THE CONVERGENCE GUARANTEES OF A NON-CONVEX APPROACH FOR SPARSE RECOVERY USING REGULARIZED LEAST SQUARES

Laming Chen and Yuantao Gu[*]

State Key Laboratory on Microwave and Digital Communications
Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA

ABSTRACT

Existing literatures suggest that sparsity is more likely to be induced with non-convex penalties, but the corresponding algorithms usually suffer from multiple local minima. In this paper, we introduce a class of sparsity-inducing penalties and provide the convergence guarantees of a non-convex approach for sparse recovery using regularized least squares. Theoretical analysis demonstrates that under some certain conditions, if the non-convexity of the penalty is below a threshold (which is in inverse proportion to the distance between the initialization and the sparse signal), the sparse signal can be stably recovered. Numerical simulations are implemented to verify the theoretical results in this paper and to compare the performance of this approach with other references.

Index Terms— Sparse recovery, weak convexity, non-convex optimization, convergence analysis.

1. INTRODUCTION

Special attention has been paid to exploiting the characteristic of sparsity in the field of signal processing in recent years, especially along with the emerging compressive sensing (CS) [2, 3]. One of the key issues arises is the problem of sparse recovery. Suppose we observe M linear measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}, \tag{1}$$

where $\mathbf{x}^* = (x_i^*) \in \mathbb{R}^N$ is an unknown sparse signal to be recovered, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a sensing matrix with more columns than rows, and $\mathbf{e}$ is the additive noise to the measurement vector. The problem of finding the sparse solution to (1) can be recast to the following regularized least squares problem

$$\operatorname*{argmin}_{\mathbf{x}} \left\{ G(\mathbf{x}) = J(\mathbf{x}) + \lambda \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \right\}, \tag{2}$$

where $J(\cdot)$ is a sparsity-inducing penalty and λ is a parameter to balance the sparsity and the accuracy of measurements.

When the ℓ_1 norm is adopted as the penalty $J(\cdot)$, problem (2) is known as *basis pursuit denoising* (BPDN) [4]. Theoretical analysis [5–7] reveals that BPDN can stably identify $\mathbf{x}^*$ if $\mathbf{x}^*$ is sufficiently sparse and the noise $\mathbf{e}$ has bounded ℓ_2 norm or is white Gaussian. Various efficient algorithms are proposed to solve BPDN, such as gradient projection for sparse reconstruction (GPSR) [8], Bregman iterative regularization [9], and sparse reconstruction by separable approximation (SpaRSA) [10]. When non-convex sparsity-inducing penalty is adopted in (2), it has been demonstrated that more accurate signals are tended to be derived [11–13]. However, the deficiency of multiple local minima in non-convex optimization limits its practical usage, where improper initialization might cause the solution trapped into the wrong ones. To the best of our knowledge, there is no rigorous theoretical analysis of algorithms for non-convex optimization (2) converging from the initialization to the sparse signal, which is the main motivation of our work.

Combining the concepts of sparseness measure [14] and weak convexity [15], a class of sparsity-inducing penalties is introduced in this paper with characterization of the non-convexity. A simple algorithm, which adopts the generalized gradient [16] as the update direction, is proposed to solve the non-convex optimization problem (2). Theoretical analysis shows that under some certain conditions, if the non-convexity of the penalty is below a threshold (which is in inverse proportion to the distance between the initialization and the sparse signal), the sparse signal can be stably recovered. Therefore, we can easily choose the initialization (e.g. the zero point) and determine the appropriate non-convexity of the penalty to guarantee convergence of the algorithm. Numerical simulations are implemented to show the influence of non-convexity on the recovery performance, and to compare the performance of this approach with other references.

2. PRELIMINARY

2.1. Null Space Constant

Several quantities are introduced in literatures to characterize the performance of sparse recovery problems and algorithms, e.g. mutual coherence [17], restricted isometry constant [18], and null space constant [19]. In this paper we adopt null space constant since it provides more tight conditions for sparse recovery than the other two [14]. Define $\mathbf{x}_S$ as the vector generated by setting the entries of $\mathbf{x}$ indexed by $S^c = \{1, 2, \ldots, N\} \setminus S$ to zeros.

