FROM LEAST SQUARES TO SPARSE: A NON-CONVEX APPROACH WITH GUARANTEE

*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

This paper aims to provide theoretical guarantees via non-convex optimization for sparse recovery. It is shown that the sparse signal is the unique local optimal solution within a neighborhood, which contains the least squares solution if the sparsity-inducing penalties are not too non-convex. The idea of projected subgradient method is generalized to solve this non-convex optimization problem. A uniform approximate projection is applied in the projection step to make the algorithm more computationally tractable. The theoretical convergence analysis of the proposed method, approximate projected generalized gradient (APGG), is performed in the noisy scenario. The result reveals that if the non-convexity of the penalties is under a threshold, the bound of the recovery error is linear in both the noise bound and the step size. Numerical simulations are performed to test the performance of APGG and verify its theoretical analysis.

***Index Terms*—** Non-convex optimization, sparsity-inducing penalty, least squares solution, approximate projected generalized gradient, convergence analysis.

1. INTRODUCTION

Since the introduction of compressive sensing (CS) [2,3], sparse signal recovery has received much attention [4, 5]. Suppose one observes a group of linear measurements $\mathbf{y} \in \mathbb{R}^M$,

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

where $\mathbf{x}^* = (x_i^*) \in \mathbb{R}^N$ is an unknown K-sparse signal with $T = \{i | x_i^* \neq 0\}$ as its support set, and $\mathbf{A} \in \mathbb{R}^{M \times N}$ is a full row rank sensing matrix with more columns than rows. Numerous researches [6, 7] have shown that it may be sufficient to solve (1) by recasting it as a convex optimization problem

$$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{2}$$

The problem (2) is also known as *basis pursuit* (BP). It is certified [8] that under some certain conditions, the solution of (2) is identically the sparsest one. This conclusion greatly reduces the computational complexity, since (2) can be reformulated as a linear program (LP), and be solved by many efficient algorithms [9].

Later, another family of sparse recovery algorithms [10–16] is put forward based on non-convex optimization,

$$\min J(\mathbf{x}) \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{3}$$

where $J(\cdot)$ belongs to any sparsity-inducing penalties. It is theoretically proved [17–19] and experimentally verified [10–16, 18, 19] that (3) tends to derive the sparsest solution with looser conditions than (2), i.e., less measurements or more non-zero unknowns. However, the inherent deficiency of multiple local minima in non-convex optimization limits its practical usage, where improper initial criteria may cause the solution trapped into them.

The least squares (LS) solution $\mathbf{A}^\dagger\mathbf{y}$, where $\mathbf{A}^\dagger$ denotes the pseudo-inverse matrix of $\mathbf{A}$, has often been adopted to initialize the non-convex optimization based sparse recovery algorithms [10, 12, 14–16]. Therefore, a question, which naturally appears and mainly motivates this paper, is raised as follows.

> **Question**. *Initialized as the LS solution, does any algorithm guarantee to find the sparsest solution to the non-convex optimization (3)?*

In this paper, focusing on a class of *weakly convex* functions with parameter ρ less than zero [20], where larger $(-\rho)$ implies $J(\cdot)$ is more non-convex, the mentioned question is replied as follows.

> **Answer**. *The APGG method is proposed with guarantees that it converges from the LS solution to the sparsest solution provided that $(-\rho)$ is below a threshold.*

Specifically, this paper generalizes the idea of projected subgradient method [21, 22] to solve the optimization problem (3). For the class of sparsity-inducing penalties introduced in Section 2.1, their generalized gradients can be applied as the step direction. The initial LS solution and the projection step involve $\mathbf{A}^\dagger$, while exact calculation of pseudo-inverse may be computationally intractable or even impossible because of its large scale in practical applications. Thus we adopt a uniform approximate $\mathbf{A}^\dagger$, which greatly reduces the computational complexity of the method. We term it as *approximate projected generalized gradient* (APGG) method. The theoretical convergence analysis of APGG in the noisy scenario is demonstrated. It reveals that as long as $(-\rho)$ is below a threshold, i.e. the penalty $J(\cdot)$ is not too non-convex, the iterative solution will get into the neighborhood of the original sparse signal with radius linear in both the noise bound and the step size. In the noiseless scenario, for sufficiently small step size, the solution will approach the sparse signal with any precision.

