Approximate Affine Linear Relationship Between L_1 Norm Objective Functional Values and L_2 Norm Constraint Bounds

Zhijing Yang
Telephone: +86 20 3932 2246 Email: yzhj@gdut.edu.cn
School of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China.
*Bingo Wing-Kuen Ling
Telephone: +86 20 3932 2246 Fax: +86 20 3932 2252 Email: yongquanling@gdut.edu.cn
School of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China.
Chris Bingham
Telephone: +44 (0)15 2283 7912 Fax: +44 (0)15 2288 6489 Email: cbingham@lincoln.ac.uk
School of Engineering, University of Lincoln, Lincoln, Lincolnshire, LN6 7TS, United Kingdom.

Abstract

For an optimization problem with an L_1 norm objective function subject to an L_2 norm inequality constraint, this paper shows that there is an approximately linear relationship between the L_1 norm objective functional values and the L_2 norm specifications. This relationship is verified through the use of random and real world industrial data. The obtained results can be employed for 1) estimating the L_1 norm objective functional value without solving the optimization problem numerically; 2) providing an insight for defining the L_2 norm specification in which a simple method is proposed in this paper; and 3) testing whether the obtained solutions are the globally optimal solutions or not. These advantages are demonstrated via the use of random data.

Keywords: Sparse optimization, L_0 norm optimization, L_1 norm optimization, approximate linear relationship.

1. Introduction

Denote $\Re^{a\times b}$ and $\Re^d$ as the $a\times b$ real valued matrix space and the d dimensional real valued column vector space, respectively. Let $A=\begin{bmatrix}\overline{A}_1 & \cdots & \overline{A}_n\end{bmatrix}\in\Re^{m\times n}$ be an overcomplete dictionary with $\overline{A}_i\in R^m$ for $i=1,\cdots,n$ denotes the atoms. Let $y\in R^m$ be a signal with $m<n$. Let $x\in\Re^n$ be a coefficient vector. Denote $\|x\|_0$ as the L_0 norm of x and it refers to the total number of the nonzero elements in x. The problem of finding a sparse representation of y using A is to find x such that $y=Ax$ and $\|x\|_0$ is minimized. Denote x_0^* as the optimal solution of the L_0 norm optimization problem. That is:

$$x_0^*=\arg\min_x \|x\|_0 \qquad s.t. \qquad y=Ax. \tag{1}$$

Since only few coefficients are required for the representations of the signals, hardware implementations of these representations are very efficient. Actually, the optimal representations of the signals of many industrial databases can be formulated as sparse optimization problems [1]-[3]. Hence, the signal representations via the sparse optimizations has drawn a great attention in recent decade [11]-[13].

Denote $rank(A)$ as the total number of independent rows or the total number of independent columns of A. It is worth noting that x_0^* may not be unique for any arbitrary $y \in R^m$ and $rank(A) = m$. Let $\hat{A}$ be a matrix containing only $\hat{m}$ columns of A in which $\hat{m} < m$ and denote $col(\hat{A})$ as the column space of $\hat{A}$. For $y \in col(\hat{A}) \subset \Re^m$, finding x_0^* requires combinational searches. On the other hand, as the L_1 norm operator is the convex relaxation of the L_0 norm operator, it is a common practice to replace the L_0 norm operator in (1) by the L_1 norm operator [4]-[7]. Denote $\| x \|_1$ as the L_1 norm of x. Let x_1^* be the optimal solution of the L_1 norm optimization problem. That is:

$$x_1^* = \arg\min_x \| x \|_1 \qquad s.t. \qquad y = Ax. \tag{2}$$

Here, This L_1 norm problem can be solved via linear programming approaches. For using the L_1 norm operator, several conditions that guarantee $x_0^* = x_1^*$ have been derived [5]-[7]. However, these conditions are derived based on deterministic approaches, so they are only sufficient conditions which are too tight to be applied to practical industrial problems. With the probabilistic framework, some relaxed conditions are derived [4]. Denote $\alpha_0 \in R^n$ as the solution of (1) and ρ as a constant between 0 and 1. It is shown in [4] that if

"*there exists* y *such that* $y = A\alpha_0$ *and* α_0 *has fewer than* ρm *nonzero components, then* x_0^* *is unique and* $x_0^* = \alpha_0 = x_1^*$." (3)

Under ideal circumstances, the constraint in (2) can be satisfied exactly. However, if y contains a significant amount of noise which is a typical condition in industrial environments, then satisfying the constraint in (2) exactly is not meaningful. For this case, the equality constraint is relaxed to an inequality constraint. Denote $\| x \|_2$ as the L_2 norm of x. Denote ε as the acceptable user defined bound on the L_2 norm of the difference between the original signal and the reconstructed signal. Let x_ε^* be the optimal solution of the L_1 norm inequality constrained optimization problem. That is:

$$x_\varepsilon^* = \arg\min_x \| x \|_1 \qquad s.t. \qquad \| Ax - y \|_2 \le \varepsilon. \tag{4}$$

Since $\| x_\varepsilon^* \|_1$ characterizes the scarcity of the signal representation and ε determines the L_2 norm of the difference between the original signal and the reconstructed signal, both $\| x_\varepsilon^* \|_1$ and ε should be as small as possible. Although selecting a small value of ε can guarantee a small bound on $\| Ax_\varepsilon^* - y \|_2$, it will result to a large value of $\| x_\varepsilon^* \|_1$. Therefore, a tradeoff exists between $\| x_\varepsilon^* \|_1$ and ε. To address this tradeoff issue, the optimization problems minimizing both $\| Ax_\varepsilon^* - y \|_2$ and $\| x_\varepsilon^* \|_1$ [8] or the optimization problems minimizing $\| Ax_\varepsilon^* - y \|_2$ subject to the specifications on $\| x_\varepsilon^* \|_1$ [9] are solved instead by the least angle regression (LARS) algorithms or the least absolute shrinkage and selection operator (LASSO) algorithms. However, these optimization problems are different from the original optimization problems (the optimization problems with the L_1 norm objective functions subject to the L_2 norm specifications). Hence, the obtained solutions of these optimization problems are different. Denote z_0 as a standard white noise. Define

$\sigma = 0.2\frac{\|Ax_0\|_2}{\|z_0\|_2}$. Here, x_0 is the original sparse signal and $y = Ax_0 + \sigma z_0$. In this case, ε is selected as $\varepsilon^2 = \sigma^2\left(m + 2\sqrt{2m}\right)$ [14]. In this case, the probability that $\|\sigma z_0\|^2$ exceeds ε^2 is equal to the probability that a chi square with m degrees of freedom exceeds its mean by at least two standard deviations. This quantity is about 2.5% when m is not too small. However, this approach is based on a statistical model and the noise characteristics are known. Nevertheless, the system model may not be probabilistic. Even though the system model is probabilistic, the noise characteristics are unknown in practical situations. Hence, it is important to derive a methodology to determine the value of ε when the system model is not probabilistic or the noise characteristics are unknown. This paper is to address this issue.

Since it is of significantly important to adaptively choose ε to obtain the best representation, it is required to investigate the relationship between $\| x_\varepsilon^* \|_1$ and ε. Although the empirical relationship between the L_2 norm objective functional values and the L_∞ norm specifications has been recently investigated [10], the relationship between the L_1 norm objective functional values and the L_2 norm specifications has not been studied yet.

Unlike the optimization problems with the L_2 norm objective functions subject to the L_1 norm specifications where their feasible sets are the convex hulls of a set of vertices and bounded by linear hyperplanes, the feasible sets of the optimization problems with the L_1 norm objective functional values subject to the L_2 norm specifications are not characterized by the convex hulls of a set of vertices and not bounded by linear hyperplanes. Hence, this kind of optimization problems cannot be solved via conventional LARS algorithms or LASSO algorithms. In fact, these optimization problems are highly non-traceable. It is very difficult to find their analytical solutions and to characterize the relationship between $\| x_\varepsilon^* \|_1$ and ε analytically. This paper is to address this very important issue.

The outline of this paper is as follows. The approximate linear relationship between the L_1 norm objective functional values and the L_2 norm specifications is presented in Section 2. A design of the L_2 norm specification is proposed in Section 3. Finally, conclusions are drawn in Section 4.