Definition 1. *Define null space constant $\gamma(J, \mathbf{A}, K)$ as the smallest quantity such that*

$$J(\mathbf{z}_S) \le \gamma(J, \mathbf{A}, K) J(\mathbf{z}_{S^c}) \tag{3}$$

holds for any set $S \subset \{1, 2, \ldots, N\}$ with $\#S \le K$ and for any vector $\mathbf{z} \in \mathcal{N}(\mathbf{A})$, where $\mathcal{N}(\mathbf{A})$ denotes the null space of $\mathbf{A}$.

2.2. Weak Convexity

A real valued function $F(\cdot)$ defined on a convex subset $S \subseteq \mathbb{R}$ is ρ-convex if there exists some real number ρ which is the largest quan-

Parts of this work are presented in [1]. This work was partially supported by the National Program on Key Basic Research Project (973 Program 2013CB329201) and the National Natural Science Foundation of China (NSFC 61371137). The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

tity such that the inequality

$$F(\lambda t_1+(1-\lambda)t_2) \leq \lambda F(t_1)+(1-\lambda)F(t_2)-\rho\lambda(1-\lambda)(t_1-t_2)^2$$

holds for any $t_1, t_2 \in S$ and for any $\lambda \in [0,1]$. $\rho > 0$, $\rho = 0$ and $\rho < 0$ correspond to strong convexity, convexity and weak convexity, respectively. The following proposition reveals that $F(\cdot)$ can be decomposed into the sum of a convex function and a square.

Proposition 1. *(Proposition 4.3 from [15]). Function $F : S \to \mathbb{R}$ is ρ-convex if and only if there exists a convex function $H : S \to \mathbb{R}$ such that $F(t) = H(t) + \rho t^2$ for all $t \in S$.*

The concept of subgradient for convex (possibly nondifferentiable) functions can be generalized for weakly convex functions, and it is termed as generalized gradient [16]. Mathematically, for any $t \in \text{int}S$ which denotes the interior of S, the generalized gradient $f(t)$ of $F(\cdot)$ at point t is any element of the generalized gradient set $\partial F(t)$. It can be derived that $\partial F(t) = \partial H(t) + 2\rho t$, where $H(\cdot)$ is the convex function in the decomposition of $F(\cdot)$ and $\partial H(\cdot)$ is the subgradient set of $H(\cdot)$.

3. MAIN CONTRIBUTION

The main contributions of this paper are twofold. First, by exploiting the concepts of sparseness measure and weak convexity, a class of sparsity-inducing penalties is introduced. Second, the generalized gradient method is proposed to solve (2) with convergence guarantees from the initialization to the sparse signal, as is revealed in Theorem 1.

3.1. Sparsity-inducing Penalty

First, a class of sparsity-inducing penalties is introduced. The penalty $J(\cdot)$ is defined as

$$J(\mathbf{x}) = \sum_{i=1}^{N} F(x_i), \tag{4}$$

where $F(\cdot)$ is a weakly convex sparseness measure satisfying the following Definition 2.

Definition 2. *The weakly convex sparseness measure $F(\cdot)$ satisfies*
1) $F(0) = 0$, $F(\cdot)$ is even and not identically zero;
2) $F(\cdot)$ is non-decreasing on $[0, +\infty)$;
3) The function $t \mapsto F(t)/t$ is non-increasing on $(0, +\infty)$;
4) $F(\cdot)$ is a weakly convex function on $[0, +\infty)$.

Most commonly used non-convex sparsity-inducing penalties are formed by weakly convex sparseness measures, e.g. those penalties in [20–23]. The following lemma reveals some important properties of weakly convex sparseness measure. Define $\partial F(0) = \{0\}$.

Lemma 1. *The weakly convex sparseness measure $F(\cdot)$ satisfies the following properties:*

1) $F(\cdot)$ is continuous and there exists $\alpha > 0$ such that $F(t) \leq \alpha|t|$ holds for all $t \in \mathbb{R}$;

2) For any $\beta > 0$, $F(\beta t)$ is also a weakly convex sparseness measure, and its parameters are $\rho_\beta = \beta^2\rho$ and $\alpha_\beta = \beta\alpha$.

