A wonderful triangle in compressed sensing

Jun Wang[a,*]

[a]*School of Science, Jiangsu University of Science and Technology, Zhenjiang* 212003*, China*

Abstract

In order to determine the sparse approximation function which has a direct metric relationship with the ℓ_0 quasi-norm, we introduce a wonderful triangle whose sides are composed of $\|\mathbf{x}\|_0$, $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_\infty$ for any non-zero vector $\mathbf{x} \in \mathbb{R}^n$ by delving into the iterative soft-thresholding operator in this paper. Based on this triangle, we deduce the ratio ℓ_1 and ℓ_∞ norms as a sparsity-promoting objective function for sparse signal reconstruction and also try to give the sparsity interval of the signal. Considering the ℓ_1/ℓ_∞ minimization from a angle β of the triangle corresponding to the side whose length is $\|\mathbf{x}\|_\infty - \|\mathbf{x}\|_1/\|\mathbf{x}\|_0$, we finally demonstrate the performance of existing ℓ_1/ℓ_∞ algorithm by comparing it with ℓ_1/ℓ_2 algorithm.

Keywords: Compressed sensing, sparsity-promoting, the ℓ_0 minimization, iterative soft-thresholding operator, nonconvex optimization, sparse signal reconstruction, the ℓ_1/ℓ_∞ algorithm.

2010 MSC: 68W40, 65K10, 90C25, 90C51, 90C30

1. Introduction

Compressed Sensing (CS) [1, 2] is mostly known for finding the exact or approximate sparsest solutions for underdetermined linear systems of equations $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$ for $m \ll n$. CS arises in a wide range of applications such as signal and image processing [3], single pixel

*Corresponding author
Email address: jwang77@just.edu.cn (Jun Wang)

camera [4], statistics [5] and machine learning [6], etc. Mathematically, the core problem of CS can be formulated by the ℓ_0 minimization [7, 8]

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{A}\mathbf{x}=\mathbf{b}, \tag{1.1}$$

or its ℓ_0 regularized problem [9, 10]

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2+\lambda\|\mathbf{x}\|_0, \tag{1.2}$$

where $\|\mathbf{x}\|_0$ measures the number of nonzero components in $\mathbf{x}$ and $\lambda>0$.

Under the assumption that the signal is sparse, Candés, Romberg and Tao [11] and Donoho [12] provided a breakthrough work that $S_1(\mathbf{x})=\|\mathbf{x}\|_1$ can replace $\|\mathbf{x}\|_0$ of problem (1.1) and (1.2), which called be BP and BPDN, when the sensing matrix $\mathbf{A}$ satisfies some condition including the restricted isometry property (RIP) [13] or null space property (NSP) [14].

Over the next decade, many researchers have been continuously finding new sparse alternative functions. Based on the formula

$$\|\mathbf{x}\|_0=|\operatorname{supp}(\mathbf{x})|=\sum_{i\in\operatorname{supp}(\mathbf{x})}\frac{|x_i|}{|x_i|}\approx\sum_{i=1}^{n}\frac{|x_i|}{|x_i|+\epsilon},$$

Boyd et al. [15], Zhao et al. [16] and the authors [17, 18] have replaced $\|\mathbf{x}\|_0$ with the surrogate functions $S_2(\mathbf{x})=\sum_{i=1}^n\bar{w}_i|x_i|$ and $S_3(\mathbf{x})=\sum_{i=1}^n\widehat{w}_i x_i^2$ with the positive weights $\bar{w}_i$ and $\widehat{w}_i$, respectively. Taking into account the following

$$\|\mathbf{x}\|_0=\lim_{p\to0}\|\mathbf{x}\|_p^p=\lim_{p\to0}\sum_{i=1}^{n}|x_i|^p,$$

some authors [19, 20, 21, 22] supersede $\|\mathbf{x}\|_0$ by the ℓ_p quasi-norm $S_4(\mathbf{x})=\|\mathbf{x}\|_p^p$ with the parameter $0<p<1$. Recently, many researchers have started to consider mixed norms as an alternative function to the ℓ_0 quasi-norm. The authors [23, 24, 25, 26, 27] considered the difference and the rate of ℓ_1 and ℓ_2 norms, i.e. $S_5(\mathbf{x})=\|\mathbf{x}\|_1-\|\mathbf{x}\|_2$ and $S_6(\mathbf{x})=\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$, respectively. Wang [28, 29] considered $S_7(\mathbf{x})=\|\mathbf{x}\|_1-\|\mathbf{x}\|_\infty$ and $S_8(\mathbf{x})=\sum_{i\in\mathbb{S}}1+\sum_{i\in\mathbb{S}^c}|x_i|$, where $\|\mathbf{x}\|_\infty\triangleq\max_{1\le i\le n}|x_i|$ and the support $\mathbb{S}\subseteq\mathbb{N}:=\{1,2,\ldots,n\}$ is estimated iteratively and $\mathbb{S}^c$ denotes the complement of $\mathbb{S}$ with respect to $\mathbb{N}$.

However, there are two notable key points here from the above researches and corresponding conclusions.

(1) The original signal is required to be sparse.

(2) These surrogate functions $S_i(\mathbf{x}), i = 1, 2, \ldots, 9$ can replace $\|\mathbf{x}\|_0$ for sparse reconstructions.

For these two key points, we have been confused with doubts (i) and (ii) related to (1) and (2), respectively. Next, we will show our specific ideas in detail.

To the best of our knowledge, we first noticed that in the numerical simulations about successful recovery rate, the sparsity tested does not generally exceed 50. For instance, Xu et al. [26] demonstrated that the maximum sparsity of recovery rate is 24 with the reason that the successful recovery rate is 0 after 25. Although we have known that if the dimension n and sparsity s of the signal are given in advance, then the number of measurements m satisfying $m \geq Cs\log(n)$ [30] for some constant C, we are still confused at the following question:

(i) for given dimension n of some signal $\mathbf{x}^o \in \mathbb{R}^n$, what value the sparsity $s = \|\mathbf{x}^o\|_0 (s \ll n)$ should be set to claim that the signal $\mathbf{x}^o$ is sparse?

Specifically, assume that the dimension and sparsity of some signal is $n = 500$ and $s = 60$ respectively, then can we call this signal is sparse ?

