ITERATIVE ℓ_1 MINIMIZATION FOR NON-CONVEX COMPRESSED SENSING

Penghang Yin * Jack Xin †

April 28, 2016

Abstract

An algorithmic framework, based on the difference of convex functions algorithm, is proposed for minimizing a class of concave sparse metrics for compressed sensing problems. The resulting algorithm iterates a sequence of ℓ_1 minimization problems. An exact sparse recovery theory is established to show that the proposed framework always improves on the basis pursuit (ℓ_1 minimization) and inherits robustness from it. Numerical examples on success rates of sparse solution recovery illustrate further that, unlike most existing non-convex compressed sensing solvers in the literature, our method always out-performs basis pursuit, no matter how ill-conditioned the measurement matrix is. Moreover, the iterative ℓ_1 algorithms lead by a wide margin the state-of-the-art algorithms on $\ell_{1/2}$ and logarithimic minimizations in the strongly coherent (highly ill-conditioned) regime, despite the same objective functions.

1 Introduction

Compressed sensing (CS) techniques [8, 5, 6, 15] enable efficient reconstruction of a sparse signal under linear measurements far less than its physical dimension. Mathematically, CS aims to recover an n-dimensional vector $\bar{x} \in \mathbb{R}^n$ with few non-zero components from an under-determined linear system $Ax = A\bar{x}$ of just $m \ll n$ equations, where $A \in \mathbb{R}^{m\times n}$ is a known measurement matrix. The first CS technique is the convex ℓ_1 minimization or the so-called basis pursuit [13]:

$$\min_{x\in\mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad Ax = A\bar{x}. \tag{1}$$

*Department of Mathematics, UC Irvine, Irvine, CA 92697, USA, (penghany@uci.edu).
†Department of Mathematics, UC Irvine, Irvine, CA 92697, USA, (jxin@math.uci.edu).

Breakthrough results [8] have established that when matrix A satisfies certain restricted isometry property (RIP), the solution to (1) is exactly $\bar{x}$. It was shown that with overwhelming probability, several random ensembles such as random Gaussian, random Bernoulli, and random partial Fourier matrices, are of RIP type [8, 11, 29]. Note that (1) is just a minimization principle rather than an algorithm for retrieving $\bar{x}$. Algorithms for solving (1) and its associated ℓ_1 regularization problem [33]:

$$\min_{x\in\mathbb{R}^n} \ \frac{1}{2}\|Ax-b\|^2+\lambda\|x\|_1 \tag{2}$$

include Bregman methods [40, 21], alternating direction algorithms [3, 37, 16], iterative thresholding methods [12, 1] among others [22].

Inspired by the success of basis pursuit, researchers then began to investigate various non-convex CS models and algorithms. More and more empirical studies have shown that non-convex CS methods usually outperform basis pursuit when matrix A is RIP-like, in the sense that they require fewer linear measurements to reconstruct signals of interest. Instead of minimizing ℓ_1 norm, it is natural to consider minimization of non-convex (concave) sparse metrics, for instance, ℓ_q (quasi-)norm ($0 < q < 1$) [9, 10, 24], capped-ℓ_1 [42, 27], and transformed-ℓ_1 [25, 41]. Another category of CS methods in spirit rely on support detection of $\bar{x}$. To name a few, there are orthogonal matching pursuit (OMP) [34], iterative hard thresholding (IHT) [2], (re)weighted-ℓ_1 scheme [7], iterative support detection (ISD) [35], and their variations [28, 43, 23].

On the other hand, it has been proved that even if A is not RIP-like and contains highly correlated columns, basis pursuit still enables sparse recovery under certain conditions of $\bar{x}$ involving its support [4]. In this scenario, most of the existing non-convex CS methods, however, are not that robust to the conditioning of A, as suggested by [38]. Their success rates will drop as columns of A become more and more correlated. In [38], based on the difference of convex functions algorithm (DCA) [31, 32], the authors propose DCA-ℓ_{1-2} for minimizing the difference of ℓ_1 and ℓ_2 norms [17, 39]. Extensive numerical experiments [38, 26, 27] show that DCA-ℓ_{1-2} algorithm consistently outperforms ℓ_1 minimization, irrespective of the conditioning of A.

Stimulated by the empirical evidence found in [38, 26, 27], we propose a general DCA-based CS framework for the minimization of a class of concave sparse metrics. More precisely, we consider the reconstruction of a sparse vector $\bar{x} \in \mathbb{R}^n$ by minimizing sparsity-promoting metrics:

$$\min_{x \in \mathbb{R}^n} P(|x|) \quad \text{s.t.} \quad Ax = A\bar{x}. \tag{3}$$

Throughout the paper, we assume that $P(x)$ always takes the form $\sum_{i=1}^n p(x_i)$ unless otherwise stated. To promote sparsity, p defined on $[0, +\infty)$ generally satisfies:

- p is continuous, concave and increasing.
- The right derivative $p'(0+)$ exists with $p'(0+) > 0$.

