Efficiency of $\ell_1$-minimization for $\ell_0$-minimization problems: Analysis via the Range Space Property

Yun-Bin Zhao

School of Mathematics, University of Birmingham
http://web.mat.bham.ac.uk/Y.Zhao
(E-mail: y.zhao.2@bham.ac.uk)

2013/2014
Outline

- $\ell_0$-minimization problem/the sparsest solution to a linear system
- $\ell_0$-minimization and Uniqueness of $\ell_1$-minimizers
- Efficiency analysis for $\ell_1$-minimization through a range space property (RSP)
- RSP-based theory for sparse signal recovery
- Sparsest optimal solution of linear programs
- Conclusions
**(l_0**)-minimization problem

\[
(l_0) \quad \min \{ \|x\|_0 : Ax = b \}
\]

where \( A \) is an \( m \times n \) matrix with \( m < n \).

- Solving this problem has become a common request in science, biology, and engineering (e.g., signal and image processing [compression, reconstruction, denoising, inpainting, separation and transmission], statistical model selection, compressed sensing, etc.)

- \( l_0 \)-problem is an NP-hard [Natarajan, 1995]. When is it computationally tractable, and how might it be solved? [Continuous approximation, heuristic method (orthogonal matching pursuit), thresholding-type methods, \( l_1 \)-method, and weighted \( l_1 \)-method].
$\ell_1$-minimization

- Replacing $\|x\|_0$ by $\|x\|_1$ yields the $\ell_1$-minimization:

$$\begin{align*}
(\ell_1) \quad & \text{Minimize} \quad \|x\|_1 \\
& \text{s.t.} \quad Ax = b,
\end{align*}$$

- $\ell_1$-norm

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

is the convex envelope of $\|x\|_0$ over the region

$$\{x : \|x\|_\infty \leq 1\}.$$ 

- **When are $\ell_0$- and $\ell_1$-problems equivalent?**
Recent $\ell_1$-norm related problems:

- Basis pursuit denoising:
  
  \[
  \text{Minimize } \lambda \|x\|_1 + \|Ax - y\|_2^2
  \]

- Quadratically constrained basis pursuit:
  
  \[
  \min \{ \|x\|_1 : \|Ax - y\|_2 \leq \varepsilon \}
  \]

- LASSO:
  
  \[
  \min \{ \|Ax - y\|_2 : \|x\|_1 \leq \tau \}
  \]

- The Dantzig selector:
  
  \[
  \min \{ \|x\|_1 : \|A^T(Ax - y)\|_\infty \leq \tau \}
  \]
Equivalence & Strong Equivalence

Definition:

(i) $\ell_0$- and $\ell_1$-problems are said to be **equivalent** if there exists a solution to $\ell_0$-problem that coincides with the unique solution to the $\ell_1$-problem.

(ii) $\ell_0$- and $\ell_1$-problems are said to be **strongly equivalent** if the $\ell_0$-problem has a unique solution which coincides with the unique solution to the $\ell_1$-problem.

At the moment, the understanding of the relationship between $\ell_0$- and $\ell_1$-problems is mainly focused on the strong equivalence.

Some Strong Equivalence Conditions

- **Mutual Coherence condition** [Donoho and Elad (2003)]:

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

where \(\mu(A)\) is the mutual coherence defined as

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

- **Restricted Isometry Property (RIP)** [Candès and Tao (2005)]: *The matrix \(A\) has the restricted isometry property (RIP) of order \(k\) if there exists a constant \(0 < \delta_k < 1\) such that*

\[
(1 - \delta_k)\|z\|^2_2 \leq \|Az\|^2_2 \leq (1 + \delta_k)\|z\|^2_2
\]

*for all \(k\)-sparse vector \(z\).*
Null Space Property (NSP) [Cohen et al (2009), Zhang (2008), etc.]: The matrix $A$ has the NSP of order $k$ if

$$\|h_\Lambda\|_1 < \|h_{\Lambda_c}\|_1 \quad \left(\text{i.e., } \|h_\Lambda\|_1 \leq \frac{1}{2} \|h\|_1\right)$$

holds for all $0 \neq h \in \mathcal{N}(A)$ and all $\Lambda \subseteq \{1, 2, \ldots, n\}$ such that $|\Lambda| \leq k$.