2. Approximate linear relationship between L_1 norm objective functional values and L_2 norm specifications

The explanation of having the approximate data independent linear relationship between the L_1 norm objective functional values and the L_2 norm specifications is as follow. Since the optimization problem is convex and the globally optimal solution of the corresponding unconstrained optimization problem is not in the feasible set of the original optimization problem, the globally optimal solution of the original optimization problem is on the boundary of its feasible set. This implies that the globally optimal solution of the original optimization problem can be governed by an equality constraint defined by the specification bound multiplied by the unit vector. By representing the globally optimal solution into two subvectors $x_1^\bullet$ and $x_2^\bullet$, the above equality constraint can be used to eliminate one subvector $x_1^\bullet$ in terms of another subvector $x_2^\bullet$ and the unit vector. Also, the L_1 norm objective function can be expressed in terms of one subvector $x_2^\bullet$ and the unit vector. On the other hand, since the globally optimal solution is sparse, the globally optimal solution can also expressed as two subvectors z_1 and z_2 in which one subvector z_1 is not sparse and the other subvector z_2 is very small. Here, we have two sets of subvectors $\{ x_1^\bullet, x_2^\bullet \}$ and $\{ z_1, z_2 \}$

in which they are related by a simple row operation. As one subvector z_2 is very small, we can assume that this subvector z_2 is the zero vector. This implies that we have another equality constraint governing the relationship between the unit vector and the original subvector $x_2^{\bullet}$ via the row operation. As a result, we have a relationship between another subvector z_1 and the unit vector. Hence, the original optimization problem can be approximated by an optimization problem with the L_1 norm objective function of the unit vector subject to this unit vector constraint. This optimization problem can be represented by an optimization problem with a linear objective function subject to the unit vector constraint. By using the Lagrange approach, the form of the unit vector is obtained. Hence, the form of the approximated globally optimal solution of the original optimization problem can be obtained accordingly. In fact, the obtained solution is a linear combination of the specification bound multiplied by a constant vector and y. Hence, if this constant vector and y have the same direction, then the L_1 norm objective functional values and the L_2 norm specifications have a near affine linear relationship. The detail explanations are in Appendix A.

Here, we have assumed that $\|y\|_2 > \varepsilon$ as well as both A_1 and $R_4 - R_3A_1^{-1}A_2$ are full rank matrices. Since there are many inverse operators in Appendix A, it requires to study the invertibility of these matrices. The analysis is shown in Appendix B. These assumptions are usually valid in practical situations. From the above, we can see that the accuracy of the approximation depends on how small z_2 is, whether ι, R_1, R_2, R_3 and R_4 are remained unchanged or not, as well as

$$\begin{bmatrix} \dfrac{-\varepsilon\left(I + A_1^{-1}A_2\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}} \\ \dfrac{\varepsilon\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}} \end{bmatrix}$$

and $\begin{bmatrix} \left(I + A_1^{-1}A_2\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y \\ -\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y \end{bmatrix}$ are in the same direction or not. If $\|z_2\|_1$ is small,

$$\begin{bmatrix} \dfrac{-\varepsilon\left(I + A_1^{-1}A_2\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}} \\ \dfrac{\varepsilon\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}} \end{bmatrix}$$

and $\begin{bmatrix} \left(I + A_1^{-1}A_2\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y \\ -\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y \end{bmatrix}$ are in the same direction, as well as only considering ε within a small neighborhood, then the approximation should be accurate enough for most practical situations.

To verify the approximate linear relationship between $\|x_\varepsilon^*\|_1$ and ε, many test pairs of A and y are generated. After performing these tests, it is found that the approximate linear relationship between $\|x_\varepsilon^*\|_1$ and ε is valid for all these pairs of A and y. Only three types of experiments are presented below to support the proposed assertions because of the page limit.

Experiment 1: Signals satisfying the condition in (3)

Here, the relationship between $\|x_\varepsilon^*\|_1$ and ε when the condition in (3) is satisfied is investigated. In this experiment, $m=40$ and $n=100$. All the elements of A are independently drawn from a zero mean Gaussian distribution. Then, $\overline{A}_i$ for $i=1,\cdots,100$ is further normalized to a unit energy vector. There are 5 nonzero elements in α_0 and the locations of these elements are selected randomly with their polarities also being chosen randomly. The amplitude of the smallest nonzero element of α_0 is set to *0.2* and the amplitudes of other elements are set to 1. Using this α_0, y is obtained. Obviously, $x_0^*=\alpha_0$. x_0^* is shown in Fig. 1(a). When $\varepsilon=0,\frac{\|y\|_2}{20},\frac{\|y\|_2}{10}$ and $\frac{3\|y\|_2}{20}$, the corresponding x_ε^* are shown in Figs. 1(b)-(e), respectively. It is shown in [4] that $x_0^*=x_1^*$ when $\rho\le\frac{1}{5}$. This can also be verified from Figs. 1(a)-(b). Although a large value of ε allows a large bound on $\|Ax_\varepsilon^*-y\|_2$, it results to a small value of $\|x_\varepsilon^*\|_1$. In other words, we have a high scarcity of x_ε^*. It can be seen in Fig. 1(e) that the total number of the nonzero elements in x_ε^* is 4 when $\varepsilon=\frac{3\|y\|_2}{20}$. This further demonstrates the reason for the relaxation on the L_2 norm specification.

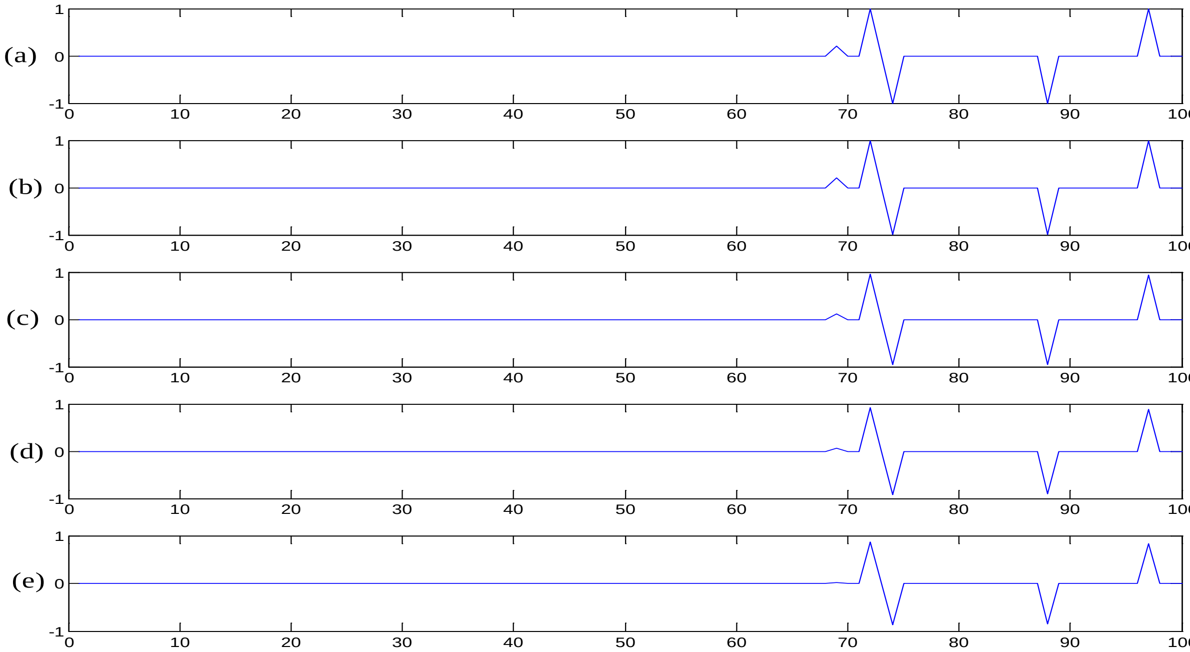

Fig. 1. (a) x_0^*. (b)-(e): x_ε^* for $\varepsilon=0,\frac{\|y\|_2}{20},\frac{\|y\|_2}{10}$ and $\frac{3\|y\|_2}{20}$, respectively.

To further investigate the relationship between $\|x_\varepsilon^*\|_1$ and ε, 100 test pairs of A and y are generated as described in the above. For each test pair of A and y, various values of ε, denoted as ε_i, are obtained via uniformly sampling in the range $[0,\|y\|_2]$. Denote the corresponding solution of (4) as $x_{\varepsilon_i}^*$ for $i=1,2,\cdots,N$. It is found that an approximate linear relationship exists between $\|x_{\varepsilon_i}^*\|_1$ and ε_i for all these 100 test pairs of A and y. For an illustration purpose, $(\varepsilon_i,\|x_{\varepsilon_i}^*\|_1)$ of one of the test pairs of A and y are shown in Fig. 2. A straight line linear regression fit of $(\varepsilon_i,\|x_{\varepsilon_i}^*\|_1)$ denoted as $f(\varepsilon_i)=k^*\varepsilon_i+b^*$ is also shown in Fig. 2. The relative L_2 norm error is defined as

$$\frac{\sqrt{\sum_{i=1}^{N} | f(\varepsilon_i) - \| x_{\varepsilon_i}^* \|_1 |^2}}{\sqrt{\sum_{i=1}^{N} (\| x_{\varepsilon_i}^* \|_1)^2}}. \tag{5}$$

It is found that the L_2 norm error is around 2.3%. On the other hand, it is also found that an approximate linear relationship exists between $\left\| x_{\varepsilon_i}^* - x_0^* \right\|_2$ ($\left\| x_{\varepsilon_i}^* - x_0^* \right\|_2$ is the root mean squares error between $x_{\varepsilon_i}^*$ and x_0^*) and ε_i for all these 100 test pairs of A and y. For an illustration purpose, $(\varepsilon_i, \left\| x_{\varepsilon_i}^* - x_0^* \right\|_2)$ of one of the test pairs of A and y are shown in Fig. 3. A linear regression fit of $(\varepsilon_i, \left\| x_{\varepsilon_i}^* - x_0^* \right\|_2)$ denoted as $g(\varepsilon_i) = \bar{k}^* \varepsilon_i + \bar{b}^*$ is also shown in Fig. 3. The relative L_2 norm error is defined as

$$\frac{\sqrt{\sum_{i=1}^{N} | g(\varepsilon_i) - \| x_{\varepsilon_i}^* \|_1 |^2}}{\sqrt{\sum_{i=1}^{N} (\| x_{\varepsilon_i}^* \|_1)^2}}. \tag{6}$$

It is found that the relative L_2 norm error is around 1.8%. All these 100 test pairs of A and y show similar results.