Suppose that the signal is sparse, although there are many algorithms that handle the models (1.1) or (1.2), we still hope to find a computable objective function that yields a better approximate solutions without the additional conditions on the sensing matrix $\mathbf{A}$. In addition, most of the above objective functions considered are related to the ℓ_p $(0 < p \leq \infty)$ norm. Thus, a natural question is that

(ii) is there any other quantitative relationship between ℓ_0 and ℓ_p $(0 < p \leq \infty)$?

These two doubts (i) and (ii) have always surrounded the CS theory, so that we are unconsciously thinking about how explain them to have a clear understanding. The contributions of the paper are three-fold:

(1) We demonstrate that a signal is called to be sparse when its sparsity is in a certain interval $(0, 50]$.

(2) We reveal the connection of the rate of ℓ_1 and ℓ_∞ to the ℓ_0 quasi-norm. As the sparsity $s \geq 5$ of the signal increases, we theoretically give a metric relationship between $\|\mathbf{x}\|_1/\|\mathbf{x}\|_\infty$ and $\|\mathbf{x}\|_0$, i.e.

$$\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} \approx \|\mathbf{x}\|_0.$$

And if the sparsity $s \leq 5$, we can get very good results by directly using the BP or BPDN, because that the more sparse the signal, the better the recovery of these two models. Specially, it is noteworthy that this idea is new in CS.

(3) We construct an iterative DCA with linearization and our empirical results shed light about the effects of sparse recovery.

The rest of the paper is organized as follows. We try to give some heuristic answers that do not completely solve these two questions, but at least give us a beginning of understanding by using the ℓ_1 norm of the iterative soft-thresholding operator in Section 2. To verify the validity of problem (2.7), we construct an algorithm to test whether the sparse solution can be recovered by imitating the literature [24] in Section 3. In Section 4, we summarize our findings as well as point out some possible directions for future investigation.

2. Main results

Before continuing, we provide here some notations used throughout the paper. We use $\mathbb{R}^n$ to denote the n-dimensional Euclidean space. Matrices are bold capital, vectors are bold lower-cases, and scalars or entries are not bold. For instance, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$ is a vector and x_i its i-th component. For

any $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$ is their inner product. The Euclidean norm of $\mathbf{x}$ is denoted by $\|\mathbf{x}\|_2$. The ℓ_1 and ℓ_∞ norm of $\mathbf{x}$ is denoted by $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_\infty$, respectively.

The shrinkage operator is the most critical core component of ISTA/FISTA [31, 32] for recovering the sparse solutions. For any non-zero vector $\mathbf{y} \in \mathbb{R}^n$, we have that the shrinkage operator

$$\mathbf{S}_\sigma(\mathbf{y}) := \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \ \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \sigma \|\mathbf{x}\|_1, \tag{2.1}$$

where $\sigma \geq 0$. The non-negative real number σ adjusts the sparsity of the recovered solutions. Hence, it is crucial to set the value of the parameter σ, which forces us to do more in-depth exploration for the shrinkage operator $\mathbf{S}_\sigma(\mathbf{y})$. From (2.1), we obtain

$$\mathbf{S}_\sigma(\mathbf{y}) = \begin{cases} \mathbf{y}, \ \sigma = 0 \\ \mathbf{0}, \ \sigma \geq \sigma_{\max} \triangleq \|\mathbf{y}\|_\infty, \end{cases}$$

and therefore require that $\sigma \in [0, \sigma_{\max}]$ for making the shrinkage operator meaningful.

Define $\varphi : [0, \sigma_{\max}] \to [0, \|\mathbf{y}\|_1]$ by

$$\varphi(\sigma) := \|\mathbf{S}_\sigma(\mathbf{y})\|_1. \tag{2.2}$$

Fornasier [33, Lemma 4.12] showed that $\varphi(\sigma)$ is a piecewise linear, continuous, strictly decreasing function of σ. Here, we will further enrich these results about $\varphi(\sigma)$ in Proposition 2.1 which is deduced from [18, Lemma 2.1, Lemma 2.3].

Proposition 2.1. *For any given nonzero vector $\mathbf{y} \in \mathbb{R}^n$, let $\varphi(\sigma)$ be defined by (2.2). Then*

(*i*) $\varphi(0) = \|\mathbf{y}\|_1$ *and* $\varphi(\sigma_{\max}) = 0$;

(*ii*) $\varphi(\sigma)$ *is a strictly monotonically decreasing and continuous convex function of* σ;

(*iii*) *The right-hand derivative of* $\varphi(\sigma)$ *with respect to* σ *is that*

$$\varphi_r^{'}(\sigma) = -\|\boldsymbol{S}_\sigma(\mathbf{y})\|_0$$

with two special cases of $\varphi_r^{'}(0) = -\|\mathbf{y}\|_0$ *and* $\varphi_r^{'}(\sigma_{\max}) = 0$.

Proof. Using the definition of $\mathbf{S}_\sigma(\mathbf{y})$ in (2.1), it is easy to check that (i) is valid. We then obtain from [18, Lemma 2.1, Lemma 2.3] that (ii) holds.

Now, it follows from [18, Lemma 2.1] that, for any $\epsilon \geq 0$,

$$\epsilon\|\mathbf{S}_{\sigma+\epsilon}(\mathbf{y})\|_0 \leq \varphi(\sigma) - \varphi(\sigma+\epsilon) \leq \epsilon\|\mathbf{S}_\sigma(\mathbf{y})\|_0,$$

and therefore

$$\|\mathbf{S}_{\sigma+\epsilon}(\mathbf{y})\|_0 \leq \frac{\varphi(\sigma) - \varphi(\sigma+\epsilon)}{\epsilon} \leq \|\mathbf{S}_\sigma(\mathbf{y})\|_0,$$

which educes that

$$\varphi_{\mathrm{r}}^{'}(\sigma) = \lim_{\epsilon\to+0}\frac{\varphi(\sigma+\epsilon) - \varphi(\sigma)}{\epsilon} = -\|\mathbf{S}_\sigma(\mathbf{y})\|_0.$$

Finally, if $\sigma = 0$ or $\sigma = \sigma_{\max}$, then we can directly obtain $\varphi_{\mathrm{r}}^{'}(0) = -\|\mathbf{y}\|_0$ and $\varphi_{\mathrm{r}}^{'}(\sigma_{max}) = 0$, respectively. □