A number of sparse metrics in the literature enjoy the above properties, including smoothly clipped absolute deviation (SCAD) [18], capped-ℓ_1, transformed-ℓ_1, and of course ℓ_1 itself. Although ℓ_q ($q \in (0,1)$) and logarithm functional do not meet the second condition, their smoothed versions $p(t) = (t+\varepsilon)^q$ and $p(t) = \log(t+\varepsilon)$ are differentiable at zero. These proposed properties will be essential in the algorithm design as well as in the proof of main results.

Our proposed algorithm calls for solving a sequence of minimization subproblems. The objective of each subproblem is $\|x\|_1$ plus a linear term, which is convex and tractable. We further validate robustness of this framework, by showing theoretically and numerically that it performs at least as well as basis pursuit in terms of uniform sparse recovery, independent of the conditioning of A and sparsity metric.

The paper is organized as follows. In section 2, we overview RIP and coherence of sensing matrices, as well as descent property of DCA. In section 3, we provide the iterated ℓ_1 framework for non-convex minimization, with worked out examples on representaitive sparse objectives including the total variation. In section 4, we prove the main exact recovery results based on unique recovery property of ℓ_1 minimization instead of RIP, which forms a theoretical basis of the better performance of DCA algorithms. In section 5, we compare iterative ℓ_1 algorithms with two state-of-the-art non-convex CS algorithms, IRLS-ℓ_q [24] and IRL$_1$ [7], and ADMM-ℓ_1, in CS test problems with varying degree of coherence. We found that iterative ℓ_1 out-performs ADMM-ℓ_1 independent of the sensing matrix coherence, and leads IRLS-ℓ_q [24] and IRL$_1$ [7] in the highly coherent regime. This is consistent with earlier findings of DCA-ℓ_{1-2} algorithm [38, 26, 27] to which our theory also applies. Concluding remarks are in section 6.

Notations. Let us fix some notations. For any $x, y \in \mathbb{R}^n$, $\langle x, y\rangle = x^{\mathrm{T}} y$ is their inner product. $\mathbf{0} \in \mathbb{R}^n$ is the vector of zeros, and similar to $\mathbf{1}$. $\circ$

is Hadamard (entry-wise) product, meaning that $x \circ y = (x_1 y_1, \cdots, x_n y_n)^{\mathrm{T}}$. I_m is the identity matrix of dimension m. For any function g on $\mathbb{R}^n$, $\nabla g(x) \in \partial g(x)$ is a subgradient of g at x. The $\operatorname{sgn}(x)$ is the signum function on $\mathbb{R}^n$ defined as

$$(\operatorname{sgn}(x))_i := \begin{cases} \frac{x_i}{|x_i|} & \text{if } x_i \neq 0, \\ 0 & \text{if } x_i = 0. \end{cases}$$

For any set $\Omega \subseteq \mathbb{R}^n$, $\iota_\Omega(x)$ is given by

$$\iota_\Omega(x) := \begin{cases} 0 & \text{if } x \in \Omega, \\ \infty & \text{if } x \notin \Omega. \end{cases}$$

2 Preliminaries

The well-known restricted isometry property (RIP) is introduced by Candès *et al.* [8] characterizing matrices that are nearly orthonormal. The rigorous definition of RIP condition is as follows.

Definition 2.1. *For each number s, s-restricted isometry constant of A is the smallest $\delta_s \in (0,1)$ such that*

$$(1-\delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_s)\|x\|_2^2$$

for all $x \in \mathbb{R}^n$ with sparsity of s. The matrix A is said to satisfy the s-RIP with δ_s.

Another commonly used CS concept is the mutual coherence [13].

Definition 2.2. *The coherence of a matrix A is the maximum absolute value of the cross-correlations between the columns of A, namely*

$$\mu(A) := \max_{i \neq j} \frac{|A_i^{\mathrm{T}} A_j|}{\|A_i\|_2 \|A_j\|_2}.$$

A matrix satisfying RIP condition tends to have small coherence or to be incoherent. Conversely, a highly coherent matrix is unlikely to be RIP-like.

Next we give a brief review on the difference of convex functions algorithm (DCA). For an objective function $F(x) = G(x) - H(x)$ on the space $\mathbb{R}^n$, where $G(x)$ and $H(x)$ are lower semicontinuous proper convex functions, we call $G - H$ a DC decomposition of F.

DCA takes the following form

$$\begin{cases} y^{(k)} \in \partial H(x^{(k)}) \\ x^{(k+1)} = \arg\min_{x \in \mathbb{R}^n} G(x) - (H(x^{(k)}) + \langle y^k, x - x^k \rangle) \end{cases}$$

Since $y^{(k)} \in \partial H(x^{(k)})$, by the definition of subgradient, we have

$$H(x^{k+1}) \geq H(x^k) + \langle y^k, x^{k+1} - x^k \rangle.$$

Consequently,

$$G(x^{(k)}) - H(x^{(k)}) \geq G(x^{(k+1)}) - (H(x^{(k)}) + \langle y^{(k)}, x^{(k+1)} - x^{(k)} \rangle) \geq G(x^{(k+1)}) - H(x^{(k+1)}).$$

The fact that $x^{(k+1)}$ minimizes $G(x) - (H(x^{(k)}) + \langle y^k, x - x^{(k)} \rangle)$ was used in the first inequality above. Therefore, DCA permits a decreasing sequence $\{F(x^{(k)})\}$, leading to its convergence provided $F(x)$ is bounded from below.