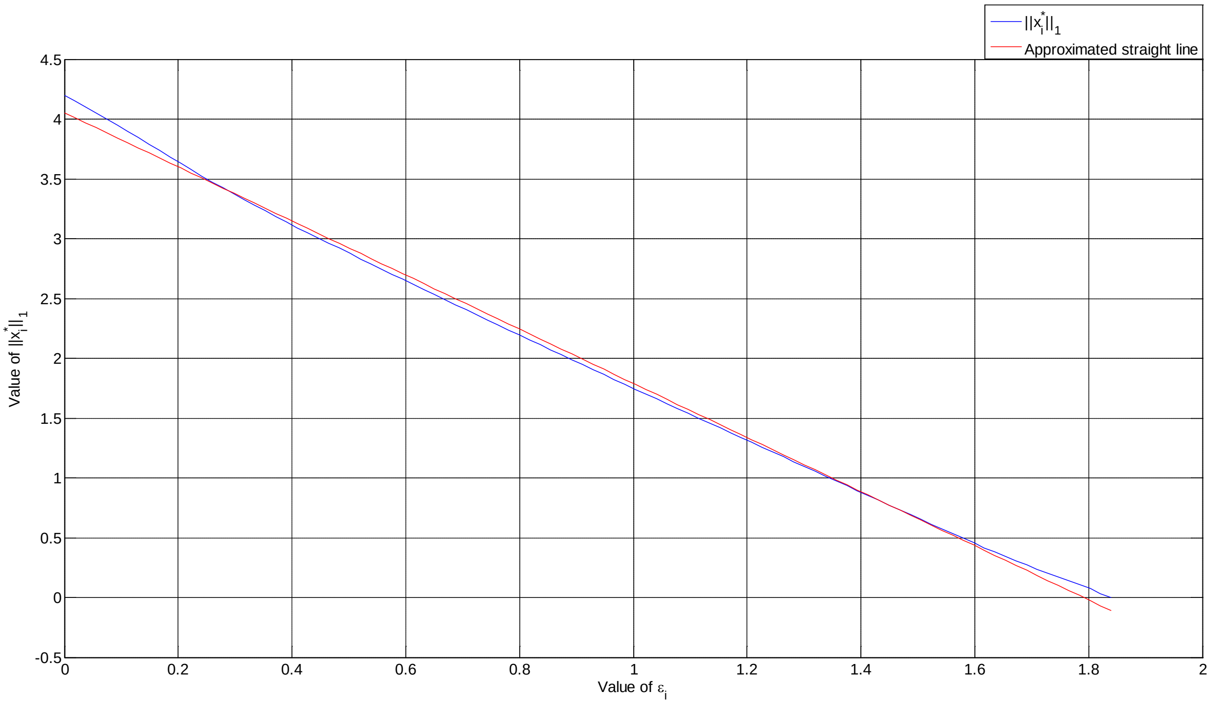

Fig. 2. $(\varepsilon_i, \| x_{\varepsilon_i}^* \|_1)$ and the linear regression fit.

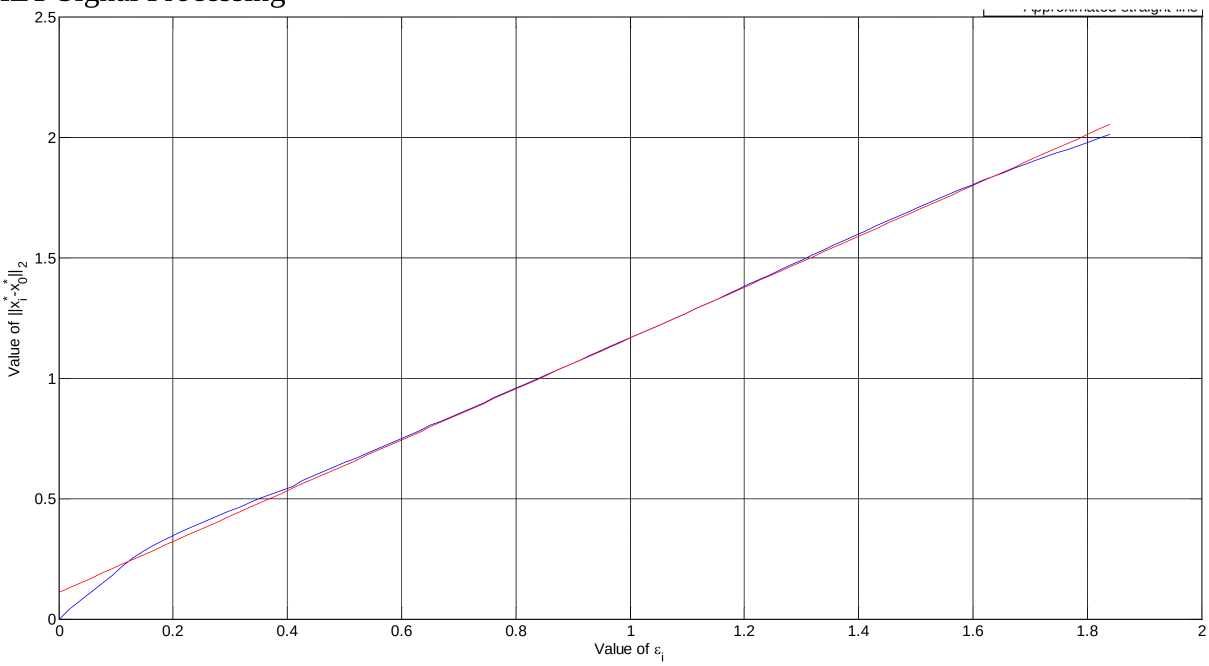

Fig. 3. $(\varepsilon_i, \left\| x_{\varepsilon}^* - x_0^* \right\|_2)$ and the linear regression fit.

Experiment 2: Trials using randomly generated data

Since condition in (3) is difficult to satisfy in many practical situations, $x_0^* = x_1^*$ only happens in ideal scenarios. It is therefore necessary to investigate the relationship between $\| x_{\varepsilon}^* \|_1$ and ε when $x_0^* \neq x_1^*$. Here, $m = 40$ and $n = 400$ as well as A is randomly generated as above and the elements in y are also independently and randomly generated from a zero mean Gaussian distribution. For $0 \le \varepsilon_i \le \| y \|_2$, $(\varepsilon_i, \| x_{\varepsilon_i}^* \|_1)$ are shown in Fig. 4. Similarly, 100 test pairs of matrices A and y have been generated. It is found that the relationship between $\| x_{\varepsilon_i}^* \|_1$ and ε_i is again approximately linear for all these 100 test pairs of A and y. A straight line linear regression fit of $(\varepsilon_i, \| x_i^* \|_1)$ denoted as $f(\varepsilon_i) = k^* \varepsilon_i + b^*$ is also shown in Fig. 4. The relative L_2 norm error is found to be around 2.6%. On the other hand, it is also found that an approximate linear relationship exists between $\left\| x_{\varepsilon_i}^* - x_0^* \right\|_2$ and ε_i for all these 100 test pairs of A and y. For an illustration purpose, $(\varepsilon_i, \left\| x_{\varepsilon_i}^* - x_0^* \right\|_2)$ of one of the test pairs of A and y are shown in Fig. 5. A linear regression fit of $(\varepsilon_i, \left\| x_{\varepsilon_i}^* - x_0^* \right\|_2)$ denoted as $g(\varepsilon_i) = \bar{k}^* \varepsilon_i + \bar{b}^*$ is also shown in Fig. 5. The relative L_2 norm error is found to be 0.000046%. Again, all these 100 test pairs of A and y show similar results.