Remark 2.1. *Proposition 2.1 extends to a better result that* $\varphi(\sigma)$ *is a convex function and obtains the right derivative exactly with respect to the* ℓ_0 *norm. This may give us more insight into other related findings.*

(1) *We obtain from the fact* $\varphi(\sigma)$ *is a convex function that [18, Theorem 2.5]*

$$\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty} \leq \frac{\|\mathbf{y}\|_1 - \|\boldsymbol{S}_\sigma(\mathbf{y})\|_1}{\sigma} \leq \|\mathbf{y}\|_0,\ \sigma \in (0, \sigma_{\max}];$$

(2) *Proposition 2.1 shows that how to set the value of the parameter* σ *in ISTA/FISTA [31, 32]. In general, we obtain from* $\varphi_r^{'}(0) = -\|\mathbf{y}\|_0$ *that the parameter* σ *is very small, such as,* $\sigma = 10^{-4}$. *In fact, the sparsity* $\|\boldsymbol{S}_\sigma(\mathbf{y})\|_0$ *of the next iteration* $\boldsymbol{S}_\sigma(\mathbf{y})$ *does show a stepwise decrease as the parameter* σ *is gradually increased from* 0 *to the maximum* $\sigma_{\max}$. *The typical dynamics of the sparsity* $\|\boldsymbol{S}_\sigma(\mathbf{y})\|_0$ *are illustrated in Figure 1. In*

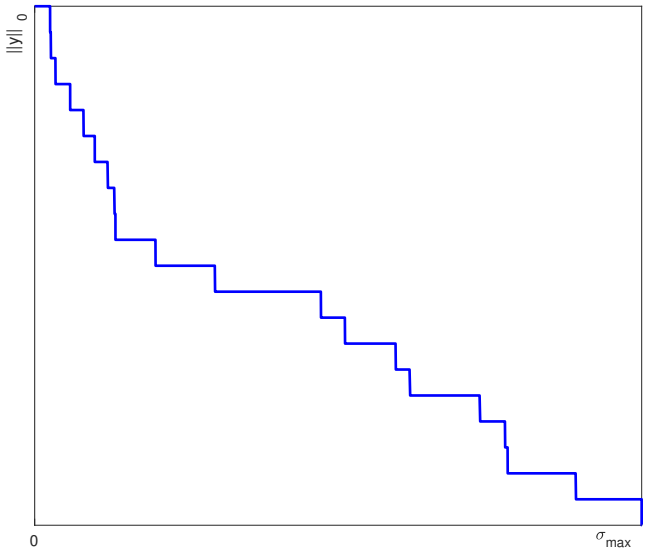


Figure 1: The echelon form of $\|\mathbf{S}_\sigma(\mathbf{y})\|_0$ for any given non-zero vector $\mathbf{y} \in \mathbb{R}^n$ with the parameter $\sigma \in [0, \sigma_{\max}]$.

addition, if the parameter σ is taken too large, it can cause many more entries in $\boldsymbol{S}_\sigma(\mathbf{y})$ to be made zero. However, these indices corresponding to them maybe the non-zero part of the solution that we need to recover. Then, this will lead to a final solution that is not an approximate sparse solution of problem (1.1) *and* (1.2).

Next, we will use Proposition 2.1 to construct a wonderful triangle in CS. Firstly, we draw $\varphi(\sigma)$ in Cartesian coordinate system (σ, φ). A schematic illustration for Proposition 2.1 is given in Figure 2. Note that the coordinates of the three points A, B and C are $(0, \|\mathbf{y}\|_1)$, $\left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_0}, 0\right)$ and $(\|\mathbf{y}\|_\infty, 0)$, respectively. Then, the lengths of all three sides of ΔABC can be calculated by the followings

$$\begin{cases} |AB| &= \sqrt{\|\mathbf{y}\|_1^2 + \frac{\|\mathbf{y}\|_1^2}{\|\mathbf{y}\|_0^2}} > 0 \\ |AC| &= \sqrt{\|\mathbf{y}\|_1^2 + \|\mathbf{y}\|_\infty^2} > 0 \\ |BC| &= \|\mathbf{y}\|_\infty - \frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_0} \geq 0 \end{cases} \tag{2.3}$$

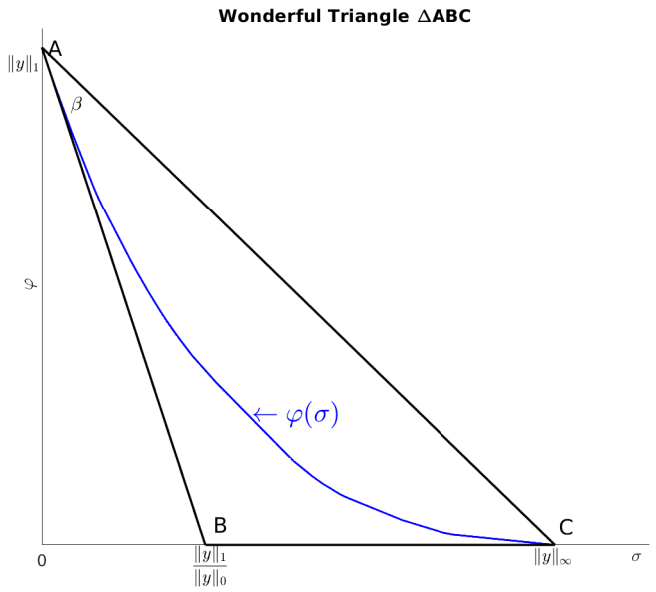


Figure 2: The wonderful triangle ΔABC formed by $\|\mathbf{y}\|_0$, $\|\mathbf{y}\|_1$ and $\|\mathbf{y}\|_\infty$ for any given non-zero vector $\mathbf{y} \in \mathbb{R}^n$. The side AB of ΔABC is a segment of the tangent line of the curve $\varphi(\sigma)$ at the vertex A.

for any n-dimensional nonzero vector $\mathbf{y}$.