Relation to prior work: Projected subgradient method [21,22] is a classical and typical approach to minimize a convex function with constraints. Later in [23], the inexact projections are adopted into these methods, but the inexact projections require approaching the exact projection in the course of the algorithm. APGG can be considered as a generalization of zero-point attracting projection (ZAP) algorithm [16], whose convex variant, ℓ_1-ZAP, has been thoroughly analyzed [24]. There is a somewhat similar sparse recovery algorithm termed SL0 [14]. Considering the penalties of SL0 don't belong to

More details including proofs for the lemmas and theorems in this paper can be found in [1]. The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

Table 1. Sparseness Measures with ρ and α_F in Definition 1

No.	$F(t)$	Param. Require.	ρ	α_F
1.	$\lvert t\rvert$	—	0	1
2.	$\dfrac{\lvert t\rvert}{(\lvert t\rvert+\sigma)^{1-p}}$	$0\le p<1$, $\sigma>0$	$(p-1)\sigma^{p-2}$	σ^{p-1}
3.	$1-\mathrm{e}^{-\sigma\lvert t\rvert}$	$\sigma>0$	$-\sigma^2/2$	σ
4.	$\ln(1+\sigma\lvert t\rvert)$	$\sigma>0$	$-\sigma^2/2$	σ
5.	$\mathrm{atan}(\sigma\lvert t\rvert)$	$\sigma>0$	$\dfrac{-3\sqrt{3}\sigma^2}{16}$	σ
6.	$(2\sigma\lvert t\rvert-\sigma^2t^2)\mathcal{X}_{\lvert t\rvert\le\frac{1}{\sigma}}$ $+\mathcal{X}_{\lvert t\rvert>\frac{1}{\sigma}}$	$\sigma>0$	$-\sigma^2$	2σ

those introduced in this paper, its parameter σ needs to be decreasing over iterations. Its convergence analysis is provided in [25] using a family of spline functions to approximate ℓ_0 norm. In [26], the solution sequence of IRLS [12] for ℓ_1 norm minimization is proved to converge to the sparse signal, while for ℓ_p norm with $p\in(0,1)$, a local convergence result is established without taking into account the initial solution.

2. SPARSENESS MEASURES AND THE APGG METHOD

In this section, the sparsity-inducing penalties and the APGG method are introduced. Please refer to [1] for detailed proofs of the results in this section and the next section.

2.1. Sparsity-inducing Penalties

First, a class of sparsity-inducing penalties is introduced so that APGG can be applied to solve (3). This class of penalties is quite general and covers many cost functions in sparse recovery literatures. The penalty $J(\mathbf{x})$ is defined as

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

where $F(\cdot)$ belongs to a class of sparseness measures satisfying the following Definition 1. The definitions and properties of ρ-convex function $F(\cdot)$ and its generalized gradient set $\partial F(\cdot)$ can be found in [20]. Define $\partial F(0)=\{0\}$.

Definition 1. *The function $F:\mathbb{R}\to\mathbb{R}$ satisfies the following properties:*

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 ρ-convex function on $[0,+\infty)$;
5) There exists a constant α_F such that for any $t\in(0,+\infty)$ and for any $f(t)\in\partial F(t)$, $|f(t)|\le\alpha_F$.

Some commonly used sparseness measures [13, 16] satisfying Definition 1 and their corresponding constants ρ and α_F are demonstrated in Table 1, where $\mathcal{X}_P$ denotes the indicator function of P.

It needs to be pointed out that Definition 1.1)-3) is almost the same as the definition of sparseness measures in [17], while two additional requirements are imposed so that APGG becomes applicable to solve (3). Most sparseness measures satisfying Definition 1.1)-3) also satisfy Definition 1.4)-5) as well. One exception is the function

$$F(t)=|t|^p \quad p\in[0,1), \tag{5}$$

which has been widely discussed in the literatures [10, 12]. It satisfies Definition 1.1)-3), but goes against the ρ-convexity and boundedness of its generalized gradient set, i.e. Definition 1.4)-5). However, approximations are made in some literatures to improve its robustness. For example, measure (5) is approximated by

$$F(t)=\frac{|t|}{(|t|+\sigma)^{1-p}} \tag{6}$$

with $\sigma>0$ [18]. The approximation satisfies Definition 1.4)-5), and its constants are shown in Table 1. This confirms that Definition 1.4)-5) are reasonable and are implicated assumptions when some robust algorithms or the theoretical analysis are taken into consideration.