3 Iterative ℓ_1 framework

Our proposed iterative ℓ_1 framework for solving (3) is built on ℓ_1 minimization and DCA. Note that (3) can be equivalently written as

$$\min_{x \in \mathbb{R}^n} P(|x|) + \iota_{\{x : Ax = A\bar{x}\}}(x).$$

We then rewrite the above objective in DC decomposition form:

$$P(|x|) + \iota_{\{x:Ax=A\bar{x}\}}(x) = (p'(0+)\|x\|_1 + \iota_{\{x:Ax=A\bar{x}\}}(x)) - (p'(0+)\|x\|_1 - \sum_{i=1}^{n} p(|x_i|))$$

Clearly the first term on the right-hand side is convex in terms of x. We show below that the second term is also a convex function.

Proposition 3.1. $p'(0+)\|x\|_1 - \sum_{i=1}^n p(|x_i|)$ *is convex in* x.

Proof. For notational convenience, define $f(t) := p'(0+)t - p(t)$ on $[0, \infty)$. Since p is concave on $[0, \infty)$, we have that f is convex on $[0, \infty)$. We only need to show that $f(|\cdot|)$ is convex on $\mathbb{R}$, or equivalently, for all $t_1, t_2 \in \mathbb{R}$, $a \in (0, 1)$,

$$f(|at_1 + (1-a)t_2|) \leq af(|t_1|) + (1-a)f(|t_2|).$$

Case 1. If t_1 and t_2 have the same sign or one of them is 0. Since $f(|at_1 + (1-a)t_2|) = f(a|t_1| + (1-a)|t_2|)$ and f is convex on $[0, \infty)$, then the above inequality holds.

Case 2. If t_1 and t_2 are of the opposite sign. By the concavity of p on $[0, \infty)$, we have

$$p(t) \leq p(0) + p'(0+)t, \ \forall t > 0,$$

that is, $f(t) \geq f(0)$ for all $t > 0$. Without loss of generality, we suppose $a|t_1| \geq (1-a)|t_2|$. Then

$$
\begin{aligned}
& f(|at_1 + (1-a)t_2|) = f(a|t_1| - (1-a)|t_2|) \\
\leq & \frac{(1-a)(|t_1| + |t_2|)}{|t_1|} f(0) + \frac{a|t_1| - (1-a)|t_2|}{|t_1|} f(|t_1|) \\
\leq & (1-a) f(|t_2|) + \frac{(1-a)|t_2|}{|t_1|} f(|t_1|) + \frac{a|t_1| - (1-a)|t_2|}{|t_1|} f(|t_1|) \\
= & a f(|t_1|) + (1-a) f(|t_2|)
\end{aligned}
$$

In the first inequality above, we used the convexity of f on $[0, \infty)$, whereas in the second one, we used the fact that $f(t) \geq f(0)$ for $t > 0$. □

At the $(k+1)^{\text{th}}$ iteration, DCA calls for linearization of the second convex term at the current guess $x^{(k)}$, and solving the resulting convex subproblem for $x^{(k+1)}$. After converting back the linear constraint and removing the constant and the factor of $p'(0+)$, we iterate:

$$
x^{(k+1)} = \arg\min_x \|x\|_1 - \langle R(x^{(k)}), x \rangle \quad \text{s.t.} \quad Ax = A\bar{x}, \tag{4}
$$

where

$$
R(x) := \operatorname{sgn}(x) \circ (\mathbf{1} - \frac{\nabla P(|x|)}{p'(0+)}) \in \partial(\| \cdot \|_1 - \frac{P(| \cdot |)}{p'(0+)})(x).
$$

Be aware that $\nabla P(|x|) \in \partial P(\cdot)(|x|)$ denotes subgradient of P at $|x|$ instead of subgradient of $P(| \cdot |)$ at x. In this way, the subproblem reduces to minimizing $\|x\|_1$ plus a linear term of x, which can be effciently solved by a variety of state-of-the-art algorithms for basis pursuit (with minor modifications). In Table 1, we list some non-convex metrics and the corresponding iterative ℓ_1 algorithm.