Let $\beta = \angle$BAC, which is the angle corresponding to side BC of ΔABC, then we obtain from (2.3), the cosine rule and the law of sines that

$$\begin{cases} \sin\beta &= \dfrac{1-\frac{1}{\|\mathbf{y}\|_0}\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty}}{\sqrt{\left(1+\frac{1}{\|\mathbf{y}\|_0^2}\right)\left(1+\frac{\|\mathbf{y}\|_1^2}{\|\mathbf{y}\|_\infty^2}\right)}}, \\ \cos\beta &= \dfrac{\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty}+\frac{1}{\|\mathbf{y}\|_0}}{\sqrt{\left(1+\frac{1}{\|\mathbf{y}\|_0^2}\right)\left(1+\frac{\|\mathbf{y}\|_1^2}{\|\mathbf{y}\|_\infty^2}\right)}}, \\ \tan\beta &= \dfrac{1-\frac{1}{\|\mathbf{y}\|_0}\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty}}{\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty}+\frac{1}{\|\mathbf{y}\|_0}}. \end{cases} \tag{2.4}$$

It follows from

$$1 \le t \triangleq \frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty} \le s \triangleq \|\mathbf{y}\|_0 \le n \tag{2.5}$$

and (2.4) that

$$\begin{cases} \sin\beta &= \frac{s-t}{\sqrt{(1+s^2)(1+t^2)}} \in \left[0, \frac{1}{\sqrt{1+t^2}}\right], \\ \cos\beta &= \frac{1+st}{\sqrt{(1+s^2)(1+t^2)}} \in \left[\frac{t}{\sqrt{1+t^2}}, 1\right], \\ \tan\beta &= \frac{s-t}{1+st} \in \left[0, \frac{1}{t}\right]. \end{cases} \tag{2.6}$$

From (2.4), we can see that the dynamic parameter β is composed of $\|\mathbf{y}\|_0$, $\|\mathbf{y}\|_1$ and $\|\mathbf{y}\|_\infty$. As these three terms change, will this parameter be a very small immobile constant? If this is in case, then perhaps this parameter implies a certain quantitative relationship among $\|\mathbf{y}\|_0$, $\|\mathbf{y}\|_1$ and $\|\mathbf{y}\|_\infty$.

Next, we will demonstrate our interesting observation from the following two aspects.

2.1. $\tan\beta$ *approximates* $\sin\beta$ *in three dimensions*

Firstly, it is important to note that if $s = 1$, then it follows from (2.5) that $t = 1$ and if $t = 1$, then it follows from $\|\mathbf{y}\|_1 = \|\mathbf{y}\|_\infty$ and [28, Lemma 2.1] that $s = 1$. Then, we draw them in Maple[1] or Mathematica[2] with the condition that the sparsity s ranges from 2 to some given positive integer and t is set from $1.0 + 10^{-6}$ to s. Finally, these results are given in Figure 3, 4, 5 and 6

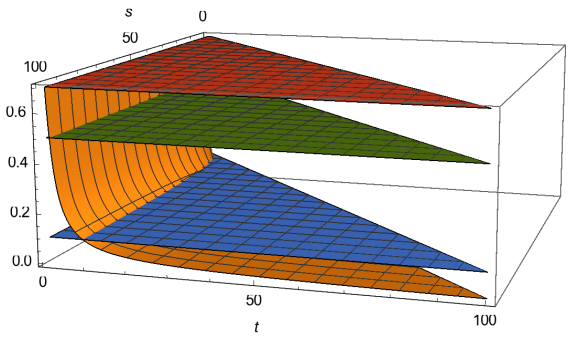


Figure 3: $\sin\beta$ is plotted with the sparsity s ranges from 2 to 100 and t is set from $1.0 + 10^{-6}$ to s. The top (red) plant with 0.7, middle (green) plant with 0.5 and bottom (blue) plant with 0.1.

[1]Maple: $plot3d((s-t)/sqrt((1+s*s)*(1+t*t), s = 2..100, t = 1.0000001..s)$

[2]Mathematica: $Plot3D[(s-t)/(Sqrt[1+t*t]*Sqrt[1+s*s]), \{s, 2, 100\}, \{t, 1.0000001, s\}]$

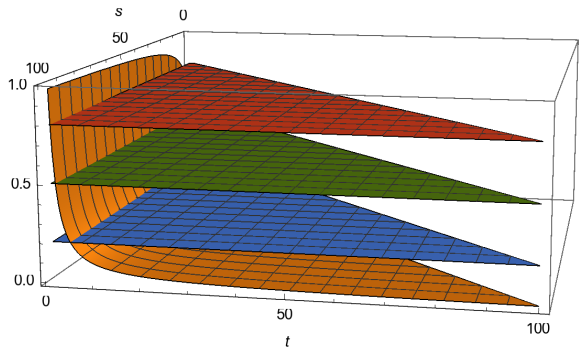


Figure 4: $\tan\beta$ is plotted with the sparsity s ranges from 2 to 100 and t is set from $1.0+10^{-6}$ to s. The top (red) plant with 0.8, middle (green) plant with 0.5 and bottom (blue) plant with 0.2.

It can seen from Figure 3 and 4 that $\sin\beta \approx \tan\beta \in [0, 0.2]$ as s and t increase. More specifically, Figure 5 shows that $\sin\beta \approx \beta \approx \tan\beta \in [0, 0.1]$ when $s \geq t \geq 5$. Then, we obtain from (2.4) that

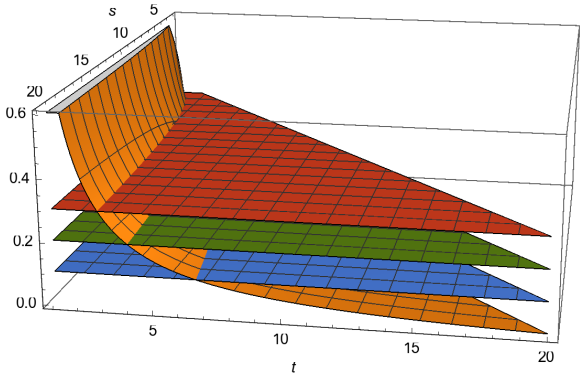


Figure 5: $\tan\beta$ is plotted with the sparsity s ranges from 2 to 20 and t is set from $1.0+10^{-6}$ to s. The top (red) plant with 0.3, middle (green) plant with 0.2 and bottom (blue) plant with 0.1.

$$\sqrt{\left(1+\frac{1}{\|\mathbf{y}\|_0^2}\right)\left(1+\frac{\|\mathbf{y}\|_1^2}{\|\mathbf{y}\|_\infty^2}\right)} \approx \frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty}+\frac{1}{\|\mathbf{y}\|_0}$$

which means that

$$\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty} \approx \|\mathbf{y}\|_0.$$