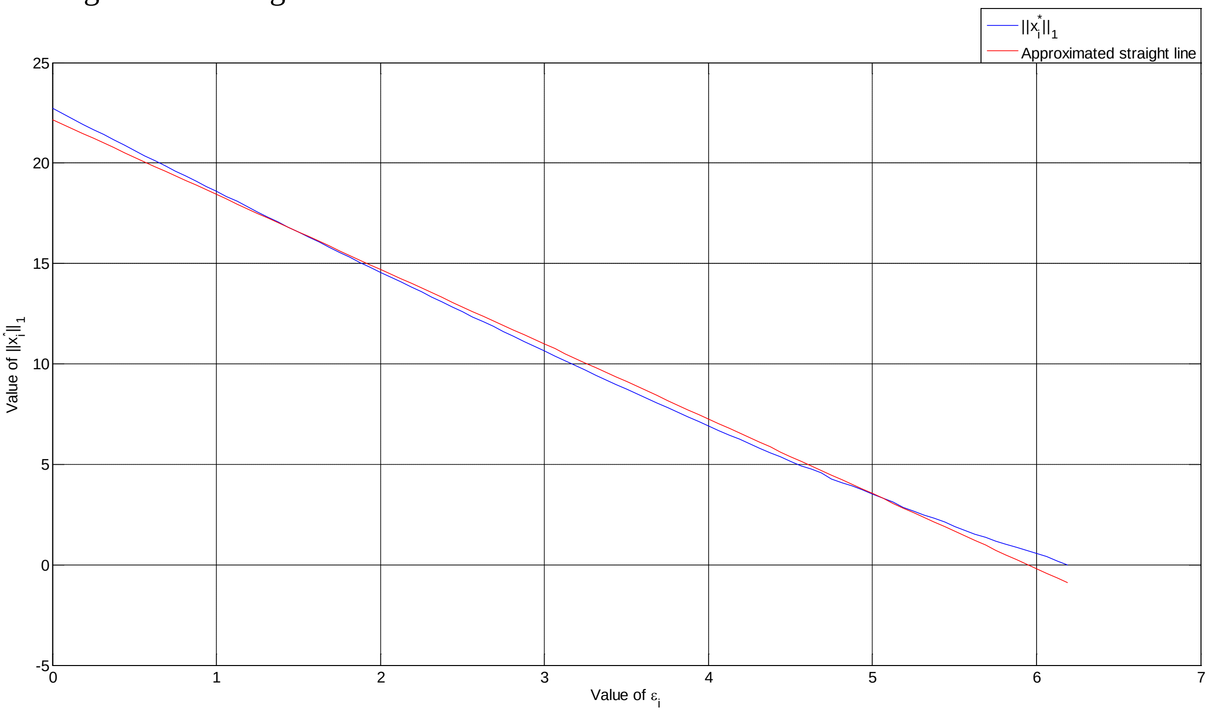

Fig. 4. $(\varepsilon_i, \| x_i^* \|_1)$ and the linear regression fit.

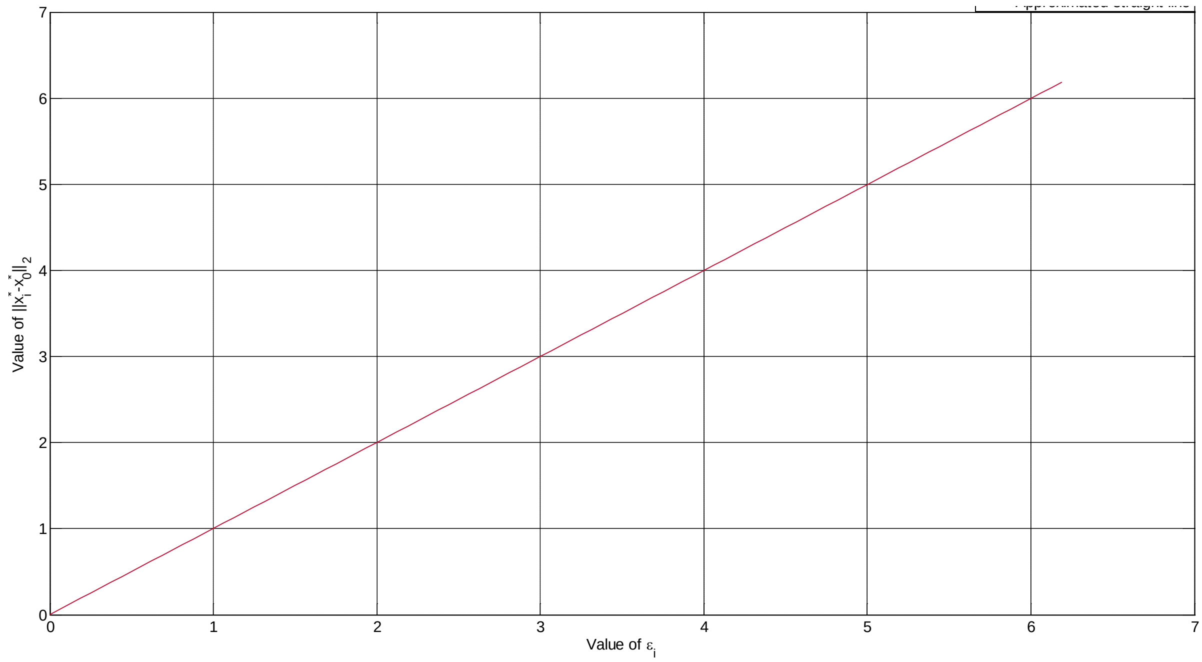

Fig. 5. $(\varepsilon_i, \left\| x_\varepsilon^* - x_0^* \right\|_2)$ and the linear regression fit.

Experiment 3: Trials using data from experimental measurements

To more readily demonstrate the application potential of the proposed relationships, vibration data from an industrial gas turbine taken over a 1 month period is illustrated as shown in Fig. 6. y is taken directly from the first 100 samples of data and then normalized to unit energy. On the other hand, $\overline{A}_1$ is taken from the second 100 samples of the data and then normalized to unit energy. Similarly, successive columns of A is taken from successive 100 samples of the data and then normalized to unit energy. For $0.05 \le \varepsilon_i \le \| y \|_2$, $(\varepsilon_i, \| x_{\varepsilon_i}^* \|_1)$ are shown in Fig. 7. A linear regression fit of $(\varepsilon_i, \| x_{\varepsilon_i}^* \|_1)$ is also shown in Fig. 7. It can be seen that the relationship between $\left\| x_{\varepsilon_i}^* \right\|_1$ and ε_i is again approximately linear with the relative L_2 norm error about 0.2%. Similarly, all these 100 test pairs of A and y show similar results.

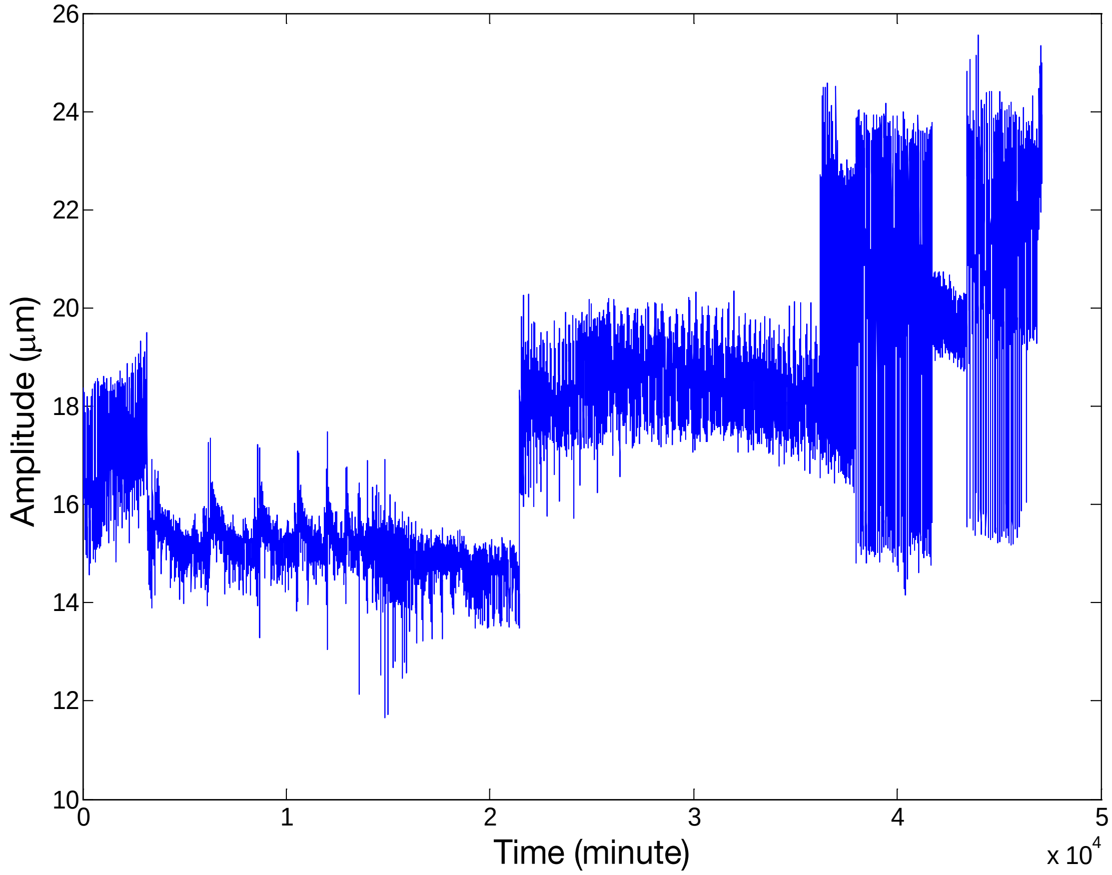

Fig. 6. Vibration measurements from an industrial gas turbine taken over a 1 month period.