Table 1: Examples of sparse metrics and associated iterative ℓ_1 scheme

sparse metric	$p(t)$	$p'(0+)$	$(R(x))_i$
capped-ℓ_1	$\min\{t, \theta\},\ \theta > 0$	1	$\operatorname{sgn}(x_i) 1_{\lvert x_i \rvert \geq \theta}$
transformed-ℓ_1	$\frac{(\theta+1)t}{t+\theta},\ \theta > 0$	$\frac{\theta+1}{\theta}$	$\operatorname{sgn}(x_i)(1 - (\frac{\theta}{\lvert x_i \rvert + \theta})^2)$
smoothed log	$\log(t + \varepsilon),\ \varepsilon > 0$	$\frac{1}{\varepsilon}$	$\operatorname{sgn}(x_i)(1 - \frac{\varepsilon}{\lvert x_i \rvert + \varepsilon})$
smoothed ℓ_q	$(t + \varepsilon)^q,\ \varepsilon > 0$	$q\varepsilon^{q-1}$	$\operatorname{sgn}(x_i)(1 - (\frac{\varepsilon}{\lvert x_i \rvert + \varepsilon})^{1-q})$

For initialization, we take $x^{(0)} = R(x^{(0)}) = 0$, which is basically to solve ℓ_1 minimization. The proposed algorithm is thus summarized in Algorithm 1 below. Due to the descending property of DCA, Algorithm 1 produces a convergent sequence $\{P(x^{(k)})\}$. Beyond that, we shall not prove any stronger convergence result in this paper. We refer the readers to [38], in which subsequential convergence of $\{x^{(k)}\}$ is established for DCA-ℓ_{1-2}.

Algorithm 1 Iterative ℓ_1 minimization

Initialize: $x^{(0)} = \mathbf{0}$.

for $k = 1, 2, \ldots$ **do**

$y^{(k)} = \mathrm{sgn}(x^{(k)}) \circ (\mathbf{1} - \frac{\nabla P(|x^{(k)}|)}{p'(0+)})$

$x^{(k+1)} = \arg\min_x \|x\|_1 - \langle y^{(k)}, x\rangle \quad \text{s.t.} \quad Ax = A\bar{x}$

end for

Extensions. Two natural extensions of (3) are regularized model:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda P(|Dx|), \tag{5}$$

and denoising model:

$$\min_{x\in\mathbb{R}^n} P(|Dx|) \quad \text{s.t.} \quad \|Ax - b\|_2 \le \sigma, \tag{6}$$

where b is the measurement, D is a general matrix, and $\lambda, \sigma > 0$ are parameters. They find applications in magnetic resonance imaging (MRI) [5], total variation (TV) denoising [30] and so on. We can show that DC decomposition of $P(|Dx|)$ is

$$P(|Dx|) = p'(0+)\|Dx\|_1 - (p'(0+)\|Dx\|_1 - P(|Dx|)).$$

The iterative ℓ_1 frameworks are detailed in Algorithms 2 and 3 respectively.

Algorithm 2 Iterative ℓ_1 regularization

Initialize: $x^{(0)} = \mathbf{0}$.

for $k = 1, 2, \ldots$ **do**

$y^{(k)} = D^{\mathrm{T}}\left(\mathrm{sgn}(Dx^{(k)}) \circ (\mathbf{1} - \frac{\nabla P(|Dx^{(k)}|)}{p'(0+)})\right)$

$x^{(k+1)} = \arg\min_x \frac{1}{2}\|Ax - b\|_2^2 + \lambda p'(0+)(\|Dx\|_1 - \langle y^{(k)}, x\rangle)$

end for

Algorithm 3 Iterative ℓ_1 denoising

Initialize: $x^{(0)} = \mathbf{0}$.

for $k = 1, 2, \dots$ **do**

$y^{(k)} = D^{\mathrm{T}} \left(\mathrm{sgn}(Dx^{(k)}) \circ (\mathbf{1} - \frac{\nabla P(|Dx^{(k)}|)}{p'(0+)}) \right)$

$x^{(k+1)} = \|Dx\|_1 - \langle y^{(k)}, x \rangle \quad \text{s.t.} \quad \|Ax - b\|_2 \le \sigma$

end for

4 Recovery results

Although in general global minimum is not guaranteed in minimization, we can show that its performance is provably robust to the conditioning of measurement matrix A, by proving that it always tends to sharpen ℓ_1 solution.

Let us take another look at the assumptions on p which were crucial in the proof of Proposition 3.1. Since p is concave and increasing on $[0, \infty)$, $0 \le (\frac{\nabla P(|x|)}{p'(0+)})_i \le 1$, $\forall x \in \mathbb{R}, 1 \le i \le n$, and thus $\|R(x)\|_\infty \le 1$. Now we are ready to show the main results.

Theorem 4.1 (Support-wise uniform recovery)**.** *Let $T \subseteq \{1, \dots, n\}$ be an arbitrary but fixed index set. If basis pursuit uniquely recovers all $\bar{x}$ supported on T, so does (4).*

Proof. By the assumption that basis pursuit uniquely recovers all $\bar{x}$ supported on T, and by the well-known null space property [20] for ℓ_1 minimization, we must have

$$\|h_T\|_1 < \|h_{T^c}\|_1, \ \forall h \in \mathrm{Ker}(A) \setminus \{\mathbf{0}\},$$

and $x^{(1)} = \bar{x}$ in (4). The 2^{nd} step of DCA reads

$$x^{(2)} = \arg\min \|x\|_1 - \langle R(\bar{x}), x \rangle \quad \text{s.t.} \quad Ax = A\bar{x}.$$