The above discussion may give us an idea that maybe we consider the ℓ_1/ℓ_∞ minimization

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.7}$$

or its regularized problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty}, \tag{2.8}$$

where the parameter $\lambda > 0$. Besides, we can obtain the same idea from Figure 6. In fact, when $s \geq t \geq 5$, we have that $\cos\beta \approx 1$ which means

$$s \approx t \Longleftrightarrow \frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty} \approx \|\mathbf{y}\|_0.$$

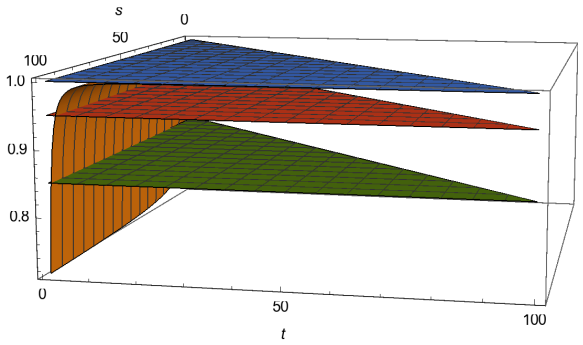


Figure 6: $\cos\beta$ is plotted with the sparsity s ranges from 2 to 100 and t is set from $1.0 + 10^{-6}$ to s. The middle (red) plant with 0.95, bottom (green) plant with 0.85 and top (blue) plant with 1.0.

2.2. Average $\tan\beta$ *about the sparsity in two dimensions*

From (2.3), it follows that

$$\|\mathbf{y}\|_0 = \frac{\tan\beta + \frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty}}{1 - \tan\beta\left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty}\right)}. \tag{2.9}$$

If there exits an approximate parameter $\beta_s := \tan\beta$ corresponding to each sparsity s, then using (2.9), we can construct an optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{\beta_s + \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty}}{1 - \beta_s\left(\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty}\right)} \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{2.10}$$

If $s = 1$, then $\beta_s = 0$. In this case, problem (2.10) reduces (2.7). Hence, we assume that $s \geq 2$ in subsequent analysis and discussion.

Dinkelbach [34] provided an iteration method whose key feature is the solution of a series of the following problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ \beta_s + \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} - \alpha_k \left(1 - \frac{\beta_s \|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty}\right) \ \text{s.t.} \ \mathbf{A}\mathbf{x} = \mathbf{b},$$

which is equivalent to

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ (1 + \alpha_k \beta_s) \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} \ \text{s.t.} \ \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.11}$$

where the parameter

$$\alpha_k = \frac{\beta_s + \frac{\|\mathbf{x}^k\|_1}{\|\mathbf{x}^k\|_\infty}}{1 - \beta_s \left(\frac{\|\mathbf{x}^k\|_1}{\|\mathbf{x}^k\|_\infty}\right)}.$$

It is easy to see that problem (2.11) and (2.7) have the same solution set. Hence, we can directly solve (2.7) instead of (2.10). Then how to determine the parameter β_s corresponding to each sparsity s?

There are two basic idea that arithmetic and geometric means that removing the variable t from $\tan\beta$ of (2.6). By (2.5) and (2.6), we define arithmetic mean

$$\beta_A(s) := \frac{1}{s-1} \int_1^s \frac{s-t}{1+st} \mathrm{d}t, \tag{2.12}$$

and geometric mean

$$\beta_G(s) := \exp\left(\frac{\int_1^s \ln\left(\frac{s-t}{1+st}\right) \mathrm{d}t}{s-1}\right), \tag{2.13}$$

with the special case $\beta_G(1) = 0$ and $\beta_A(1) = 0$. By the fact that the arithmetic mean is greater than the geometric mean, we have that $\beta_G(s) \leq \beta_A(s)$.

In Figure 7, we see that if the sparsity $s \geq 60$, then the means $\beta_G(s)$ and $\beta_A(s)$ slowly decrease. When $s = 100$, we have that $\beta_A(s) \approx 0.036422$ and $\beta_G(s) \approx 0.00944564$. If $s < 60$, $\beta_G(s)$ and $\beta_A(s)$ first increase to a maximum value and then decrease, followed by a gentle decrease as the sparsity s increases. In addition, the increase and decrease of $\beta_G(s)$ and $\beta_A(s)$ are very rapid when the sparsity $s \leq 10$, and we can also see this phenomenon as a manifestation of the sparsity. Please see more details about $\beta_G(s)$ and $\beta_A(s)$ in Figure 7.

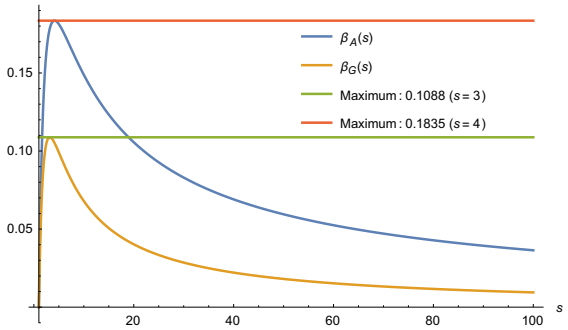


Figure 7: Arithmetic mean $\beta_A(s)$ and Geometric mean $\beta_G(s)$ with the sparsity s ranges from 1 to 100.

Next, we further demonstrate the confusion about (2.10) you may have encountered. It is because that $\|\mathbf{y}\|_0 \geq 1$. Therefore we obtain from (2.10) that

$$1 - \beta_s \left(\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} \right) > 0,$$

which means that $\frac{\|\mathbf{x}\|_\infty}{\|\mathbf{x}\|_1} > \beta_s$. In fact, problem (2.10) should be rewritten by

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{\beta_s + \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty}}{1 - \beta_s \left(\frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} \right)}$$
$$\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \ \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} < \frac{1}{\beta_s}.$$

In order to make sense of the above problem, then we should require the parameter β_s to be as small as possible. Moreover, we need a relatively small parameter β_s, while we don't want it to be too large. Therefore, we construct the above two mean values $\beta_G(s)$ and $\beta_A(s)$ which limit the range of β_s and then use the arithmetic mean $\beta_A(s)$ to replace the unknown parameter β_s for some given sparsity s. It follows from $\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_0 \|\mathbf{x}\|_\infty = s\|\mathbf{x}\|_\infty$ that $1/s \geq \beta_s$. When $s = 100$, then we have that $0.01 > \beta_G(s) \approx 0.00944564$. Therefore, we can take this constant β_s as the geometric mean $\beta_G(s)$. In this setting, problem (2.10) can be replaced by problem (2.11) which is equivalent to the ℓ_1/ℓ_∞ minimization (2.7).