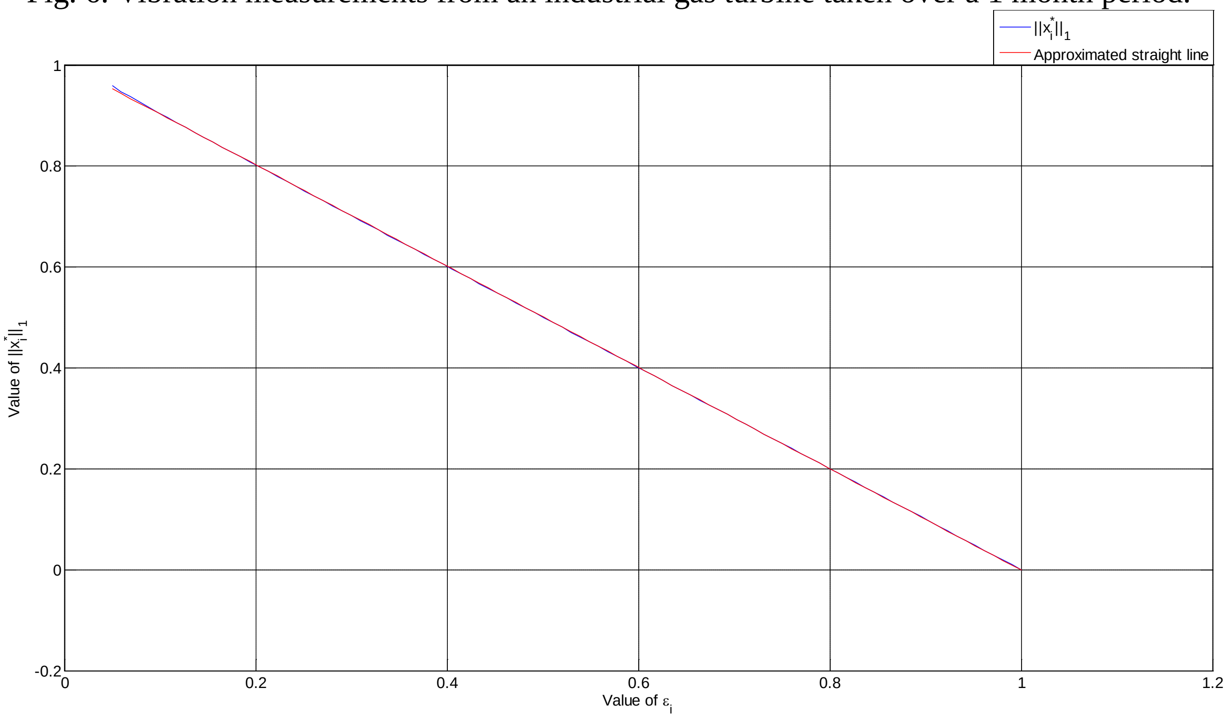

Fig. 7. $(\varepsilon_i, \| x_i^* \|_1)$ and the linear regression fit.

By exploring the approximately linear relationship between ε and $\| x_\varepsilon^* \|_1$, $\| x_\varepsilon^* \|_1$ can be estimated without solving (4) numerically. Given any ε, where $0 \le \varepsilon \le \| y \|_2$, an estimate of $\| x_\varepsilon^* \|_1$ is given by:

$$\| x_\varepsilon^* \|_1 = k^* \varepsilon + b^* . \quad (7)$$

From Appendix A, we can see that theoretically

$$k^* = \left\| \begin{bmatrix} \dfrac{-\left(I + A_1^{-1}A_2\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}} \\ \dfrac{\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1 - \left(R_2 - R_1A_1^{-1}A_2\right)\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}} \end{bmatrix} \right\|_1$$

and

$b^* = \left\| \begin{bmatrix} \left(I + A_1^{-1}A_2\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y \\ -\left(R_4 - R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y \end{bmatrix} \right\|_1$. Practically, we can compute two points in the straight line to find the values of k^* and b^*. Thereby, our result provides substantial reductions in computational overhead compared to traditional methods that rely on numerical approaches. Since the estimation error has been shown to be very small, the estimation of $\| x_\varepsilon^* \|_1$ by (7) will be very close to the true value.

3. Design of L_2 norm specification

The obtained linear relationship between $\| x_\varepsilon^* \|_1$ and ε provides an insight for defining the L_2 norm specification. First, we have the following property:

Property 1 Assume that $\overline{A}_i$ for $i = 1, \cdots, n$ have unit energy. For a given y and $0 \le \varepsilon \le \| y \|_2$, we have:

$$\| x_\varepsilon^* \|_1 \approx k^* \| Ax_\varepsilon^* - y \|_2 + b^*, \tag{8}$$

where k^* and b^* are constants and they are dependent on A and y.

Proof: Assume that $\| Ax_\varepsilon^* - y \|_2 = \varepsilon_1 < \varepsilon$. Let $\tilde{x}_\varepsilon^* = (1-\lambda)x_\varepsilon^*$, where $0 < \lambda \le 1$. Then, we have:

$$\begin{aligned} \| A\tilde{x}_\varepsilon^* - y \|_2^2 &= \| A(1-\lambda)x_\varepsilon^* - y \|_2^2 \\ &= (A(1-\lambda)x_\varepsilon^* - y)^T (A(1-\lambda)x_\varepsilon^* - y) \\ &= (1-\lambda)^2 (Ax_\varepsilon^*)^T (Ax_\varepsilon^*) - 2(1-\lambda)(Ax_\varepsilon^*)^T y + y^T y \\ &= (Ax_\varepsilon^*)^T (Ax_\varepsilon^*) - 2(Ax_\varepsilon^*)^T y + y^T y + (\lambda^2 - 2\lambda)(Ax_\varepsilon^*)^T (Ax_\varepsilon^*) + 2\lambda (Ax_\varepsilon^*)^T y \\ &= \| Ax_\varepsilon^* - y \|_2^2 + \lambda\left((\lambda - 2)\| Ax_\varepsilon^* \|_2^2 + 2(Ax_\varepsilon^*)^T y\right) \\ &\le \| Ax_\varepsilon^* - y \|_2^2 + \lambda\left(2\| Ax_\varepsilon^* \|_2^2 + 2|(Ax_\varepsilon^*)^T y|\right) \\ &= \| Ax_\varepsilon^* - y \|_2^2 + 2\lambda\left(\| Ax_\varepsilon^* \|_2^2 + |(Ax_\varepsilon^*)^T y|\right) \\ &= \varepsilon_1^2 + 2\lambda\left(\| Ax_\varepsilon^* \|_2^2 + |(Ax_\varepsilon^*)^T y|\right) \end{aligned}, \tag{9}$$

Since A, x_ε^* and y are known, let

$$\lambda = \min\{\frac{\varepsilon^2 - \varepsilon_1^2}{2(\| Ax_\varepsilon^* \|_2^2 + |(Ax_\varepsilon^*)^T y|)}, 1\}.$$

From (9), we have $\| A\tilde{x}_\varepsilon^* - y \|_2 \le \varepsilon$. This implies that $\tilde{x}_\varepsilon^*$ satisfies the constraint $\| A\tilde{x}_\varepsilon^* - y \|_2 \le \varepsilon$. Since $\tilde{x}_\varepsilon^* = (1-\lambda)x_\varepsilon^*$ and $0 < \lambda \le 1$, then $\| \tilde{x}_\varepsilon^* \|_1 = \| (1-\lambda)x_\varepsilon^* \|_1 < \| x_\varepsilon^* \|_1$. This contradicts the fact that x_ε^* is an optimal solution of (4). Hence, we have $\| Ax_\varepsilon^* - y \|_2 = \varepsilon$. According to the obtained approximately linear relationship between ε and $\| x_\varepsilon^* \|_1$, we have $\| x_\varepsilon^* \|_1 \approx k^* \| Ax_\varepsilon^* - y \|_2 + b^*$, where k^* and b^* are constants and they are dependant on A

and y. This completes the proof. ■

Since there is a tradeoff between $\| x_\varepsilon^* \|_1$ and ε, it is of significantly important to determine ε to obtain a better signal representation. In general, the choice of ε is application dependent. However, for most applications, ε should not be chosen as the extreme values such as $\varepsilon = 0$ and $\varepsilon = \| y \|_2$. On the other hand, ε should be chosen in the "middle" between 0 and $\| y \|_2$ in order to achieve a better tradeoff. By knowing the linear relationship between ε and $\| x_\varepsilon^* \|_1$, the "middle" between 0 and $\| y \|_2$ can be defined as the point such that

$$\| x_\varepsilon^* \|_1 \times \| A x_\varepsilon^* - y \|_2 \tag{10}$$

is maximized. That is, the area under the line $\| x_\varepsilon^* \|_1 \approx k^* \| A x_\varepsilon^* - y \|_2 + b^*$ is maximized. Obviously, this criterion can avoid to obtain a point at the extrema. Denote the optimal operating point as $(\varepsilon^*, \| x_{\varepsilon^*}^* \|_1)$. For a non-optimal point operating at $(\varepsilon_1, \| x_{\varepsilon_1}^* \|_1) = (\varepsilon_1, 0)$, we have $-\frac{b^*}{k^*} = \| y \|_2$. Hence, we have:

$$\varepsilon^* = -\frac{b^*}{2k^*} = \frac{\| y \|_2}{2}. \tag{11}$$

Now, (11) can be used to choose ε. Beside, since the problems in (4) are not traceable, it is difficult to guarantee that the obtained solutions are the globally optimal solutions. By using the obtained results in Property 1, it is easier to test whether the obtained solutions are the globally optimal solutions or not.