Proof. Please refer to Section VI-A in [1]. □

Besides ρ, the parameter α also plays an important role in characterizing the non-convexity of weakly convex sparseness measure $F(\cdot)$ or sparsity-inducing penalty $J(\cdot)$. Lemma 1.2) derives these parameters when the scale of the variable of $F(\cdot)$ varies.

3.2. Convergence Analysis of the Generalized Gradient Method

First, the generalized gradient method is proposed to solve the non-convex optimization problem (2). Mathematically, initialized as the zero point $\mathbf{x}(0) = \mathbf{0}$, this method iterates as

$$\mathbf{x}(n+1) = \mathbf{x}(n) - \kappa\nabla G(\mathbf{x}(n)), \tag{5}$$

where positive κ denotes the step size and $\nabla G(\mathbf{x}) = \nabla J(\mathbf{x}) + 2\lambda\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y})$ where $\nabla J(\mathbf{x})$ is a column vector whose ith element is the generalized gradient $f(x_i) \in \partial F(x_i)$.

Now let's turn to the convergence analysis of the generalized gradient method. Define $\sigma_{\min}(\mathbf{A})$ as the smallest nonzero singular value of $\mathbf{A}$. The following lemma is established for preparation.

Lemma 2. *For any $(J, \mathbf{A}, K)$ with $J(\cdot)$ formed by weakly convex sparseness measure $F(\cdot)$ and $\gamma(J, \mathbf{A}, K) < 1$, and for any positive constant M_0, if the regularization parameter $\lambda = C_1C_2\|\mathbf{e}\|_2^{-1}/2$, the inequality*

$$G(\mathbf{x}) - G(\mathbf{x}^*) \geq (C_1/3)\|\mathbf{x} - \mathbf{x}^*\|_2 \tag{6}$$

holds for all vectors $\mathbf{x}^*$ and $\mathbf{x}$ satisfying $\|\mathbf{x}^*\|_0 \leq K$, $\|\mathbf{x} - \mathbf{x}^*\|_2 \leq M_0$, and $\|\mathbf{x} - \mathbf{x}^*\|_2 \geq 3C_2\|\mathbf{e}\|_2$, where

$$C_1 = \frac{F(M_0)}{M_0}\frac{1-\gamma(J,\mathbf{A},K)}{1+\gamma(J,\mathbf{A},K)}, \quad C_2 = \frac{\alpha\sqrt{N}+C_1}{C_1\sigma_{\min}(\mathbf{A})}. \tag{7}$$

Proof. See Section 4.1. □

According to Lemma 2, if the gap between $G(\mathbf{x})$ and $G(\mathbf{x}^*)$ is small, $\mathbf{x}$ would not be far away from the sparse vector $\mathbf{x}^*$. Lemma 2 also gives a choice of the regularization parameter λ. As is shown, λ is in inverse proportion to the noise term $\|\mathbf{e}\|_2$, and the proportionality constant can be well approximated by $\alpha\sqrt{N}/(2\sigma_{\min}(\mathbf{A}))$ (which is easier to calculate) with relative error less than $N^{-\frac{1}{2}}$. The following Lemma 3 demonstrates the main result on the local minima of problem (2).

Lemma 3. *For any $(J, \mathbf{A}, K)$ with $J(\cdot)$ formed by weakly convex sparseness measure $F(\cdot)$ and $\gamma(J, \mathbf{A}, K) < 1$, and for any positive constant M_0, if the regularization parameter $\lambda = C_1C_2\|\mathbf{e}\|_2^{-1}/2$, the inequality*

$$(\mathbf{x} - \mathbf{x}^*)^{\mathrm{T}}\nabla G(\mathbf{x}) > 0 \tag{8}$$

holds for all vectors $\mathbf{x}^*$ and $\mathbf{x}$ satisfying $\|\mathbf{x}^*\|_0 \leq K$,

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \min\{M_0, C_1/(-4\rho)\}, \tag{9}$$

and $\|\mathbf{x} - \mathbf{x}^*\|_2 \geq 3C_2\|\mathbf{e}\|_2$, where C_1 and C_2 are specified as (7).