Range Space Property (RSP) of order $K$ [Zhao 2013]: The matrix $A^T$ is said to satisfy the range space property of order $K$ if for any disjoint subsets $S_1, S_2$ of $\{1, \ldots, n\}$ with $|S_1| + |S_2| \leq K$, there is a vector $\eta \in \mathcal{R}(A^T)$ such that

$$\eta_i = 1 \ \forall i \in S_1; \quad \eta_i = -1 \ \forall i \in S_2; \quad \text{otherwise } |\eta_i| < 1.$$
Restriction of ‘strong equivalence’ Criteria

- The strong equivalence between $\ell_0$- and $\ell_1$-problems is essential to the compressed-sensing theory [Candés (2006), Donoho (2006)].

- However, the strong-equivalence-type conditions can only partially explain the numerical behavior of the $\ell_1$-method in many situations. [Candés (2008), Elad (2010), Zhao (2012)].

- The probabilistic analysis [Candès and Romberg, 2005] demonstrates that the $\ell_1$-method is more powerful of finding sparse solutions of linear systems than what is indicated by various strong-equivalence criteria.

- From a mathematical point of view, it is also important to understand the equivalence between $\ell_0$- and $\ell_1$-problems.
Questions:

1. How to deterministically interpret the actual numerical performance of $\ell_1$-minimization more efficiently than strong-equivalence type criteria?

2. How to deterministically understand the limitation of $\ell_1$-minimization for locating the sparsest solution of linear systems?

3. If a linear system has multiple sparsest solutions, when can $\ell_1$-minimization guarantee to find one of them?

The key step is to completely characterize the uniqueness of the solution to $\ell_1$-problems, which is central to both recovering sparse signals and solving $\ell_0$-problems.
Strict complementarity property of LP

Consider the LP problem

\[(P) \quad \min \{ c^T x : \; Qx = p, \; x \geq 0 \}, \]

and its dual problem

\[(DP) \quad \max \{ p^T y : \; Q^T y + s = c, \; s \geq 0 \}, \]

By optimality, \((x^*, (y^*, s^*))\) is a solution pair of \((P)\) and \((DP)\) if and only if it satisfies the conditions

\[Qx^* = p, \; x^* \geq 0, \; Q^T y^* + s^* = c, \; s^* \geq 0, \; (x^*)^T s^* = 0.\]

**Theorem** [Schrijver(1989)]. Let \((P)\) and \((DP)\) be feasible. Then there exists a pair \((x^*, (y^*, s^*))\) of strictly complementary solutions of \((P)\) and \((DP)\), i.e., \((x^*)^T s^* = 0\) and \(x^* + s^* > 0\).
The uniqueness of $\ell_1$-minimizers

**Theorem 1 (Necessary Conditions) [Zhao 2013]**

If $x$ is the unique solution to the $\ell_1$-problem, then

(i) the matrix $(A_{J_+} A_{J_-})$ has full column rank, where

$$J_+ = \{ i : x_i > 0 \}, \quad J_- = \{ i : x_i < 0 \}.$$ 

(ii) there exists a vector $\eta$ such that

\[
\begin{cases}
\eta & \in \mathcal{R}(A^T), \\
\eta_i & = 1 \quad \text{for all} \quad x_i > 0, \\
\eta_i & = -1 \quad \text{for all} \quad x_i < 0, \\
|\eta_i| & < 1 \quad \text{for all} \quad x_i = 0.
\end{cases}
\]
The uniqueness of $\ell_1$-minimizers

Merging with Fuchs’ sufficient condition (2004) yields

**Theorem 2 (Necessary and Sufficient Condition)**

$x$ is the unique solution to the $\ell_1$-problem, if and only if the following two conditions hold:

(i) the matrix \(( A_{J_+} A_{J_-} )\) has full column rank, where

\[ J_+ = \{ i : x_i > 0 \}, \quad J_- = \{ i : x_i < 0 \}. \]

(ii) there exists a vector $\eta$ such that

\[
\begin{cases}
\eta & \in \mathcal{R}(A^T), \\
\eta_i &= 1 \quad \text{for all} \quad x_i > 0, \\
\eta_i &= -1 \quad \text{for all} \quad x_i < 0, \\
|\eta_i| &< 1 \quad \text{for all} \quad x_i = 0.
\end{cases}
\]

(1)
Guaranteed recovery

**Corollary 3.** If $x$ is the unique solution to $\ell_1$-minimization, then $\|x\|_0 \leq m$, i.e., $x$ must be at least $m$-sparse.