Let $x^{(2)} = \bar{x} + h^{(2)}$, then

$$\begin{aligned}
&\|\bar{x}\|_1 - \langle R(\bar{x}), \bar{x} \rangle \ge \|\bar{x} + h^{(2)}\|_1 - \langle R(\bar{x}), \bar{x} + h^{(2)} \rangle \\
\Longrightarrow &\|\bar{x}\|_1 - \langle R(\bar{x}), \bar{x} \rangle \ge \|\bar{x}\|_1 + \langle \mathrm{sgn}(\bar{x}), h_T^{(2)} \rangle + \|h_{T^c}^{(2)}\|_1 - \langle R(\bar{x}), \bar{x} + h_T^{(2)} \rangle \\
\Longleftrightarrow & - \langle \mathrm{sgn}(\bar{x}) - R(\bar{x}), h_T^{(2)} \rangle \ge \|h_{T^c}^{(2)}\|_1 \\
\Longleftrightarrow & - \langle \mathrm{sgn}(\bar{x}) \circ \frac{\nabla P(|\bar{x}|)}{p'(0+)}, h_T^{(2)} \rangle \ge \|h_{T^c}^{(2)}\|_1
\end{aligned}$$

Since $\|\frac{\nabla P(|\bar{x}|)}{p'(0+)}\|_\infty \le 1$, we have $\|h_T^{(2)}\|_1 \ge \|h_{T^c}^{(2)}\|_1$. As a result, $h^{(2)}$ must be $\mathbf{0}$. □

If nonzero entries of $\bar{x}$ have the same magnitude, a stronger result holds that (4) recovers any fixed signal whenever basis pursuit does.

Theorem 4.2 (Recovery of equal-height signals)**.** *Let $\bar{x}$ be a signal with equal-height peaks supported on T, i.e. $|x_i| = |x_j|, \forall i, j \in T$. If the basis pursuit uniquely recovers $\bar{x}$, so does (4).*

Proof. If basis pursuit uniquely recovers $\bar{x}$, then for all $h \in \mathrm{Ker}(A) \setminus \{\mathbf{0}\}$, $\|\bar{x}\|_1 < \|\bar{x} + h\|_1 = \|\bar{x} + h_T\|_1 + \|h_{T^c}\|_1$. This implies that for all $h \in \mathrm{Ker}(A) \setminus \{\mathbf{0}\}$ and $\|h\|_\infty \le \min_{i \in T} |\bar{x}_i|$, $\|\bar{x}\|_1 < \|\bar{x} + h_T\|_1 + \|h_{T^c}\|_1 = \|\bar{x}\|_1 + \langle \mathrm{sgn}(\bar{x}), h_T\rangle + \|h_{T^c}\|_1$. So for all $h \in \mathrm{Ker}(A) \setminus \{\mathbf{0}\}$ and $\|h\|_\infty \le \min_{i \in T} |\bar{x}_i|$, we have $-\langle \mathrm{sgn}(\bar{x}), h_T\rangle < \|h_{T^c}\|_1$.

Therefore,

$$-\langle \mathrm{sgn}(\bar{x}), h_T\rangle < \|h_{T^c}\|_1,\ \forall h \in \mathrm{Ker}(A) \setminus \{\mathbf{0}\}, \tag{7}$$

and also $x^{(1)} = \bar{x}$.

We let $x^{(2)} = \bar{x} + h^{(2)}$, and suppose that $h^{(2)} \ne \mathbf{0}$. Repeating the argument in Theorem 4.1 and by (7), we arrive at

$$-\langle \mathrm{sgn}(\bar{x}) \circ \frac{\nabla P(|\bar{x}|)}{p'(0+)}, h_T^{(2)}\rangle \ge \|h_{T^c}^{(2)}\|_1 > -\langle \mathrm{sgn}(\bar{x}), h_T^{(2)}\rangle.$$

Since peaks of $\bar{x}$ have equal height, $(\frac{\nabla P(|\bar{x}|)}{p'(0+)})_i \in [0, 1)$ is a constant for all $i \in T$. So $-\langle \mathrm{sgn}(\bar{x}) \circ \frac{\nabla P(|\bar{x}|)}{p'(0+)}, h_T^{(2)}\rangle$ is non-negative and less than $-\langle \mathrm{sgn}(\bar{x}), h_T^{(2)}\rangle$, which leads to a contradiction. □

Remark 4.1. *We did not make any assumption on the matrix A in the above theorems.*

Remark 4.2. *It is easy to check that the recovery results can be extended to DCA-ℓ_{1-2} proposed in [38]. For DCA-ℓ_{1-2}, $P(x) = \|x\|_1 - \|x\|_2$, which couples all the components together, and*

$$R(x) = \begin{cases} \frac{x}{\|x\|_2} & \text{if } x \ne \mathbf{0}, \\ \mathbf{0} & \text{if } x = \mathbf{0}. \end{cases}$$

Theorem 4.1 provides a theoretical explanation for the experimental observations made in [38, 26, 27] that DCA-ℓ_{1-2} performs consistently better than ℓ_1 minimization.