Proof. See Section 4.2. □

According to Lemma 3, for any local minimum $\mathbf{x}$ in the area of (9), it also satisfies $\|\mathbf{x} - \mathbf{x}^*\|_2 \leq 3C_2\|\mathbf{e}\|_2$. The following Lemma 4 summarizes the convergence performance of the generalized gradient method in a single iteration. For simplicity, let $\mathbf{x}$ and $\mathbf{x}^+$ represent $\mathbf{x}(n)$ and $\mathbf{x}(n+1)$, respectively.

Lemma 4. *For any $(J, \mathbf{A}, K)$ with $J(\cdot)$ formed by weakly convex sparseness measure $F(\cdot)$ and $\gamma(J, \mathbf{A}, K) < 1$, positive constant M_0, and vector $\mathbf{x}^*$ with $\|\mathbf{x}^*\|_0 \leq K$, if the regularization parameter $\lambda = C_1C_2\|\mathbf{e}\|_2^{-1}/2$ and the previous iterative solution $\mathbf{x}$ satisfies (9) and*

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \geq 6(d+1)C_1^{-1}\kappa + 3C_2\|\mathbf{e}\|_2, \tag{10}$$

where $d = \max_{\|\mathbf{x}-\mathbf{x}^*\|_2 \le M_0} \|\nabla G(\mathbf{x})\|_2^2$ is a finite quantity and C_1 and C_2 are specified as (7), the next iterative solution $\mathbf{x}^+$ satisfies

$$\|\mathbf{x}^+ - \mathbf{x}^*\|_2^2 \le \|\mathbf{x} - \mathbf{x}^*\|_2^2 - \kappa^2. \tag{11}$$

Proof. The proof is similar to that of Lemma 5 in [1], and it is omitted here to save space. □

According to Lemma 4, if $\mathbf{x}(n)$ lies within a neighborhood of the sparse signal $\mathbf{x}^*$ as (9), as long as the distance between $\mathbf{x}(n)$ and $\mathbf{x}^*$ is larger than a quantity linear in both the step size κ and the noise term $\|\mathbf{e}\|_2$, the next iterative solution $\mathbf{x}(n+1)$ will definitely get closer to $\mathbf{x}^*$. Therefore, in finite iterations, the solution will get into the $(O(\kappa) + O(\|\mathbf{e}\|_2))$-neighborhood of $\mathbf{x}^*$. To ensure that the generalized gradient method converges, we require the sufficient condition (9) satisfied for the initialization. We can simply choose

$$M_0 = \|\mathbf{x}(0) - \mathbf{x}^*\|_2 \le C_1/(-4\rho). \tag{12}$$

The following lemma reveals that penalties with small non-convexity will result in (12).

Lemma 5. *For any $(J, \mathbf{A}, K)$ with $J(\cdot)$ formed by weakly convex sparseness measure $F(\cdot)$ and $\gamma(J, \mathbf{A}, K) < 1$, and for any positive constant M_0, constraint (12) holds if the non-convexity of $J(\cdot)$ satisfies*

$$\frac{-\rho}{\alpha} \le \frac{1}{M_0} \frac{1 - \gamma(J, \mathbf{A}, K)}{5 + 3\gamma(J, \mathbf{A}, K)}. \tag{13}$$

Proof. Please refer to Section VI-H in [1]. □

In this paper $-\rho/\alpha$ is utilized to characterize the non-convexity, where $-\rho$ divided by α is to remove the scaling effect on the penalty. According to Lemma 1.2), the non-convexity of $J(\beta\mathbf{x})$ is

$$\frac{-\rho_\beta}{\alpha_\beta} = \beta \frac{-\rho}{\alpha} \tag{14}$$

for all $\beta > 0$. As a result, for any $J(\cdot)$ formed by weakly convex sparseness measure, we can always generate a non-convex penalty satisfying (13). Based on Lemma 4 and Lemma 5, the convergence of the generalized gradient method is guaranteed as follows.