This Corollary justifies the role of $\ell_1$-method as a sparsity-seeking method.

**Definition.** A solution $x$ of the system $Ax = b$ is said to have a guaranteed recovery by the $\ell_1$-method if $x$ is the unique solution to the $\ell_1$-problem.

Any $x$ with sparsity $\|x\|_0 > m$ is definitely not the unique solution of $\ell_1$-problem. Thus there is no guaranteed recovery for such a solution by $\ell_1$-minimization.
Theorem 4. $\ell_0$- and $\ell_1$-problems are equivalent if and only if the range space property (RSP) holds at a sparsest solution $x$ to the linear system, i.e.,

$$\begin{align*}
\eta & \in R(A^T), \\
\eta_i & = 1 \quad \text{for all } x_i > 0, \\
\eta_i & = -1 \quad \text{for all } x_i < 0, \\
|\eta_i| & < 1 \quad \text{for all } x_i = 0.
\end{align*}$$

Corollary 5. $\ell_0$-problem is computationally tractable when the RSP holds at a sparsest solution of the linear system $Ax = b$. 
Efficiency of $\ell_1$-method

Some features of RSP

- All existing sufficient conditions for the strong equivalence of $\ell_0$- and $\ell_1$-problems imply RSP property.
- The RSP does not require the uniqueness of the sparsest solution.
- When $\ell_0$-problem has multiple sparsest solutions, $\ell_1$-method can still solve the $\ell_0$-problem, when the RSP holds at a sparsest solution.

Categories of linear systems:

Class 1: Both $\ell_1$- and $\ell_0$-problems have a unique solution

Class 2: $\ell_1$-problem has a unique solution, but $\ell_0$-problem has multiple solutions.

Class 3: Both $\ell_1$- and $\ell_0$-problems have multiple solutions.
RSP doesn’t require uniqueness of sparsest solutions

Example 6. Consider the linear system $Ax = b$ with

$$A = \begin{pmatrix} 1 & 0 & -2 & 5 \\ 0 & 1 & 4 & -9 \\ 1 & 0 & -2 & 5 \end{pmatrix}, \text{ and } b = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

The system $Ax = b$ has multiple sparsest solutions:

$$x^{(1)} = (1, -1, 0, 0)^T,$$
$$x^{(2)} = (0, 1, -1/2, 0)^T,$$
$$x^{(3)} = (0, 4/5, 0, 1/5)^T,$$
$$x^{(4)} = (0, 0, 2, 1)^T,$$
$$x^{(5)} = (1/2, 0, -1/4, 0)^T,$$
$$x^{(6)} = (4/9, 0, 0, 1/9)^T.$$
The mutual coherence, RIP and NSP criteria do not apply to this example, since the system has multiple sparsest solutions.

The RSP holds at \( x^{(6)} \). Indeed, by taking \( u = (1, \frac{4}{9}, 0)^T \), we have that

\[
\eta = A^T u = \left(1, \frac{4}{9}, -\frac{2}{9}, 1\right)^T,
\]

which satisfies the RSP at \( x^{(6)} \).

By Theorem 4, \( \ell_0 \)- and \( \ell_1 \)-problems are equivalent.

This linear system is in Class 2.
Example 7. Consider the system $Ax = b$ with

$$A = \begin{pmatrix}
\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
\frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{2}} \\
0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\
\end{pmatrix}, \quad b = \begin{pmatrix}
1 \\
1 \\
\end{pmatrix}$$

Clearly, $x^* = (0, 0, \sqrt{3}, 0, 0, 0)$ is the unique sparsest solution to this linear system.

The mutual coherence condition

$$\|x^*\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(A)}) = 1$$

fails.