5 Numerical experiments

To demonstrate effectiveness of the proposed framework, we reconstruct sparse vector $\bar{x}$ using iterative ℓ_1 algorithm (Algorithm 2 with $D = I_n$) for minimizing the regularized model (5) with smoothed ℓ_q norm (IL$_1$-ℓ_q) and smoothed logarithm functional (IL$_1$-log), and compare them with two state-of-the-art non-convex CS algorithms, namely IRLS-ℓ_q [24] and IRL$_1$ [7]. Note that IRLS-ℓ_q and IRL$_1$ attempt to minimize ℓ_q and logarithm, respectively, and both involve a smoothing strategy in minimization. So it would be particularly interesting to compare IL$_1$-ℓ_q with IRLS-ℓ_q, and IL$_1$-log with IRL$_1$. $q = 0.5$ is chosen for IRLS-ℓ_q and IL$_1$-ℓ_q in all experiments. We shall also include ADMM-ℓ_1 [3] for solving ℓ_1 regularization (LASSO) in comparison.

Experiments are carried out as follows. We first sample a sensing matrix $A \in \mathbb{R}^{m\times n}$, and generate a test signal $\bar{x} \in \mathbb{R}^n$ of sparsity s supported on a random index set with i.i.d. Gaussian entries. We then compute the measurement $A\bar{x}$ and apply each solver to produce a reconstruction x^* of $\bar{x}$. The reconstruction is called a success if

$$\frac{\|x^* - \bar{x}\|_2}{\|\bar{x}\|_2} < 10^{-3}.$$

We run 100 independent realizations and record the corresponding success rates at different sparsity levels.

Matrix for test. We test on random Gaussian matrix whose columns satisfy

$$A_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, I_m/m), \quad i = 1, \cdots, n$$

Gaussian matrices are RIP-like and have uncorrelated (incoherent) columns. For Gaussian matrix, we choose $m = 64$ and $n = 256$.

We also use more ill-conditioned sensing matrix of significantly higher coherence. Specifically, a randomly oversampled partial DCT matrix A is defined as

$$A_i = \frac{1}{\sqrt{m}} \cos(2i\pi\xi/F), \quad i = 1, \cdots, n$$

where $\xi \in \mathbb{R}^m \sim \mathcal{U}([0,1]^m)$ whose components are uniformly and independently sampled from [0,1]. $F \in \mathbb{N}$ is the refinement factor. Coherence $\mu(A)$ goes up as F increases. In this setting, it is still possible to recover the sparse vector $\bar{x}$ if its spikes are sufficiently separated. Specifically, we randomly select a T (support of $\bar{x}$) so that

$$\min_{j,k\in T} |j - k| \geq L,$$

where L is called the minimum separation. It is necessary for L to be at least 1 Rayleigh length (RL) which is unity in the frequency domain [19, 14]. In our case, the value of 1 RL equals F. The testing matrix $A \in \mathbb{R}^{100\times 1500}$, i.e. $m = 100$, $n = 1500$. We test at three coherence levels with $F = 5, 10, 15$. Note that $\mu(A) \approx 0.95$ for $F = 5$, $\mu(A) \approx 0.998$ for $F = 10$, and $\mu(A) \approx 0.9996$ for $F = 15$. We also set $L = 2F$ in experiments.

Algorithm implementation. For ADMM-ℓ_1, we let $\lambda = 10^{-6}$, $\beta = 1$, $\rho = 10^{-5}$, $\epsilon^{\text{abs}} = 10^{-7}$, $\epsilon^{\text{rel}} = 10^{-5}$, and the maximum number of iterations `maxiter` = 5000 [3, 38]. For IRLS-ℓ_q, `maxiter` = 1000, `tol`= 10^{-8}. For rewighted ℓ_1, the smoothing parameter ε is adaptively updated as introduced in [7], and the outer iteration criterion is stopped if the relative error between two consecutive iterates is less than 10^{-2}. The weighted ℓ_1 minimization subproblems is solved by the YALL1 solver (available at http://yall1.blogs.rice.edu/). The tolerance for YALL1 was set to 10^{-6}. All other settings of the algorithms are set to default ones.

For IL$_1$-ℓ_q, we let $\lambda = 10^{-6}$, and the smoothing parameter $\varepsilon = \max\{\frac{|x^{(1)}|_{(d)}}{3}, 0.01\}$, where $x^{(1)}$ is the output from the first iteration, which is also the solution to LASSO. $|x|_{(d)}$ denotes the d^{th} largest entry of $|x|$. We set d to $\lfloor \frac{m}{4} \rfloor$. For IL$_1$-log, $\varepsilon = \max\{|x^{(1)}|_{(d)}, 0.01\}$. The subproblems are solved by alternating direction method of multipliers (ADMM), which is detailed in [38]. The parameters for solving subproblems are the same as that for ADMM-ℓ_1.

Interpretation of results. The plot of success rates is shown in Figure 1. When A is Guassian, we see that all non-convex CS solvers are comparable and much better than ADMM-ℓ_1, with IRLS-ℓ_q being the best. For oversampled DCT matrices, we see that the success rates of IRLS-ℓ_q and IRL$_1$ drop as F increases, whereas the proposed IL$_1$-ℓ_q and IL$_1$-log are robust and consistently outperform ADMM-ℓ_1.