Theorem 1. *(Convergence of the generalized gradient method) For any $(J, \mathbf{A}, K)$ with $J(\cdot)$ formed by weakly convex sparseness measure $F(\cdot)$ and $\gamma(J, \mathbf{A}, K) < 1$, and for any positive constant M_0, if the regularization parameter $\lambda = C_1 C_2 \|\mathbf{e}\|_2^{-1}/2$ and the non-convexity of $J(\cdot)$ satisfies (13), the recovered solution $\hat{\mathbf{x}}$ by the generalized gradient method satisfies*

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 \le 6(d+1)C_1^{-1}\kappa + 3C_2\|\mathbf{e}\|_2 \tag{15}$$

provided that $\|\mathbf{x}^*\|_0 \le K$ and $\|\mathbf{x}(0) - \mathbf{x}^*\|_2 \le M_0$, where $d = \max_{\|\mathbf{x}-\mathbf{x}^*\|_2 \le M_0} \|\nabla G(\mathbf{x})\|_2^2$ and C_1 and C_2 are specified as (7).

According to Theorem 1, under some certain conditions, if the non-convexity of the penalty is below a threshold, the generalized gradient method returns a stably recovered solution by choosing a sufficiently small step size κ. If $\rho = 0$, $J(\cdot)$ is just a scaled version of the ℓ_1 norm, and the condition (13) always holds. Therefore, no constraint needs to be imposed on the initialization. This is consistent in the fact that problem (2) is convex now and the initialization can be arbitrary.

3.3. Discussion

Our work mainly aims to provide the convergence guarantees of a non-convex approach for sparse recovery from the initialization to the sparse signal. Some other literatures also try to address this important problem. In [11], a family of non-convex penalties which can be decomposed as a difference of convex functions is introduced, and the problem is solved based on an iterative algorithm which solves at each iteration a convex weighted Lasso problem. Theoretical analysis reveals that only the convergence to a stationary point of the objective function is guaranteed. In [12], a sufficient condition which ensures that the sparse signal is a local minimizer of the sparse recovery problem is provided. Unlike our work, only "partial" convergence is theoretically guaranteed in those existing works.

4. PROOF

4.1. Proof of Lemma 2

Proof. Define $\mathbf{u} = \mathbf{x} - \mathbf{x}^*$ and decompose $\mathbf{u}$ by $\mathbf{u} = \mathbf{z} + \mathbf{z}^\perp$, where $\mathbf{z} \in \mathcal{N}(\mathbf{A})$ and $\mathbf{z}^\perp \in \mathcal{N}(\mathbf{A})^\perp$ which denotes the orthogonal complement of $\mathcal{N}(\mathbf{A})$. Recalling the definition of $G(\cdot)$,

$$G(\mathbf{x}) - G(\mathbf{x}^*) = J(\mathbf{x}) - J(\mathbf{x}^*) + \lambda\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 - \lambda\|\mathbf{e}\|_2^2. \tag{16}$$

On the one hand, according to the proof of Lemma 3 in [1],

$$J(\mathbf{x}) - J(\mathbf{x}^*) \ge C_1\|\mathbf{u}\|_2 - C_1C_2\|\mathbf{A}\mathbf{u}\|_2. \tag{17}$$

On the other hand, it can be calculated that

$$\lambda\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 - \lambda\|\mathbf{e}\|_2^2 \ge \lambda\|\mathbf{A}\mathbf{u}\|_2^2 - 2\lambda\|\mathbf{e}\|_2\|\mathbf{A}\mathbf{u}\|_2 \tag{18}$$

Therefore, (16), (17), and (18) implies that

$$\begin{aligned} G(\mathbf{x}) - G(\mathbf{x}^*) &\ge C_1\|\mathbf{u}\|_2 - \left(2\lambda\|\mathbf{e}\|_2 + C_1C_2\right)^2/(4\lambda) &(19)\\ &= C_1\|\mathbf{u}\|_2 - 2C_1C_2\|\mathbf{e}\|_2 \ge C_1\|\mathbf{u}\|_2/3, &(20)\end{aligned}$$

which completes the proof. □

4.2. Proof of Lemma 3

Proof. According to the definition of $\nabla G(\mathbf{x})$, the proof of Lemma 4 in [1], Lemma 2, the requirement (9), and the fact that

$$2(\mathbf{x} - \mathbf{x}^*)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y}) \ge \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 - \|\mathbf{A}\mathbf{x}^* - \mathbf{y}\|_2^2, \tag{21}$$