In summary, we can find a parameter β_s for a given sparsity $s = \|\mathbf{y}\|_0$. We call β_s to be the sparse metric of the sparsity s. Then the final result shows

that we can solve the solution of problem (2.7) which approximates one sparse solution to the ℓ_0 minimization (1.1).

In end of the subsection, we present a numerical experiment to complement our previous theoretical investigation. In the settings of numerical experiments, we take $n = 3000$. Then, the sparsity s ranges from 1 to 1000 by increment of 1, and for each s we perform 100000 independent experiments with $\tan\beta$ recorded and calculate the average value $\beta(s)$ of $\tan\beta$. The entries of signal $\mathbf{y}$ with the sparsity s on its support are i.i.d. Gaussian random variables with zero mean and unit variances. The results are given in Figure 8.

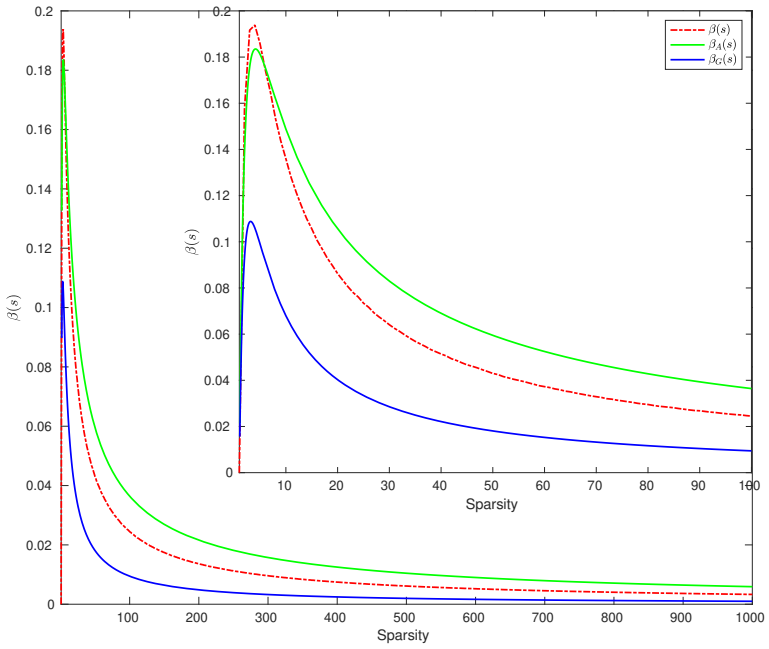


Figure 8: Arithmetic mean $\beta_A(s)$ and Geometric mean $\beta_G(s)$ with the sparsity s ranges from 1 to 1000.

3. Numerical experiments

In this section, we will consider (2.7) with the noiseless case. The noisy case (2.8) can be carried out similarly by tuning the parameter of the penalty term arising from the constraint. Note that ℓ_1/ℓ_∞ minimization (2.7) is non-convex, our algorithms solve the problem approximately. In particular, we utilize the semblable algorithms from [24], which is essentially the Inverse Power Method [35] with extra augmented quadratic term in $\mathbf{x}$-update.

We briefly explain how the algorithm works: it was observed in [24, 34] that subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$, minimizing $\|\mathbf{x}\|_1/\|\mathbf{x}\|_\infty$ is equivalent to minimizing $\|\mathbf{x}\|_1 - \gamma^\infty\|\mathbf{x}\|_\infty$ for some $\gamma^\infty \in [1, \|\mathbf{x}\|_0]$, where γ^∞ is some case-dependent parameter. Following the Difference of Convex Algorithms (DCA) framework [36] to minimizing $\|\mathbf{x}\|_1 - \gamma^\infty\|\mathbf{x}\|_\infty$, we will consider the following problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{\|\mathbf{x}\|_1 - \gamma^\infty\|\mathbf{x}\|_\infty : \mathbf{A}\mathbf{x} = \mathbf{b}\right\}.$$

By linearizing the second term $\|\mathbf{x}\|_\infty$, the DCA iterates as follows

$$\mathbf{x}^{k+1} = \mathrm{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left\{\|\mathbf{x}\|_1 - \gamma_k^\infty\langle\xi^k, \mathbf{x}\rangle : \mathbf{A}\mathbf{x} = \mathbf{b}\right\}, \tag{3.1}$$

where $\xi^k \in \partial\|\mathbf{x}^k\|_\infty$ and $\gamma_k^\infty = \|\mathbf{x}^k\|_1/\|\mathbf{x}^k\|_\infty$. Notice that (3.1) can be formulated as a linear programming (LP) problem [12, IV.1, pp. 1296], which unfortunately has no guarantee that the optimal solution exists [24] (as the problem can be unbounded). To increase the robustness of the above method , we try to incorporate a quadratic term into (3.1) and then obtain that

$$\mathbf{x}^{k+1} = \mathrm{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left\{\|\mathbf{x}\|_1 - \gamma_k^\infty\langle\xi^k, \mathbf{x}\rangle + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^k\|_2^2 : \mathbf{A}\mathbf{x} = \mathbf{b}\right\}.$$