The RIP of order 2 fails, and the NSP of order 2 also fail.
However, the RSP holds at $x^\ast$.

In fact, taking $y = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ yields

$$\eta = A^T y = \left(\sqrt{\frac{2}{3}}, \sqrt{\frac{2}{3}}, 1, \frac{1 - \sqrt{3} - \sqrt{2}}{3\sqrt{2}}, 0, -\sqrt{\frac{2}{3}}\right),$$

which satisfies the RSP at $x^\ast$.

This example shows that even for problems in Class 1, existing strong equivalence conditions may still fail to confirm the strong equivalence of $\ell_1$- and $\ell_0$-problems, but the RSP can.
Summary

1. The strong equivalence conditions only apply to problems in Class 1, and hence cannot explain the success of $\ell_1$-method for solving $\ell_0$-problems in Class 2.

2. The numerical performance of the $\ell_1$-method can be broadly explained by the RSP-based analysis.

3. The RSP-based analysis shows that the equivalence of $\ell_0$- and $\ell_1$-problems can be achieved not only for a subclass of problems in Class 1, but also for a subclass of problems in Class 2.

4. Moreover, the RSP-based theory also sheds light on the limitation of $\ell_1$-methods. Failing to satisfy the RSP, a sparsest solution definitely has no guaranteed recovery by the $\ell_1$-method.
Application to Compressed sensing

- Suppose that we would like to recover a sparse signal $x^*$. To serve this purpose, the so-called sensing matrix $A \in \mathbb{R}^{m \times n}$ with $m < n$ is constructed, and the measurements $y = Ax^*$ are taken.

- Then we solve the $\ell_1$-minimization $\min \{ \| x \|_1 : Ax = y \}$ to obtain a solution $\hat{x}$.

Two questions:

- What class of sensing matrices can guarantee the exact recovery $\hat{x} = x^*$?

- How sparse should $x^*$ be in order to be exactly recovered by $\ell_1$-method?
Uniform Recovery

Definition:

(i) The exact recovery of all $k$-sparse vectors (i.e., $\{x : \|x\|_0 \leq k\}$) by a single sensing matrix $A$ is called uniform recovery.

(ii) The spark of a given matrix, denoted by Spark$(A)$, is the smallest number of columns of $A$ that are linearly dependent (see e.g., Donoho and Elad (2003)).

Under the RIP and NSP of order $2k$, the following result was shown by Candes (2008), Cohen, et al. (2009).

Theorem. If $A$ satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2} - 1$, or if $A$ satisfies the NSP of order $2k$, then all $k$-sparse signals can be exactly recovered, where $k < \frac{1}{2} \text{Spark}(A)$. 
Range space property (RSP) of order $K$

**RSP of order $K$.** The matrix $A^T$ is said to satisfy the range space property of order $K$ if for any disjoint subsets $S_1, S_2$ of $\{1, ..., n\}$ with $|S_1| + |S_2| \leq K$, the range space $\mathcal{R}(A^T)$ contains a vector $\eta$ such that

$$\eta_i = 1 \quad \forall i \in S_1; \quad \eta_i = -1 \quad \forall i \in S_2; \quad \text{otherwise} \quad |\eta_i| < 1.$$ 

**Lemma 8.** Suppose that one of the following holds:

- $K < \frac{1}{2} \left(1 + \frac{1}{\mu(A)} \right)$.
- The matrix $A$ has the RIP of order $2K$ with constant $\delta_{2K} < \sqrt{2} - 1$.
- The matrix $A$ has the NSP of order $2K$.

Then the matrix $A^T$ has the RSP of order $K$. 
Uniform Recovery Theorem

**Theorem 9.** Any $x$ with $\|x\|_0 \leq K$ can be exactly recovered by $\ell_1$-minimization if and only if $A^T$ has the RSP of order $K$.

Thus the RSP of order $K$ is a necessary and sufficient condition for exactly recovering all $K$-sparse vectors, so the RSP of order $K$ has completely characterized the uniform recovery by the $\ell_1$-method.