it can be derived that

$$\begin{aligned} (\mathbf{x} - \mathbf{x}^*)^{\mathrm{T}}\nabla G(\mathbf{x}) &\ge G(\mathbf{x}) - G(\mathbf{x}^*) + \rho\|\mathbf{x} - \mathbf{x}^*\|_2^2 \\ &\ge C_1\|\mathbf{x} - \mathbf{x}^*\|_2/12 > 0, &(22)\end{aligned}$$

which completes the proof. □

5. NUMERICAL SIMULATION

In this section two simulations are implemented to verify the theoretical analysis and to compare the proposed approach with other references. The sensing matrix $\mathbf{A}$ is of size $M = 200$ and $N = 1000$, whose entries are independently distributed Gaussian with zero mean and variance $1/M$. The locations of the non-zero entries of the sparse signal $\mathbf{x}^*$ are randomly chosen among all possible choices, and these non-zero entries independently satisfy standard normal distribution. The white noise $\mathbf{e}$ is also Gaussian with zero mean and variance adjusted according to the measurement SNR (MSNR). We

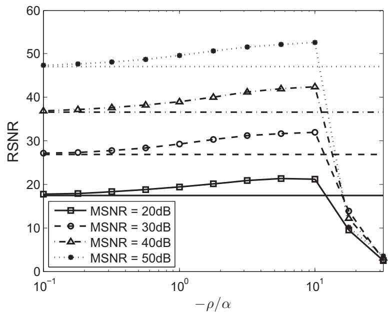

Fig. 1. The figure shows the recovery performance of the generalized gradient method versus non-convexity of the penalty when the sparsity $K = 30$. The performance with the ℓ_1 penalty is also plotted accordingly with the same types of lines as a benchmark.

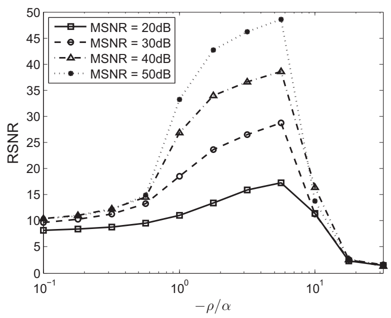

Fig. 2. The figure shows the recovery performance of the generalized gradient method versus non-convexity of the penalty when the sparsity $K = 60$.

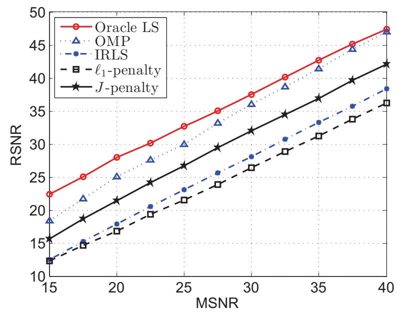

Fig. 3. The figure compares the recovery performance of different algorithms versus MSNR when $K = 30$.

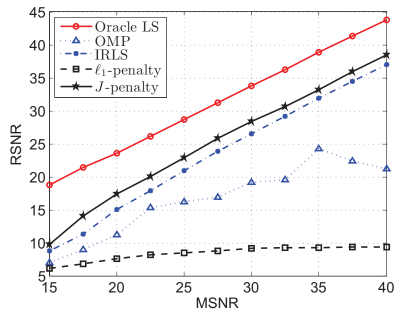

Fig. 4. The figure compares the recovery performance of different algorithms versus MSNR when $K = 60$.

adopt the non-convex sparsity-inducing penalty in [21] as $J(\cdot)$, and the regularization parameter $\lambda = (\alpha\sqrt{N}/(2\sigma_{\min}(\mathbf{A})))\|\mathbf{e}\|_2^{-1}$. The generalized gradient method is initialized with the zero point and with the step size $\kappa = 1 \times 10^{-5}$. The simulations are implemented in MATLAB on the Windows 7 operating system[1].

The first experiment tests the recovery performance of the generalized gradient method versus non-convexity of the penalty under different MSNR. The simulation is repeated 10 times to calculate the recovery SNR (RSNR) of the solution, and the results are plotted in Fig. 1 with the sparsity $K = 30$ and in Fig. 2 with $K = 60$. In Fig. 1, the performance with the ℓ_1 penalty is also plotted accordingly with the same types of lines as a benchmark. It is shown that as the non-convexity increases, the recovery performance improves at first, and degenerates rapidly when the non-convexity continues to grow. When the non-convexity approaches zero, the performance is close to that with the ℓ_1 penalty. The results in Fig. 2 reveal that when the sparsity is large, the sparse signals can only be stably recovered with moderate non-convexity. This is due to the fact that the sparse signal is no longer the minimum of the optimization problem when the non-convexity is small, and that the generalized gradient method fails to converge to the sparse signal when the non-convexity is large.