To illustrate the appropriateness of the proposed method for designing ε, a random $A \in R^{m \times n}$ with $m = 40$ and $n = 100$ as well as a random $y \in R^m$ is generated as discussed in Experiment 2. $\| x_\varepsilon^* \|_1$ corresponding to $\varepsilon = 0$ and $\varepsilon = \frac{\| y \|_2}{2}$ are shown in Fig. 8(a) and Fig. 8(b), respectively. It can be seen from Fig. 8(a) that the total numbers of elements in $x_\varepsilon^* = [x_{1,\varepsilon}, \cdots, x_{n,\varepsilon}]$ that satisfy

$$| x_{i,\varepsilon} | \geq \frac{\| x_\varepsilon^* \|_1}{n} \tag{12}$$

is 16 for $\varepsilon = \frac{\| y \|_2}{2}$ and 29 for $\varepsilon = 0$. That means, the scarcity of $\| x_\varepsilon^* \|_1$ when $\varepsilon = \frac{\| y \|_2}{2}$ is much higher than that when $\varepsilon = 0$. On the other hand, it can be seen in Fig. 9 that $A x_\varepsilon^*$ retains the main underlying characteristics of y when $\varepsilon = \frac{\| y \|_2}{2}$. Hence, the proposed method for designing ε results in a good tradeoff between the scarcity of the signal representation and the reconstruction error. For image coding, the compression ratio is dependent on the scarcity of the coefficients and the reconstruction error is dependent on the peak signal to noise ratio. In order to achieve a high coding gain, a good tradeoff between the scarcity of the coefficients and the reconstruction error is required. This problem can be approximated by the optimization problem with an L_1 norm objective function subject to the L_2 constraint. Hence, our proposed method could provide a good solution for this application.

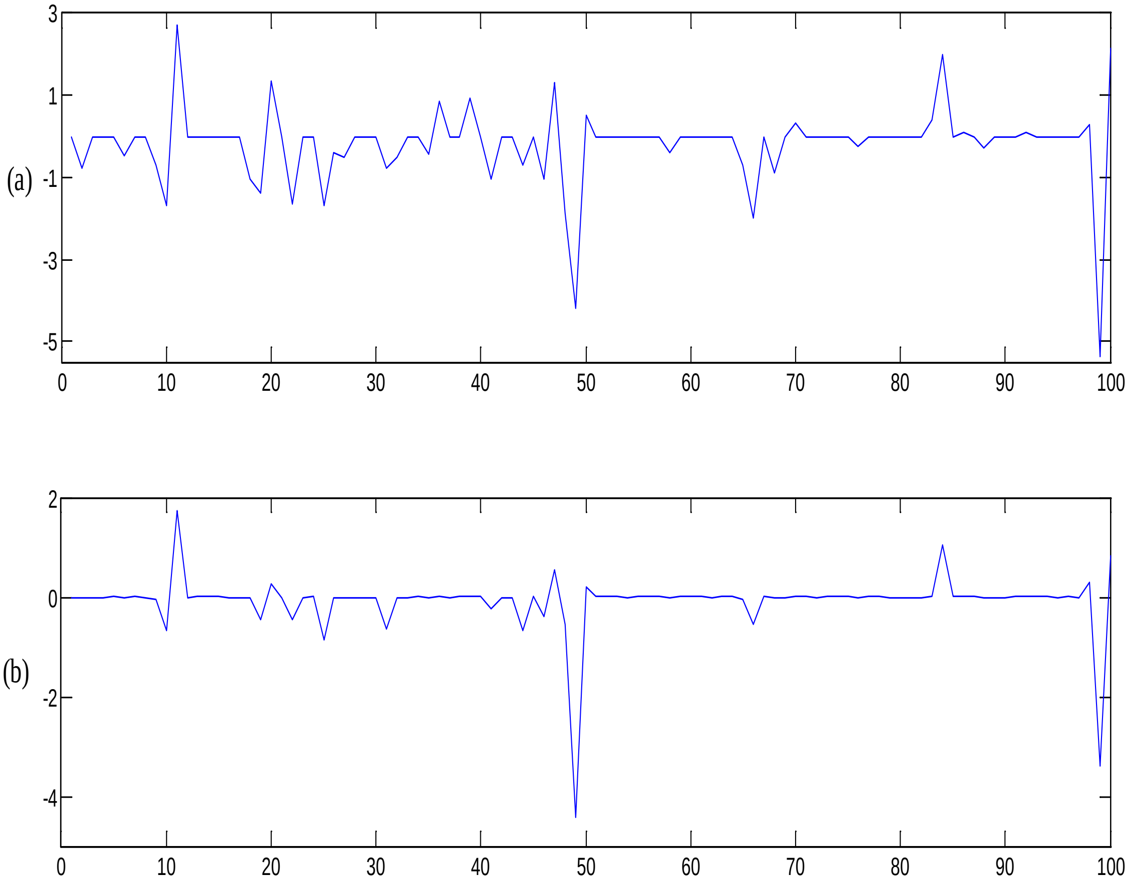

Fig. 8. (a)-(b): x_{ε}^{*} for $\varepsilon = 0$ and $\frac{\|y\|_2}{2}$, respectively.

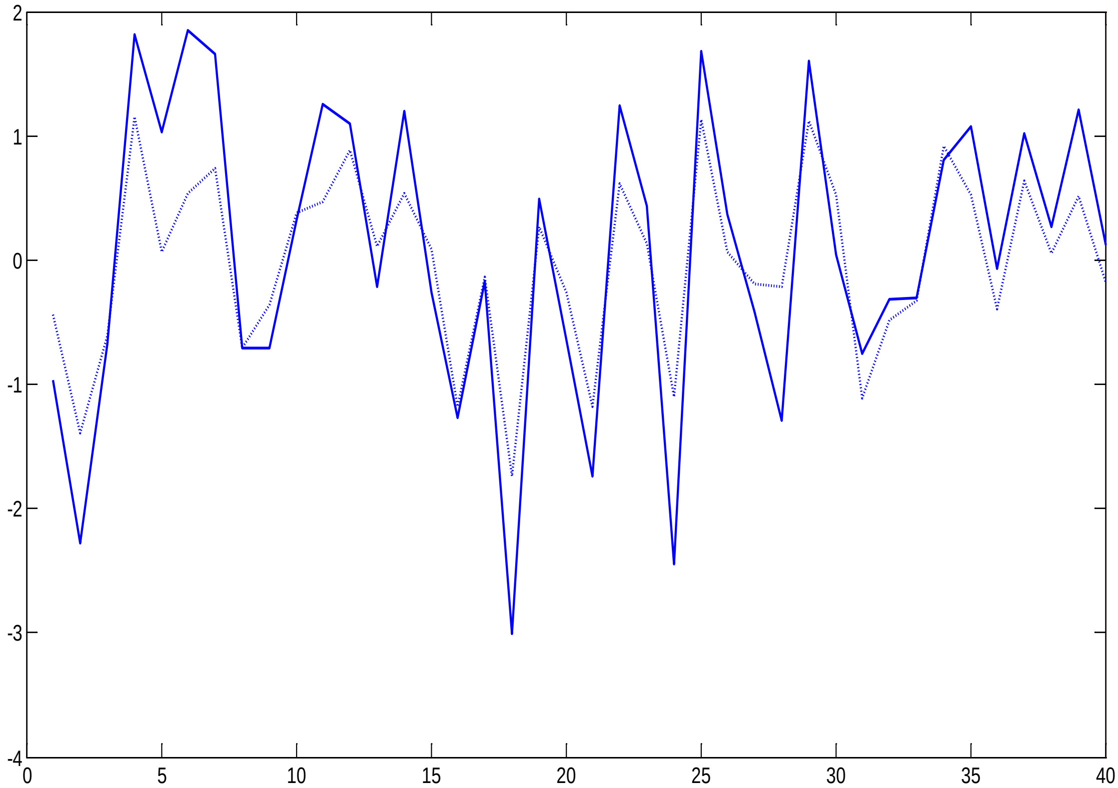

Fig. 9. y (solid line) and Ax_{ε}^{*} (dot line) for $\varepsilon = \frac{\|y\|_2}{2}$.