6 Conclusions

We developed an iterative ℓ_1 framework for a broad class of Lipschitz continuous non-convex sparsity promoting objectives, including those arising in statistics. The iterative ℓ_1 algorithm is shown via theory and computation to improve on the ℓ_1 minimization for CS problems independent of the coherence of the sensing matrices.

Acknowledgments. The authors would like to thank Yifei Lou (University of Texas at Dallas) and Jong-Shi Pang (University of Southern California)

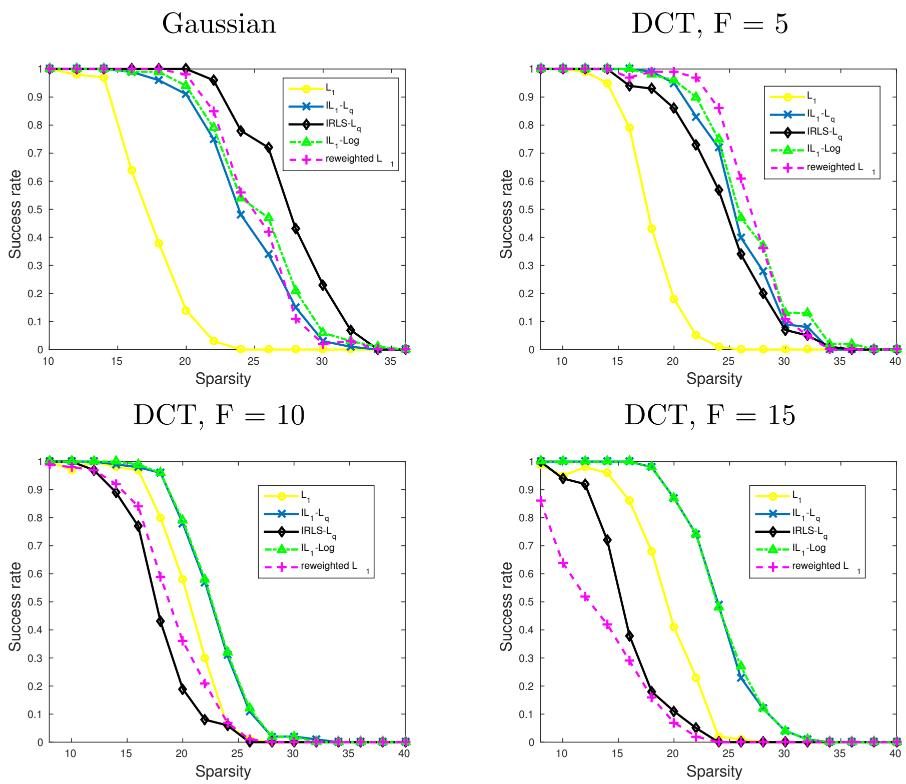

Figure 1: Plots of success rates for comparing the iterative ℓ_1 with other CS algorithms under the increasing coherence of the sensing matrices.

for helpful discussions. The work was partially supported by NSF grants DMS-1222507 and DMS-1522383.

References

[1] A. Beck and M. Teboulle, A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, *SIAM J. Imaging Sci.*, Vol. 2, No. 1, pp. 183-202, 2009.

[2] T. Blumensath and M. Davies, Iterative hard thresholding for compressed sensing, *Applied and Computational Harmonic Analysis*, 27.3 (2009): 265-274.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Foundations and Trends in Machine Learning*, 3(1):1-122, 2011.

[4] E. Candès and C. Fernandez-Granda, Towards a mathematical theory of super-resolution, *Communication on Pure and Applied Mathematics*, 67(6):906–956 (2014).

[5] E. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information, *IEEE Transactions on Information Theory*, 52(2):489–509 (2006).

[6] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, 59(8):1207–1223 (2006).

[7] E. Candès, M. Wakin, and S. Boyd, Enhancing Sparsity by Reweighted ℓ_1 Minimization, *J. Fourier Anal. Appl.*, 14(5), pp. 877-905, 2008.

[8] E. Candès and T. Tao, Decoding by Linear Programming, *IEEE Trans. Info. Theory*, 51 (2005), pp. 4203-4215.

[9] R. Chartrand and V. Staneva, Restricted isometry properties and nonconvex compressive sensing, *Inverse Problems*, 24 (2008), pp. 1-14.

[10] R. Chartrand and W. Yin, Iteratively Reweighted Algorithms for Compressive Sensing, *IEEE international conference on acoustics, speech, and signal processing*, pp. 3869-3872, 2008.

[11] A. Cohen, W. Dahmen, and R. Devore, Compressed Sensing and Best k-Term Approximation, *Journal of the American Mathematical Society*, Vol 22, No. 1, pp. 221-231, 2009.

[12] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Applied Math.*, 57(2004), pp. 1413-1457.