In the second experiment, the recovery performance of the generalized gradient method is compared with some typical algorithms, including orthogonal matching pursuit (OMP) [24], iteratively reweighted least squares (IRLS) [25], and the oracle least squares (LS) with the support known a priori. We denote J-penalty as the approach proposed in this paper with non-convexity $-\rho/\alpha = 10^{0.75}$, and ℓ_1-penalty when the ℓ_1 norm is adopted. The simulation is repeated 50 times to calculate the RSNR. Fig. 3 demonstrates the results when the sparsity $K = 30$. As can be seen, OMP is the best practical algorithm and its performance approaches that of oracle LS as MSNR increases. This is due to the fact that the support is likely to be recovered by OMP when the sparsity is small. J-penalty outperforms the rest algorithms. The results with $K = 60$ are plotted in Fig. 4, which shows that the performance of J-penalty is the best when the sparsity is large.

6. CONCLUSION

This work focuses on the performance of a non-convex approach for sparse recovery using regularized least squares. A class of non-convex sparsity-inducing penalties is introduced with characterization of the non-convexity. We prove that under some certain conditions, if the non-convexity is below a threshold, the sparse signal can be stably recovered by the generalized gradient method. Experiments show the superiority of utilizing non-convex penalties for sparse recovery, and the approach proposed in this paper enjoys the best recovery performance when the sparsity is large.

[1] The code for these simulations is available at
`http://gu.ee.tsinghua.edu.cn/publications#cl1`

7. REFERENCES

[1] L. Chen and Y. Gu, "The convergence guarantees of a nonconvex approach for sparse recovery," submitted to *IEEE Trans. Signal Processing*, arXiv:1211.7089v4 [cs.IT].

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[4] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33-61, Aug. 1998.

[5] J. Fuchs, "On sparse representations in arbitrary redundant basis," *IEEE Trans. Information Theory*, vol. 50, no. 6, pp. 1341-1344, June 2004.

[6] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Information Theory*, vol. 52, no. 1, pp. 6-18, Jan. 2006.

[7] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Information Theory*, vol. 52, no. 3, pp. 1030-1051, Mar. 2006.

[8] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586-597, Dec. 2007.

[9] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143-168, Mar. 2008.

[10] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2479-2493, July 2009.

[11] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Trans. Signal Processing*, vol. 57, no. 12, pp. 4686-4698, Dec. 2009.

[12] J. Lv and Y. Fan, "A unified approach to model selection and sparse recovery using regularized least squares," *Annals of Statistics*, vol. 37, no. 6A, pp. 3498-3528, Dec. 2009.

[13] J. Pant, W. Lu, and A. Antoniou, "Recovery of sparse signals from noisy measurements using an ℓ_p-regularized least-squares algorithm," *PacRim 2011*, pp. 48-53, Aug. 2011.

[14] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335-355, May 2007.

[15] J. Vial, "Strong and weak convexity of sets and functions," *Mathematics of Operations Research*, vol. 8, no. 2, pp. 231-259, May 1983.

[16] F. Clarke, "Generalized gradients and applications," *Trans. American Math. Society*, vol. 202, pp. 247-262, Apr. 1975.

[17] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Information Theory*, vol. 47, no. 7, pp. 2845-2862, Nov. 2001.

[18] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Information Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.

[19] A. Cohen, W. Dahmen, and R. Devore, "Compressed sensing and best k-term approximation," *Journal of the American Math. Society*, vo. 22, no. 1, pp. 211-231, Jan. 2009.

[20] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877-905, Dec. 2008.

[21] J. Jin, Y. Gu, and S. Mei, "A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 409-420, Apr. 2010.

[22] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395-407, May 2009.

[23] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Trans. Medical Imaging*, vol. 28, no. 1, pp. 106-121, Jan. 2009.

[24] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[25] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," *ICASSP 2008*, pp. 3869-3872, Apr. 2008.