4. Conclusions

The main contributions of this paper are to study the relationship between the L_1 norm objective functional values and the L_2 norm specifications of sparse optimization problems. By making some assumptions and approximations, the proof of this property is given. To verify the validity of the assumptions and the approximations, three types of experiments are conducted. The first type of experiments is based on the signals satisfying the condition in (3). The second type of experiments is based on the randomly signals in which the condition in (3) is not satisfied. Finally, the third type of experiments is based on signals from experimental measurements. It is found that the relative L_2 norm errors for the first type of experiments, the second type of experiments and the third type of experiments are around 2.3%, 2.6% and 0.2%, respectively. This demonstrates that the relationship between the L_1 norm objective functional values and the L_2 norm specifications of sparse optimization problems is approximately linear. The obtained results can be employed for 1) estimating the L_1 norm of the optimal solution without recourse to numerical algorithms; 2) providing an insight for defining the L_2 norm specification; and 3) testing whether the obtained solutions are the globally optimal solutions or not.

Appendix A

The detail explanations of the near affine linear relationship between the L_1 norm objective functional values and the L_2 norm specifications is as follows.

Let $x_1^\bullet \in \Re^m$ and $x_2^\bullet \in \Re^{n-m}$ be the subvectors of x_ε^* such that $x_\varepsilon^* = \begin{bmatrix} x_1^{\bullet T} & x_2^{\bullet T} \end{bmatrix}^T$. Let $A_1 \in \Re^{m \times m}$ and $A_2 \in \Re^{m \times (n-m)}$ be the submatrices of A such that $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$. Here, we assume that $\|y\|_2 > \varepsilon$. Since the globally optimal solution of the corresponding unconstrained optimization problem is the zero vector in which it is not in the feasible set of the above optimization problem because of $\|y\|_2 > \varepsilon$, its globally optimal solution should be on the boundary of its feasible set because of the convexity nature of the optimization problem. Hence, we can define u^* as a unit energy vector such that $Ax_\varepsilon^* - y = \varepsilon u^*$. Suppose that A_1 is a full rank matrix. Then, $Ax_\varepsilon^* - y = A_1 x_1^\bullet + A_2 x_2^\bullet - y = \varepsilon u^*$ implies that $x_1^\bullet = A_1^{-1}\left(\varepsilon u^* + y - A_2 x_2^\bullet\right)$. Since

$$\|x_\varepsilon^*\|_1 = \|x_1^\bullet\|_1 + \|x_2^\bullet\|_1 = \left\|A_1^{-1}\left(\varepsilon u^* + y - A_2 x_2^\bullet\right)\right\|_1 + \|x_2^\bullet\|_1 = \left\|\begin{bmatrix} A_1^{-1}\varepsilon & -A_1^{-1}A_2 \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} + A_1^{-1}y\right\|_1 + \left\|\begin{bmatrix} 0 & I \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix}\right\|_1$$

$$= \left\|\begin{bmatrix} \begin{bmatrix} A_1^{-1}\varepsilon & -A_1^{-1}A_2 \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} + A_1^{-1}y \\ \begin{bmatrix} 0 & I \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} \end{bmatrix}\right\|_1 = \left\|\begin{bmatrix} A_1^{-1}\varepsilon & -A_1^{-1}A_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} + \begin{bmatrix} A_1^{-1}y \\ 0 \end{bmatrix}\right\|_1,$$

the original optimization problem is equivalent to the following optimization problem:

$$\min_{(u^*, x_2^\bullet)} \left\|\begin{bmatrix} A_1^{-1}\varepsilon & -A_1^{-1}A_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} + \begin{bmatrix} A_1^{-1}y \\ 0 \end{bmatrix}\right\|_1, \text{ subject to } \|u^*\|_2^2 = 1.$$

Since $\begin{bmatrix} A_1^{-1}\varepsilon & -A_1^{-1}A_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} + \begin{bmatrix} A_1^{-1}y \\ 0 \end{bmatrix}$ is sparse, we can group the small elements in $\begin{bmatrix} A_1^{-1}\varepsilon & -A_1^{-1}A_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} + \begin{bmatrix} A_1^{-1}y \\ 0 \end{bmatrix}$ together. In fact, this is a row operation. Hence, there exists a matrix $R=\begin{bmatrix}R_1 & R_2\\ R_3 & R_4\end{bmatrix}$ such that $\begin{bmatrix}R_1 & R_2\\ R_3 & R_4\end{bmatrix}\left(\begin{bmatrix}A_1^{-1}\varepsilon & -A_1^{-1}A_2\\ 0 & I\end{bmatrix}\begin{bmatrix}u^*\\ x_2^\bullet\end{bmatrix}+\begin{bmatrix}A_1^{-1}y\\ 0\end{bmatrix}\right)=\begin{bmatrix}z_1\\ z_2\end{bmatrix}$, where z_1 is not sparse but z_2 is small. That is, $z_2\approx 0$. Therefore, we have the following approximated optimization problem:

$$\min_{(u^*,x_2^\bullet)} \|z_1\|_1, \text{ subject to } z_2=0 \text{ and } \|u^*\|_2^2=1.$$

It is worth noting that

$$\begin{bmatrix}R_1 & R_2\\ R_3 & R_4\end{bmatrix}\left(\begin{bmatrix}A_1^{-1}\varepsilon & -A_1^{-1}A_2\\ 0 & I\end{bmatrix}\begin{bmatrix}u^*\\ x_2^\bullet\end{bmatrix}+\begin{bmatrix}A_1^{-1}y\\ 0\end{bmatrix}\right)=\begin{bmatrix}R_1A_1^{-1}\varepsilon & R_2-R_1A_1^{-1}A_2\\ R_3A_1^{-1}\varepsilon & R_4-R_3A_1^{-1}A_2\end{bmatrix}\begin{bmatrix}u^*\\ x_2^\bullet\end{bmatrix}+\begin{bmatrix}R_1A_1^{-1}y\\ R_3A_1^{-1}y\end{bmatrix}$$
$$=\begin{bmatrix}R_1A_1^{-1}\varepsilon u^*+\left(R_2-R_1A_1^{-1}A_2\right)x_2^\bullet+R_1A_1^{-1}y\\ R_3A_1^{-1}\varepsilon u^*+\left(R_4-R_3A_1^{-1}A_2\right)x_2^\bullet+R_3A_1^{-1}y\end{bmatrix}=\begin{bmatrix}z_1\\ z_2\end{bmatrix}.$$

Assume that $R_4-R_3A_1^{-1}A_2$ is a full rank matrix. Then, $z_2=R_3A_1^{-1}\varepsilon u^*+\left(R_4-R_3A_1^{-1}A_2\right)x_2^\bullet+R_3A_1^{-1}y=0$ implies that

$$x_2^\bullet=-\left(R_4-R_3A_1^{-1}A_2\right)^{-1}\left(R_3A_1^{-1}\varepsilon u^*+R_3A_1^{-1}y\right)$$

and

$$\begin{aligned}z_1&=R_1A_1^{-1}\varepsilon u^*+\left(R_2-R_1A_1^{-1}A_2\right)x_2^\bullet+R_1A_1^{-1}y\\ &=R_1A_1^{-1}\varepsilon u^*-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}\left(R_3A_1^{-1}\varepsilon u^*+R_3A_1^{-1}y\right)+R_1A_1^{-1}y\\ &=\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\varepsilon u^*+\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\end{aligned}.$$

Define $\iota=sign(z_1)$. Now, the above optimization problem is equivalent to the following optimization problem:

$$\min_{u^*} \quad \iota^T\left(\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\varepsilon u^*+\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\right),$$

subject to $\|u^*\|_2^2=1$.

