershgorin disc theorem, but more precise since

$$\Big(\sum_{\substack{j=1\\ j\neq k}}^{n}|b_{ij}|^2\Big)^{1/2}\leq\sum_{\substack{j=1\\ j\neq k}}^{n}|b_{ij}|\quad\text{for } k=1,2,\ldots,n.$$

Proof of Theorem 3.1. Note that the spark and the mutual coherence of a matrix are invariant under column scaling. Without loss of generality, we can assume that $A$ is a matrix with $l_2$-normalised columns. In this case, the Gram matrix $G=A^TA$ satisfies

$$g_{ii}=1,\quad 1\leq i\leq n,$$
$$|g_{ij}|\leq\mu(A),\quad 1\leq i,j\leq n,\ i\neq j.$$

Consider an arbitrary $p\times p$ leading minor from $G=A^TA$. From Lemma 3.4, if this submatrix satisfies $\sum_{j=1,j\neq k}^{p}|g_{ij}|^2<1$ for $1\leq i\leq p$, the minor will be nonsingular. From the definition of mutual coherence, $\sum_{j=1,j\neq k}^{p}|g_{ij}|^2\leq(p-1)\mu(A)^2$. Thus, the condition $(p-1)\mu(A)^2<1$, that is, $p<1+1/\mu(A)^2$, implies the nonsingularity of every $p\times p$ leading minor. Thus, any $p$ columns from $A$ are linearly independent and (3.1) follows from the definition of the spark. $\square$

### 3.2. A new lower bound for the spark based on the coherence function.

Theorem 3.5. *If the matrix $A\in C^{m\times n}$ has $l_2$-normalised columns $a_1,a_2,\ldots,a_n$, that is, $\|a_j\|_2=1$ for $j=1,2,\ldots,n$, then*

$$\mathrm{spark}(A)\geq\min_{1\leq p\leq n}\{p\mid\mu_2(p-1)\geq 1\}.\tag{3.4}$$

https://doi.org/10.1017/S000497271700034X Published online by Cambridge University Press

PROOF. Assume that $\mu_2(p-1) < 1$. Then, in every $p \times p$ leading minor from $G = A^T A$, the quadratic sum of the absolute values of the off-diagonal entries in any row is less than one. By Lemma 3.4, $G$ has no zero eigenvalue, so it is nonsingular. Thus, any $p$ columns from $A$ are linearly independent. The inequality (3.4) follows from the definition of the spark. □

REMARK 3.6. From Definitions 2.1 and 2.3, $\mu_2(p-1) \le \sqrt{p-1}\,\mu(A)$. Suppose that $\sqrt{p-1}\,\mu(A) < 1$, that is, $p < 1 + \mu(A)^{-2}$. Then $\mu_2(p-1) < 1$. Thus,

$$\min_{1 \le p \le n} \{p \mid \mu_2(p-1) \ge 1\} \ge 1 + \frac{1}{\mu(A)^2}. \tag{3.5}$$

The inequality (3.5) shows that $\min_{1 \le p \le n}\{p \mid \mu_2(p-1) \ge 1\}$ is a better lower bound for the spark than $1 + \mu(A)^{-2}$. On the other hand, calculating $1 + \mu(A)^{-2}$ is simpler since it does not involve mass sorting.

THEOREM 3.7. *If the matrix $A \in C^{m \times n}$ has $l_2$-normalised columns $a_1, a_2, \ldots, a_n$, that is, $\|a_j\|_2 = 1$ for $j = 1, 2, \ldots, n$, then*

$$\min_{1 \le p \le n} \{p \mid \mu_2(p-1) \ge 1\} \ge \min_{1 \le p \le n} \{p \mid \mu_1(p-1) \ge 1\}. \tag{3.6}$$

PROOF. Assume that $t \in [n]$ satisfies

$$\max_{1 \le i \le n} \Big( \sum_{j=2}^{s+1} h_{ij}^2 \Big)^{1/2} = \Big( \sum_{j=2}^{s+1} h_{tj}^2 \Big)^{1/2}.$$

Then, for $s = 1, 2, \ldots, n-1$,

$$\mu_2(s) = \max_{1 \le i \le n} \Big( \sum_{j=2}^{s+1} h_{ij}^2 \Big)^{1/2} = \Big( \sum_{j=2}^{s+1} h_{tj}^2 \Big)^{1/2} \le \sum_{j=2}^{s+1} h_{tj} \le \max_{1 \le i \le n} \sum_{j=2}^{s+1} h_{ij} = \mu_1(s)$$

and this yields (3.6). □

Theorem 3.7 indicates that $\min_{1 \le p \le n}\{p \mid \mu_2(p-1) \ge 1\}$ is a better lower bound for the spark of a matrix than $\min_{1 \le p \le n}\{p \mid \mu_1(p-1) \ge 1\}$. The inequalities (3.2), (3.5) and (3.6) show that $\min_{1 \le p \le n}\{p \mid \mu_2(p-1) \ge 1\}$ is the best lower bound for the spark among the four lower bounds.