Hence, given an initialization $\mathbf{x}^0 = \mathbf{e} \triangleq (1, 1, \ldots, 1)^\top \in \mathbb{R}^{250}$ and $\gamma_0^\infty = \gamma_0^2 = 3$, the alternating algorithm for solving (2.7) can be summarized in

$$\begin{cases} \mathbf{x}^{k+1} = \mathrm{argmin}_{\mathbf{A}\mathbf{x}=\mathbf{b}} \left\{\|\mathbf{x}\|_1 + \frac{\rho}{2}\left\|\mathbf{x} - \left(\mathbf{x}^k + \frac{\gamma_k^\infty}{\rho}\xi^k\right)\right\|_2^2\right\} \\ \gamma_{k+1}^\infty = \frac{\|\mathbf{x}^{k+1}\|_1}{\|\mathbf{x}^{k+1}\|_\infty}, \end{cases} \tag{3.2}$$

while the iterations for solving $\min_{\mathbf{x}\in\mathbb{R}^n}\left\{\|\mathbf{x}\|_1-\gamma^2\|\mathbf{x}\|_2:\mathbf{A}\mathbf{x}=\mathbf{b}\right\}$, where the parameter $\gamma^2>0$, as follows

$$\begin{cases}\mathbf{x}^{k+1}=\operatorname{argmin}_{\mathbf{A}\mathbf{x}=\mathbf{b}}\left\{\|\mathbf{x}\|_1+\frac{\rho}{2}\left\|\mathbf{x}-\left(\mathbf{x}^k+\frac{\gamma_k^2\mathbf{x}^k}{\rho\|\mathbf{x}^k\|_2}\right)\right\|_2^2\right\}\\ \gamma_{k+1}^2=\frac{\|\mathbf{x}^{k+1}\|_1}{\|\mathbf{x}^{k+1}\|_2},\end{cases}\tag{3.3}$$

where $\xi^k\in\partial\|\mathbf{x}^k\|_\infty$ denotes the subdifferential of the ℓ_∞-norm and $\rho=0.5$. Since the $\mathbf{x}$-subproblem is convex programming, we solve it by using the Matlab tool CVX [37] with default settings.

All test code was written and tested in MATLAB R2020b running on the Linuxmint 20.1 ulyssa (x86_64 Linux 5.4.0-89-generic) which was installed in a Dell Inc. Precision 5820 Tower X-Series with Intel(R) Core(TM) i9-10900X (3.7GHz) and 32 (4×8) GB of UDIMM memory.

It can be seen from (9) that these two algorithms (3.2) and (3.3) can successfully recover sparse signals. In addition, the left-hand figure (a) of Figure

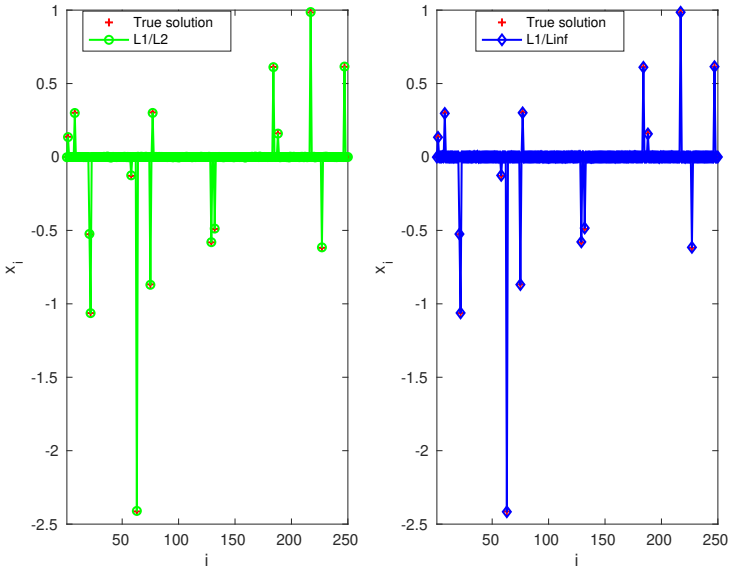
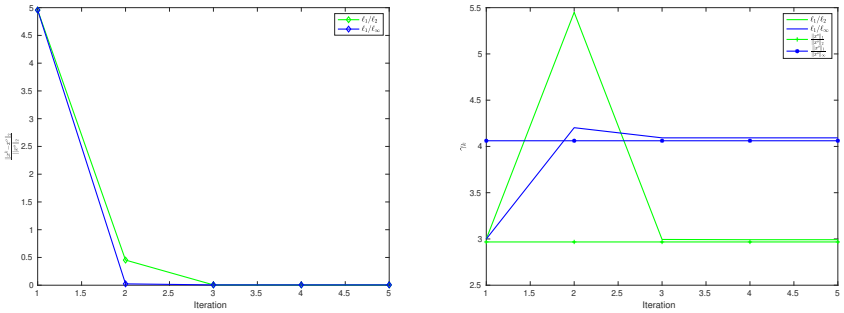


Figure 9: Comparison for recovery of the condition that the sparsity $s=10$ with the linear system $\mathbf{b}=\mathbf{A}\mathbf{x}$, where $\mathbf{A}\in\mathbb{R}^{100\times250}$ and $\mathbf{x}\in\mathbb{R}^{250}$

(10) shows that the relative error defined by

$$\text{RelErr} = \frac{\|\mathbf{x}^{\text{r}} - \mathbf{x}^{o}\|_2}{\|\mathbf{x}^{o}\|_2}$$

of (3.2) is a little better than one of (3.3), where $\mathbf{x}^{o}$ is a original signal and $\mathbf{x}^{\text{r}}$ is a recovery. However, this does not mean that the method (3.2) is better than method (3.3). It is possible that the parameters γ and ρ were set just right this time. Overall, the algorithm (3.2) is almost at the same level as (3.3) since used the same parameter ρ and the solver CVX [37] in the computation. The right-hand figure (b) of Figure (10) verifies the point.



(a) (b)

Figure 10: Comparison of the relative error (a) and the parameter γ_k (b) with the condition that the sparsity $s = 10$ with the linear system $\mathbf{b} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \mathbb{R}^{100\times 250}$ and $\mathbf{x} \in \mathbb{R}^{250}$.

4. Conclusions

We have theoretically investigated the relationship between the ℓ_1/ℓ_∞ and ℓ_0 norm. By summarizing some important objective functions, we found an interesting phenomenon that those objective functions that have a simple and direct connection with the ℓ_0 norm are able to obtain much more sparse solution to some extent. Then, we derive a expected objective function ℓ_1/ℓ_∞ from the deep investigation of the shrinkage operator. In fact, we have considered this issue after studying ℓ_1/ℓ_2 minimization. The ultimate goal is to be able to have a clear understanding of the previous doubts (i) and (ii).

Although our analysis in this article only establishes a relationship between the ℓ_0 norm and the ratio ℓ_1 and ℓ_∞ norms, it does not give a uniform recoverability condition and theoretical justification of whether problem (2.7) and (2.8) can be solved sparsely or not. More theoretical works which analogous to ones of [26] and algorithmic research which is similar to that [24] in this direction are also worth exploring in the future.