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

[14] D. Donoho. Super-resolution via sparsity constraints, *SIAM Journal on Mathematical Analysis*, 23(5):1309-1331 (1992).

[15] D. Donoho. Compressed sensing, *IEEE Transactions on Information Theory*, 52(4),pp. 1289-1306, 2006.

[16] E. Esser, Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman, CAM-report 09-31, UCLA, 2009.

[17] E. Esser, Y. Lou, and J. Xin, A Method for Finding Structured Sparse Solutions to Non-negative Least Squares Problems with Applications, *SIAM J. Imaging Sciences*, 6(4), pp. 2010-2046, 2013.

[18] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal American Statistical Association*, 96(456), pp. 1348-1360, 2001.

[19] A. Fannjiang and W. Liao, Coherence Pattern-Guided Compressive Sensing with Unresolved Grids, *SIAM J. Imaging Sciences*, Vol. 5, No. 1, pp. 179-202, 2012.

[20] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing, Springer, 2013

[21] T. Goldstein and S. Osher, The split Bregman method for L1 regularized problems, *SIAM Journal on Imaging Sciences*, 2(1), pp. 323-343, 2009.

[22] E. Hale, W. Yin, and Y. Zhang, Fixed-point continuation for ℓ_1-minimization: Methodology and convergence, *SIAM J. Optim.*, 19, pp. 1107-1130, 2008.

[23] X. Huang, L. Shi, and M. Yan, Nonconvex Sorted ℓ_1 Minimization for Sparse Approximation, *Journal of the Operations Research Society of China*, Volume 3, Issue 2, pp. 207-229, 2015.

[24] M. Lai, Y. Xu, and W. Yin, Improved Iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization, *SIAM J. Numer. Anal.*, Vol. 51, Issue 2, pp. 927-957, 2013.

[25] J. Lv, and Y. Fan, A unified approach to model selection and sparse recovery using regularized least squares, *Annals of Statistics*, 37(6A), pp. 3498-3528, 2009.

[26] Y. Lou, P. Yin, Q. He, and J. Xin, Computing Sparse Representation in a Highly Coherent Dictionary Based on Difference of L_1 and L_2, *J. Sci. Comput.*, 64 (2015), pp. 178-196.

[27] Y. Lou, P. Yin, and J. Xin, Point Source Super-resolution via Non-convex L_1 Based Methods, *J. Sci. Computing*, to appear.

[28] D. Needell and J. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmon. Anal.*, 26 (2009), 301-221.

[29] H. Rauhut, Compresssive Sensing and Structured Random Matrices, *Radon Series Comp. Appl. Math.*, Vol. 9, pp. 1-92, 2010.

[30] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, 60(1992), pp. 259–268.

[31] Pham Dinh Tao and Le Thi Hoai An, Convex analysis approach to d.c. programming: Theory, algorithms and applications, *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.

[32] Pham Dinh Tao and Le Thi Hoai An, A DC optimization algorithm for solving the trust-region subproblem, *SIAM Journal on Optimization*, 8(2), pp. 476–505, 1998.

[33] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Royal. Statist. Soc.*, 58(1):267–288, 1996.

[34] J. Tropp and A. Gilbert, Signal recovery from partial information via orthogonal matching pursuit, *IEEE Trans. Inform. Theory*, 53(12):4655–4666, 2007.

[35] Y. Wang and W. Yin, Sparse signal reconstruction via iterative support detection, *SIAM Journal on Imaging Sciences*, 3.3 (2010): 462-491.

[36] Z. Xu, X. Chang, F. Xu, H. Zhang, $L_{1/2}$ regularization: an iterative thresholding method, *IEEE Transactions on Neural Networks and Learning Systems*, 23, pp. 1013-1027, 2012.

[37] J. Yang and Y. Zhang, Alternating direction algorithms for l_1 problems in compressive sensing, *SIAM J. Sci. Comput.*, 33(1), pp. 250–278, 2011.

[38] P. Yin, Y. Lou, Q. He, and J. Xin, Minimization of ℓ_{1-2} for Compressed Sensing, *SIAM J. Sci. Comput.*, 37 (2015), pp. A536-A563.

[39] P. Yin, E. Esser, J. Xin, Ratio and Difference of L1 and L2 Norms and Sparse Representation with Coherent Dictionaries, *Commun. Information and Systems*, 14(2), 2014, pp. 87–109.

[40] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, Bregman iterative algorithms for ℓ_1 minimization with applications to compressed sensing, *SIAM J. Imaging Sci.*, 1 (2008), pp. 143-168.

[41] S. Zhang and J. Xin, Minimization of Transformed L1 Penalty: Theory, Difference of Convex Function Algorithm, and Robust Application in Compressed Sensing, preprint, arXiv:1411.5735.

[42] T. Zhang, Multi-stage convex relaxation for learning with sparse regularization, *NIPS proceedings*, 2008.

[43] S. Zhou, N. Xiu, Y. Wang, L. Kong, and H. Qi A null-space-based weighted ℓ_1 minimization approach to compressed sensing, *Information and Inference*, 2016: iaw002v1-iaw002.