## 4. Numerical examples

In this section, two examples are given to show the performance of our proposed lower bounds.

EXAMPLE 4.1. Let $A$ be the matrix

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{pmatrix}.$$

https://doi.org/10.1017/S000497271700034X Published online by Cambridge University Press

TABLE 1. The four lower bounds for the spark of the matrix $A$ in Example 4.1.

| | $1+\frac{1}{\mu(A)}$ | $\min_{1\le p\le n}\{p \mid \mu_1(p-1)\ge 1\}$ | $1+\frac{1}{\mu(A)^2}$ | $\min_{1\le p\le n}\{p \mid \mu_2(p-1)\ge 1\}$ |
|---|---|---|---|---|
| Lower bound | 4 | 5 | 7 | 8 |

TABLE 2. The four lower bounds for the spark of the random partial Fourier matrix $A$ in Example 4.2.

| | $1+\frac{1}{\mu(A)}$ | $\min_{1\le p\le n}\{p \mid \mu_1(p-1)\ge 1\}$ | $1+\frac{1}{\mu(A)^2}$ | $\min_{1\le p\le n}\{p \mid \mu_2(p-1)\ge 1\}$ |
|---|---|---|---|---|
| Lower bound | 9 | 10 | 61 | 112 |

It can be calculated that $\mu(A) = 0.4286$. The four lower bounds for the spark are shown in Table 1.

Note that $1+\mu(A)^{-1} = 3.3332$ and $1+\mu(A)^{-2} = 6.4437$. Since spark$(A)$ is a positive integer, we obtain spark$(A) \ge 4$ and spark$(A) \ge 7$ from the respective inequalities spark$(A) \ge 1+\mu(A)^{-1}$ and spark$(A) \ge 1+\mu(A)^{-2}$. Even for this small matrix, the lower bound in Theorem 3.1 can be much better than the existing lower bounds.

EXAMPLE 4.2. Suppose that $A \in C^{200\times 600}$ is a random partial Fourier matrix, that is, the rows of $A$ are randomly selected from a Fourier matrix of order 600. It can be calculated that $\mu(A) = 0.1298$. The four lower bounds for the spark are shown in Table 2.

It can be seen from Table 2 that the lower bounds in Theorems 3.1 and 3.5 are much better than the existing ones for this class of examples.

## Acknowledgement

The authors would like to thank the anonymous referee for valuable comments and suggestions.

## References

[1] E. J. Candes and T. Tao, ‘Near-optimal signal recovery from random projections: universal encoding strategies?’, *IEEE Trans. Inform. Theory* **52** (2006), 5406–5425.
[2] D. L. Donoho, ‘Compressed sensing’, *IEEE Trans. Inform. Theory* **52** (2006), 1289–1306.
[3] D. L. Donoho and M. Elad, ‘Optimally sparse representation in general (nonorthogonal) dictionaries via $l^1$ minimization’, *Proc. Natl. Acad. Sci. USA* **100** (2003), 2197–2202.
[4] D. L. Donoho and X. M. Huo, ‘Uncertainty principles and ideal atomic decomposition’, *IEEE Trans. Inform. Theory* **47** (2001), 2845–2862.
[5] M. Elad, *Sparse and Redundant Representations* (Springer, New York, 2010).
[6] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Springer, New York, 2013).
[7] R. Giryes, S. Nam, M. Elad, R. Gribonval and M. E. Davies, ‘Greedy-like algorithms for cosparse analysis model’, *Linear Algebra Appl.* **441** (2014), 22–60.
[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd edn (Cambridge University Press, Cambridge, 2013).

https://doi.org/10.1017/S000497271700034X Published online by Cambridge University Press

[9] H. Rauhut and C. Schwab, ‘Compressive sensing Petrov–Galerkin approximation of high-dimensional parametric operator equations’, *Math. Comp.* **86** (2017), 661–700.

[10] A. M. Tillmann and M. E. Pfetsch, ‘The computational complexity of the restricted isometry property, the nullspace property and related concepts in compressive sensing’, *IEEE Trans. Inform. Theory* **60** (2014), 1248–1259.

[11] J. A. Tropp, ‘Greed is good: algorithm results for sparse approximation’, *IEEE Trans. Inform. Theory* **50** (2004), 2231–2242.

[12] Y. H. Xiao and H. Zhu, ‘A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing’, *J. Math. Anal. Appl.* **405** (2013), 310–319.

HAIFENG LIU, School of Mathematics and Statistics,
Xi’an Jiaotong University, Xi’an 710049, China
e-mail: lhflkcc@126.com

JIHUA ZHU, School of Software Engineering,
Xi’an Jiaotong University, Xi’an 710049, China
e-mail: zhujh@mail.xjtu.edu.cn

JIGEN PENG, School of Mathematics and Statistics,
Xi’an Jiaotong University, Xi’an 710049, China
e-mail: jgpeng@mail.xjtu.edu.cn

https://doi.org/10.1017/S000497271700034X Published online by Cambridge University Press