Table 2. Approximate Projected Generalized Gradient Method

Initialize:	Calculate $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ as the approximate $\mathbf{A}^{\dagger}$, $n=0$, and $\mathbf{x}(0)=\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{y}$;
Repeat:	Generalized gradient step: $\tilde{\mathbf{x}}(n+1)=\mathbf{x}(n)-\kappa\nabla J(\mathbf{x}(n))$; Projection step: $\mathbf{x}(n+1)=\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{y}+(\mathbf{I}-\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A})\tilde{\mathbf{x}}(n+1)$; Iteration number increases by one: $n=n+1$;
Until:	Stop criterion satisfied;

2.2. Approximate Projected Generalized Gradient Method

The methods of computing a pseudo-inverse matrix have been developed to a mature technology [27, 28]. To save computation, especially in large scale problems, this paper adopts approximate $\mathbf{A}^{\dagger}$, which is assumed of the form $\mathbf{A}^{\mathrm{T}}\mathbf{B}$, i.e. the transpose of $\mathbf{A}$ multiplied by a matrix $\mathbf{B}\in\mathbb{R}^{M\times M}$. $\mathbf{B}$ can be considered as the approximation of $(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}$. To characterize the approximate precision of the pseudo-inverse matrix, define

$$\|\mathbf{I}-\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{B}\|_2\le\zeta, \tag{7}$$

and we assume that $\zeta<1$. The above assumptions are reasonable when iterative methods are adopted [28–30].

The APGG method is described in Table 2, where the initialization is an approximate LS solution, $\kappa>0$ denotes the step size and $\nabla J(\mathbf{x})$ is a column vector whose ith element is $f(x_i)\in\partial F(x_i)$. If $\zeta=0$, i.e. $\mathbf{A}^{\dagger}$ is exactly calculated, the method is termed *projected generalized gradient* (PGG). The following Theorem 1 demonstrates the effect of the approximate projection on the iterative solution. It reveals that for a fixed approximate precision $\zeta\in(0,1)$ and sufficiently small step size κ, as n approaches infinity, the iterative solution $\mathbf{x}(n)$ will approach the solution space at any given precision.

Theorem 1. *The iterative solution of APGG in the nth iteration, $\mathbf{x}(n)$, satisfies*

$$\|\mathbf{y}-\mathbf{A}\mathbf{x}(n)\|_2\le C_1\zeta^{n+1}+\frac{1}{2}C_2(\zeta)\kappa, \tag{8}$$

where $C_1=\|\mathbf{y}\|_2$ and $C_2(\zeta)=2\alpha_F\sqrt{N}\|\mathbf{A}\|_2\zeta/(1-\zeta)\xrightarrow{\zeta\to0}0$ are two constants.

3. MAIN CONTRIBUTIONS

In this section, the performance guarantees of APGG in the noisy scenario, i.e. $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$, are provided. The measurement noise $\mathbf{e}$ is assumed deterministic and bounded in this paper. Two lemmas are established for preparation. These lemmas are related with the optimization problem (3) and independent of specific recovery algorithms. Lemma 1 declares that if the sensing matrix $\mathbf{A}$ satisfies some certain conditions, the optimization problem (3) is locally stable. Lemma 2 reveals that if the difference between $J(\mathbf{x})$ and $J(\mathbf{x}^*)$ is small enough, $\mathbf{x}$ would not be far away from the desired solution $\mathbf{x}^*$, even though $\mathbf{x}$ does not necessarily lie in the solution space. Based on these lemmas, Theorem 2, Theorem 3, and Corollary 1 reveal that as long as $(-\rho)$ is below a threshold, the solution of APGG will definitely approach $\mathbf{x}^*$ with recovery error linear in both the noise bound and the step size. This is the Answer to the Question raised in the Introduction.

Lemma 1. *Let $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$ where $\|\mathbf{e}\|_2 \le \varepsilon$ and $\mathbf{x}^*$ is the desired K-sparse signal. Assume that the null space constant [31] $\gamma_J < 1$ for $J(\cdot)$ with a specific $F(\cdot)$ satisfying Definition 1. For any $\mathbf{x}$ satisfying $\mathbf{y} = \mathbf{A}\mathbf{x}$, $J(\mathbf{x}) \le J(\mathbf{x}^*)$ and $\|\mathbf{x} - \mathbf{x}^*\|_2 \le M_0$ where M_0 is a positive constant, there exists a positive constant C_3 such that*

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \le C_3\varepsilon, \tag{9}$$

and C_3 is independent of ε.

Lemma 2. *Let $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$ where $\|\mathbf{e}\|_2 \le \varepsilon$ and $\mathbf{x}^*$ is the desired K-sparse signal. Assume that the null space constant $\gamma_J < 1$ for $J(\cdot)$ with a specific $F(\cdot)$ satisfying Definition 1. For any $\mathbf{x}$ satisfying $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \eta$ and $2C_3(\varepsilon + \eta) \le \|\mathbf{x} - \mathbf{x}^*\|_2 \le M_0$, where C_3 is specified in Lemma 1 with the same M_0, there exists a uniform constant $c > 0$ such that*

$$J(\mathbf{x}) - J(\mathbf{x}^*) \ge c\|\mathbf{x} - \mathbf{x}^*\|_2. \tag{10}$$

The existence of the positive uniform constant c will play an important role in the theoretical convergence analysis. The following Theorem 2 demonstrates the convergence property of the APGG method in one iteration. For the convenience of theoretical analysis, define a constant N_κ such that $\forall n > N_\kappa$, $\|\mathbf{y} - \mathbf{A}\mathbf{x}(n)\|_2 \le C_2(\zeta)\kappa$. For simplicity, $\mathbf{x}$ and $\mathbf{x}^+$ represent $\mathbf{x}(n)$ and $\mathbf{x}(n+1)$, respectively.