**Lemma** (Upper bound for $K$) If $A^T$ has the RSP of order $K$, then any $K$ columns of $A$ are linearly independent, so $K < \text{Spark}(A)$. 
Beyond the uniform recovery

- RIP or NSP of order $2k$ can recover a $k$-sparse vector with $k < \text{Spark}(A)/2$. From a mathematical point of view, it is interesting to know how a vector $x$ with

$$\frac{1}{2}\text{Spark}(A) \leq \|x\|_0 < \text{Spark}(A)$$

can be possibly recovered.

- This is also motivated by some practical applications, where an unknown vector (representing a signal or an image) might not be sparse enough to be in the range $\|x\|_0 < \text{Spark}(A)/2$.

- Theorem 2 makes it possible to handle such a situation by introducing a weak-RSP concept governing the so-called non-uniform recovery of signals.
Extended to problems with nonnegativity constraints

\[
\begin{align*}
\min\{\|x\|_0 : Ax = b, \ x \geq 0\}, \\
\min\{\|x\|_1 : Ax = b, \ x \geq 0\},
\end{align*}
\]

**Theorem** If \(x\) is the unique least \(\ell_1\)-norm nonnegative solution to the system \(Ax = b\) if and only if there exists a vector \(\eta \in \mathbb{R}^n\) satisfying

\[
\eta \in \mathcal{R}(A^T), \ \eta_i = 1 \text{ for } i \in J_+, \text{ and } \eta_i < 1 \text{ for } i \notin J_+,
\]

where \(J_+ = \{i : x_i > 0\}\).
Application to linear programs

\[ \min\{ c^T x : Ax = b, x \geq 0 \} . \]

- In many situations, reducing the number of activities is vital for efficient planning, management and resource allocations.

- The sparsest optimal solution of a linear program provides the smallest number of activities to achieve the optimal objective value.

Let \( d^* \) be the optimal value of LP. The optimal solution set of the LP is given by

\[ S^* = \{ x : Ax = b, x \geq 0, c^T x = d^* \} . \]
A sparsest optimal solution to the LP is an optimal solution to the $\ell_0$-problem

$$\min \left\{ \|x\|_0 : \begin{pmatrix} A \\ c^T \end{pmatrix} x = \begin{pmatrix} b \\ d^* \end{pmatrix}, x \geq 0 \right\},$$

associated with which is the $\ell_1$-problem

$$\min \left\{ \|x\|_1 : \begin{pmatrix} A \\ c^T \end{pmatrix} x = \begin{pmatrix} b \\ d^* \end{pmatrix}, x \geq 0 \right\}.$$

**Theorem 10.** $x$ is the unique least $\ell_1$-norm optimal solution to LP if and only if the matrix $H = \begin{pmatrix} A_{J^+} \\ c_{J^+}^T \end{pmatrix}$ has full column rank, and there exists a vector $\eta \in \mathbb{R}^n$ obeying

$$\eta \in \mathcal{R}([A^T, c]), \eta_i = 1 \ \forall i \in J^+, \ and \ \eta_i < 1 \ \forall i \notin J^+$$

where $J^+ = \{i : x_i > 0\}$. Moreover, a sparsest optimal solution to LP is the unique least $\ell_1$-norm optimal solution of the LP if and only if the above RSP holds at this optimal solution.
The uniqueness of the solution to $\ell_1$-problem can be characterized: \textit{x is the unique solution to the $\ell_1$-problem if and only if the range space property (RSP) of $A^T$ and a full-rank property hold at x.}

It was shown that $\ell_0$- and $\ell_1$-problems are equivalent if and only if the range space property is satisfied at a sparest solution of linear systems.
Through the RSP-based analysis, the numerical efficiency and limitation of $\ell_1$-minimization for solving $\ell_0$-minimization can be deterministically explained.

We have shown that the equivalence of $\ell_0$- and $\ell_1$-problems exists for a broad range of linear systems in Classes 1 and 2.

Moreove, it was shown that all $k$-sparse signals can be exactly recovered if and only if the sensing matrix $A^T$ has the RSP of order $k$. 
Main references


Y.B. Zhao, Equivalence and Strong Equivalence between the Sparsest and Least $\ell_1$-Norm Nonnegative Solutions of Linear Systems and Their Applications, Technical reprot, University of Birmingham, June 2012.