Acknowledgment

The research of Jun Wang has been supported by Scientific Startup Foundation for Doctors of Jiangsu University of Science and Technology (CN) (No.1052931903) and Jiangsu Shuangchuang Program for Doctor of China (No.1054902006).

References

[1] Y. C. Eldar, G. Kutyniok, *Compressed Sensing: Theory and Applications*. U.K.: Cambridge Univ. Press, 2012.

[2] S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Cambridge MA, USA: Birkhauser, 2013.

[3] A.M. Bruckstein, D.L. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review, **51**, pp. 34–81, (2009).

[4] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, Single-pixel imaging via compressive sampling, Signal Process. Magazine, vol. 25, pp. 83–91, Mar. 2008.

[5] Q. Berthet, P. Rigollet, Optimal detection of sparse principal components in high dimension, Annal. Statistics, **41**, pp. 1780-1815, (2013).

[6] J. Wright, A.Y. Yang, G. Arvind et al., Robust face recognition via sparse representation, IEEE Trans. Pattern Analysis and Machine Intelligence, **31**, pp. 210–227, (2008).

[7] A. Beck, Y. C. Eldar, Sparsity constrained nonlinear optimization: optimality conditions and algorithms, SIAM J. Optim., **23**, pp. 1480-1509, 2013.

[8] O. P. Burdakov, Y. Kanzow, A. Schartz, Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method, SIAM J. Opim., **26**, pp. 397-425, 2016.

[9] T. Blumensath, M. E. Davies, Iterative thresholding for sparse approximation, J. Fourier Anal. Appl., vol. 14, pp. 629-654, 2008.

[10] F. Wu, W. Bian, Accelerated iterative hard thresholding algorithm for ℓ_0 regularized regression problem, J. Global Optim., vol. 76, pp. 819–840, 2020.

[11] E. J. Candés, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math., vol. 59, pp. 1207-1223, 2006.

[12] D. L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory, vol. 52, no. 4, pp. 1289-1306, 2006.

[13] E. J. Candés, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory vol. 51, pp. 4203-4215, 2005.

[14] A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and best k-term approximation, J. Am. Math. Soc., vol. 22, no. 1, pp. 211- 231, 2008.

[15] E. J. Candés, M. Wakin, S. P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, J. Fourier Anal. Appl., vol. 14, pp. 877–905, 2008.

[16] Y.-B. Zhao, D. Li, Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems, SIAM J. Optim., vol. 22, no. 3, pp. 1065–1088, 2009.

[17] R. Chartrand, W . Yin, Iteratively reweighted algorithms for compressive sensing, in: Proc. Int. Conf. Accoust. Speech, Signal Process., pp. 3869–3872, 2008.

[18] J. Wang, X. T. Wang, Sparse signal reconstruction via the approximations of ℓ_0 quasinorm, J. Indust. Manag. Optim., vol. 16, no. 4, pp. 1907–1925, 2020.

[19] S. Foucart and M. Lai, Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$, Appl. Comput. Harmon. Anal., vol. 26, no. 3, pp. 395–407, 2009.

[20] J. Zeng, S. Lin, and Y. Wang, Z. Xu, $\ell_{1/2}$ regularization: convergence of iterative half thresholding algorithm, IEEE Trans. Signal Process. vol. 62, no. 7, pp. 2317-2329, 2014.

[21] J. Wen, D. Li, F. Zhu, Stable recovery of sparse signals via ℓ_p-minimization, Appl. Comput. Harmon. Anal., vol. 38, pp. 161–176, 2015.

[22] X. Chen, D. Ge, Z. Wang, Y. Ye, Complexity of unconstrained $\ell_2 - \ell_p$ minimization, Math. Program., Ser. A, vol. 143, pp. 371–383, 2014.

[23] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of ℓ_{1-2} for compressed sensing, SIAM J. Sci. Comput., vol. 37, A536–A563, 2015.

[24] C. Wang, M. Yan, Y. Rahimi, Y. Lou, Accelerated schemes for the L1/L2 minimization, IEEE Trans. Signal Process., vol. 68, 2660–2668, 2020.

[25] C. Wang, M. Tao, J. G. Nagy. Rahimi, Y. Lou, Limited-angle CT reconstruction via the L1/L2 minimization, SIAM J. Imaging Science., vol. 14, no. 2, pp. 749–777, 2021.

[26] Y. Xu, A. Narayan, H. Tran, C. G. Webster, Analysis of the ratio of ℓ_1 and ℓ_2 norms in compressed sensing, Appl. Comput. Harmon. Anal., vol. 55, pp. 486–511, 2021.

[27] Y. Rahimi, C. Wang, H. Dong and Y. Lou, A scale invariant approach for sparse signal recovery, SIAM J. Sci. Comput., vol. 41, no. 6, pp. A3649–A3672, 2019.

[28] J. Wang, The proximal gradient methods for the $\ell_{1-\infty}$ minimization problem with the sharp estimate, submitted, 2021.

[29] J. Wang, Sparse reconstruction via the mixture optimization model with iterative support estimate, Information Sciences, vol 574, pp. 1-11, 2021.

[30] E. J. Candés, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, IEEE Trans. Inform. Theory, vol. 52, pp. 5406–5425, 2006.

[31] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sciences., vol. 2, no. 1, pp. 183–202, 2009.

[32] S. Z. Tao, B. Daniel and S.Z. Zhang, Local linear convergence of ISTA and FISTA on the LASSO problem, SIAM J. Optim., vol. 26, pp. 313-336, 2016.

[33] M. Fornasier, Numerical methods for sparse recovery, Radon Series Comp. App. Math. 9, pp. 93-200, 2010.

[34] W. Dinkelbach, On nonlinear fractional programming, Manag. Science, vol. 13, no. 7, pp. 492—498, 1967.

[35] M. Hein, T. Búhler, An inverse power method for nonlinear eigenproblems with application in 1-spectral clustering and sparse PCA, in: Advances in Neural Information Processing Systems, pp. 847—855, 2010.

[36] T. Pham-Dinh, H. A. Le-Thi, The DC (Difference of convex functions) programming and DCA models of real world nonconvex optimization problems , Ann. Oper. Res., vol. 13, no. 1-4, pp. 23-46, 2005.

[37] M. Grant, S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.2. http://cvxr.com/cvx, September 2014.