It is worth noting that in general ι, R_1, R_2, R_3 and R_4 are functions of u^*. However, as z_1 is not sparse, ι remains unchanged within a small neighborhood of ε. Hence, we can treat ι as a constant vector within this small neighborhood of ε. Similarly, we can also assume that the locations of the sparse elements in $\left[z_1^T \quad z_2^T\right]^T$ remain unchanged within this small neighborhood of ε too. As a result, we have the same row operations. That is, R_1, R_2, R_3 and R_4 remain unchanged and they can be treated as constant matrices within the small neighborhood of ε. Let λ be the Lagrange multiplier. Define the corresponding Lagrange function as

$$\begin{aligned}L(u^*,\lambda)&=\iota^T\left(\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\varepsilon u^*+\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\right)+\lambda\left(u^{*T}u^*-1\right)\\ &=\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\varepsilon u^*+\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y+\lambda\left(u^{*T}u^*-1\right)\end{aligned}.$$

Then, we have $\frac{\partial}{\partial u^*}L(u^*,\lambda)=\varepsilon\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota+2\lambda u^*$ and $\frac{\partial}{\partial\lambda}L(u^*,\lambda)=u^{*T}u^*-1$. This implies that

$$u^*=-\frac{\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}$$

and

$$\lambda=\frac{\varepsilon}{2}\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}.$$

In other words, we have

$$u^*=-\frac{\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}.$$

Since $x_2^\bullet=-\left(R_4-R_3A_1^{-1}A_2\right)^{-1}\left(R_3A_1^{-1}\varepsilon u^*+R_3A_1^{-1}y\right)$, we have

$$x_2^\bullet=\frac{\varepsilon\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}-\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y.$$

As $x_1^\bullet=A_1^{-1}\left(\varepsilon u^*+y-A_2x_2^\bullet\right)$, we have

$$\begin{aligned}x_1^\bullet&=A_1^{-1}\varepsilon u^*+A_1^{-1}y-A_1^{-1}A_2x_2^\bullet\\&=\frac{-\varepsilon\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}.\\&+\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\end{aligned}$$

Therefore, we have the analytical form of the approximate solution as

$$x_\varepsilon^*=\begin{bmatrix}\frac{-\varepsilon\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}+\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\\\frac{\varepsilon\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}-\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y\end{bmatrix}.$$

If

$$\begin{bmatrix}\frac{-\varepsilon\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}\\\frac{\varepsilon\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}\end{bmatrix}$$

and

$$\begin{bmatrix}\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\\-\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y\end{bmatrix}$$

are in the same direction,
then

$$\begin{aligned}x_\varepsilon^*&=\left\|\begin{bmatrix}\frac{-\varepsilon\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}+\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\\\frac{\varepsilon\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}-\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y\end{bmatrix}\right\|_1.\\&\approx\left\|\begin{bmatrix}\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}y\\-\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}y\end{bmatrix}\right\|_1+\varepsilon\left\|\begin{bmatrix}\frac{-\left(I+A_1^{-1}A_2\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}\\\frac{\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}{\sqrt{\iota^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)A_1^{-1}\left(A_1^{-1}\right)^T\left(R_1-\left(R_2-R_1A_1^{-1}A_2\right)\left(R_4-R_3A_1^{-1}A_2\right)^{-1}R_3\right)^T\iota}}\end{bmatrix}\right\|_1\end{aligned}$$

As a result, $\left\|x_\varepsilon^*\right\|_1$ and ε has an approximately affine linear relationship. This completes the proof. ■

Appendix B

In Appendix A, it only requires to find the inverses of A_1 and $R_4-R_3A_1^{-1}A_2$. For the invertibility of $R_4-R_3A_1^{-1}A_2$, let z_1 be a p dimensional vector. That is, $z_1\in\Re^{p\times1}$. Then, $z_2\in\Re^{(n-p)\times1}$, $R_1\in\Re^{p\times m}$, $R_2\in\Re^{p\times(n-m)}$, $R_3\in\Re^{(n-p)\times m}$ and $R_4\in\Re^{(n-p)\times(n-m)}$. In this case,

$R_4 - R_3 A_1^{-1} A_2 \in \Re^{(n-p)\times(n-m)}$. Since $A \in \Re^{m\times n}$, $Ax_\varepsilon^* - y = A_1 x_1^\bullet + A_2 x_2^\bullet - y = \varepsilon u^*$ and ε is usually very small, the total sparsity of $x_1^\bullet$ and $x_2^\bullet$ is at least $n-m$. This implies that the total number of nonzero coefficients in this vector $\begin{bmatrix} A_1^{-1}\varepsilon & -A_1^{-1}A_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} u^* \\ x_2^\bullet \end{bmatrix} + \begin{bmatrix} A_1^{-1}y \\ 0 \end{bmatrix}$ is not more than m. In other words, $p \le m$. In general, $z_2 = R_3 A_1^{-1}\varepsilon u^* + \left(R_4 - R_3 A_1^{-1} A_2\right) x_2^\bullet + R_3 A_1^{-1} y = 0$ refers to a system of $n-p$ linear equations with m variables in u^* and $n-m$ variables in $x_2^\bullet$. From here, we can see that the most common situation is the case when $p=m$ and $R_4 - R_3 A_1^{-1} A_2$ being invertible. There are only two different cases when $R_4 - R_3 A_1^{-1} A_2$ is not invertible. Consider the first case when $p<m$. In this case, the total number of linear equations ($z_2 = R_3 A_1^{-1}\varepsilon u^* + \left(R_4 - R_3 A_1^{-1} A_2\right) x_2^\bullet + R_3 A_1^{-1} y = 0$) is more than the total number of variables in $x_2^\bullet$. Here, $m-p$ variables in u^* are chosen as the dependent variables in order to make the linear equations to be feasible. In other words, we can still write the form $x_2^\bullet = \mathbf{T}u^* + q$ with $m-p$ dependent variables in u^* and the rest of the proof is still valid. Finally, consider the last case when $p=m$. In this case, the total number of linear equations is the same as the total number of variables in $x_2^\bullet$, but some equations are linear dependent. Suppose that there are s linear dependent equations. Here, s variables in u^* are chosen as the dependent variables in order to make the linear equations to be feasible. In other words, we can still write the form $x_2^\bullet = \tilde{\mathbf{T}}u^* + \tilde{q}$ with s dependent variables in u^* and the rest of the proof is still valid.

For the invertibility of A_1, since $A \in \Re^{m\times n}$, $A_1 \in \Re^{m\times m}$, $A_2 \in \Re^{m\times(n-m)}$, $x_1^\bullet \in \Re^{m\times 1}$ and $x_2^\bullet \in \Re^{(n-m)\times 1}$, $Ax_\varepsilon^* - y = A_1 x_1^\bullet + A_2 x_2^\bullet - y = 0$ refers to a system of m linear equations with m variables in $x_1^\bullet$ and $n-m$ variables in $x_2^\bullet$. A_1 being not invertible implies that there are less than m linear independent equations. As we do not have any degree of freedoms on choosing the variables in y, in general these equations are not feasible.

Acknowledgements

This work was supported in part by the National Nature Science Foundation of China (Nos. 61372173 and 61471132), the Guangdong Higher Education Engineering Technology Research Center for Big Data on Manufacturing Knowledge Patent (no. 501130144), the Hundred People Plan from the Guangdong University of Technology and the Young Thousand People Plan from the Ministry of Education of China.

References

[1] Minfen Shen, Qiong Zhang, Delai Li, Jinyao Yang and Bin Li, "Adaptive Sparse Representation Beamformer for High-Frame-Rate Ultrasound Imaging Instrument," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 5, pp. 1323-1333, 2012.

[2] Yijiu Zhao, Yu Hen Hu and Houjun Wang, "Enhanced Random Equivalent Sampling Based on Compressed Sensing," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 3, pp. 579-586, 2012.

[3] Zhijing Yang, Bingo Wing-Kuen Ling and Chris Bingham, "Joint Empirical Mode Decomposition and Sparse Binary Programming for Underlying Trend Extraction," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 10, pp. 2673-2682, 2013.

[4] Yuanqing Li and Shun-ichi Amari, “Two conditions for equivalence of 0-norm solution and 1-norm solution in sparse representation,” *IEEE Transactions on Neural Networks*, vol. 21, no. 7, pp. 1189-1196, 2010.
[5] Jean-Jacques Fuchs, “On sparse representations in arbitrary redundant bases,” *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1341-1344, 2004.
[6] Michael Elad and Alfred M. Bruckstein, “A generalized uncertainty principle and sparse representation in pairs of bases,” *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558-2567, 2002.
[7] David L. Donoho and Xiaoming Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845-2862, 2001.
[8] Leming Qu, Partha S. Routh and Phil D. Anno, “Wavelet reconstruction of nonuniformly sampled signals,” *IEEE Signal Processing Letters*, vol. 16, no. 2, pp. 73-76, 2009.
[9] David L. Donoho and Yaakov Tsaig, “Fast solution of l_1-norm minimization problems when the solution may be sparse,” *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4789-4812, 2008.
[10] Bingo Wing-Kuen Ling, Charlotte Yuk-Fan Ho, Ted Chi-Wah Kok, Wan-Chi Siu and Kok-Lay Teo, “Empirical formula for designing a class of linear phase FIR single band PCLS filters,” *Digital Signal Processing*, vol. 23, no. 1, pp. 470-477, 2013.
[11] Jianwei Ma, “Compressed sensing for surface characterization and metrology,” *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 6, pp. 1600-1615, 2010.
[12] Arash Tabibiazar and Otman Basir, “Energy-efficient compressive state recovery from sparsely noisy measurements,” *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 9, pp. 2392-2400, 2012.
[13] Charles Nader, Wendy Van Moer, Niclas Björsell, Kurt Barbé and Peter Händel, “Reducing the analog and digital bandwidth requirements of RF receivers for measuring periodic sparse waveforms,” *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 11, pp. 2960-2971, 2012.
[14] Emmanuel J. Candès, Michael B. Wakin and Stephen P. Boyd, “Enhancing sparsity by reweighting l_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 41, pp. 877-905, 2008.