Theorem 2. *Let $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$ where $\|\mathbf{e}\|_2 \le \varepsilon$ and $\mathbf{x}^*$ is the desired K-sparse signal. Assume that the null space constant $\gamma_J < 1$ for $J(\cdot)$ with a specific $F(\cdot)$ satisfying Definition 1. Suppose the previous iterative solution $\mathbf{x}$ satisfies $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le C_2(\zeta)\kappa$ and*

$$2C_3(\varepsilon + C_2(\zeta)\kappa) \le \|\mathbf{x} - \mathbf{x}^*\|_2 \le \min\{M_0, \frac{c}{-2\rho}\}, \tag{11}$$

where $C_2(\zeta)$, C_3, M_0, and c are constants specified in Theorem 1, Lemma 1, and Lemma 2, respectively. Further assume that

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \ge \frac{\mu}{c} d\kappa + C_4(\zeta)\kappa + C_5(\zeta)\varepsilon, \tag{12}$$

where $\mu > 1$ is arbitrary, $d = \max_{\mathbf{x}} \|(\mathbf{I} - \mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A})\nabla J(\mathbf{x})\|_2^2$, $C_4(\zeta) \xrightarrow{\zeta\to 0} 0$ and $C_5(\zeta) \xrightarrow{\zeta\to 0} 2\alpha_F\sqrt{N}\|\mathbf{A}\|_2\|\mathbf{B}\|_2/c$ are two constants. Then the next iterative solution $\mathbf{x}^+$ satisfies

$$\|\mathbf{x}^+ - \mathbf{x}^*\|_2^2 \le \|\mathbf{x} - \mathbf{x}^*\|_2^2 - (\mu - 1)d\kappa^2. \tag{13}$$

Since APGG adopts $\mathbf{x}(0) = \mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{y}$ as the initial solution, according to (11), one expects that

$$\|\mathbf{x}(0) - \mathbf{x}^*\|_2 \le \frac{c}{-2\rho}. \tag{14}$$

Then according to Theorem 2, the subsequent iterative solutions satisfy this constraint as well. The following theorem reveals that sparseness measures with appropriate ρ will result in (14).

Theorem 3. *For sparsity-inducing penalty $J(\cdot)$ with a specific sparseness measure $F(\cdot)$ satisfying Definition 1, consider a class of penalties*

$$J_\beta(\mathbf{x}) = \frac{1}{\beta} J(\beta\mathbf{x}), \quad \beta > 0. \tag{15}$$

If the parameters of $F(\cdot)$ are ρ and α_F, then the corresponding parameters of $F_\beta(\cdot)$ constituting $J_\beta(\cdot)$ are $\rho_\beta = \beta\rho$ and $\alpha_{F_\beta} = \alpha_F$. Furthermore, there exists a positive constant β_1 such that for any $\beta \in (0, \beta_1]$, the constraint (14) holds when penalty (15) is applied in (3).

Based on Theorem 2 and Theorem 3, the convergence guarantees of APGG are provided in the following Corollary 1.

Corollary 1. *Let $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$ where $\|\mathbf{e}\|_2 \le \varepsilon$ and $\mathbf{x}^*$ is the desired K-sparse signal. Assume that the null space constant $\gamma_J < 1$ for $J(\cdot)$ with a specific $F(\cdot)$ satisfying Definition 1. Suppose the initial solution $\mathbf{x}(0)$ is bounded, i.e., there exists a constant M_0 such that $\|\mathbf{x}(0) - \mathbf{x}^*\|_2 \le M_0$. Define $\rho_{\beta_1} = \beta_1\rho$, then for penalty (15) with parameter $(-\rho_\beta)$ below $(-\rho_{\beta_1})$, the iterative solution $\mathbf{x}(n)$ will get into the $(C_6(\zeta)\kappa + C_7(\zeta)\varepsilon)$-neighborhood of $\mathbf{x}^*$ in finite iterations, where $C_6(\zeta) = \max\{2C_3C_2(\zeta), \mu d/c + C_4(\zeta)\}$ and $C_7(\zeta) = \max\{2C_3, C_5(\zeta)\}$.*

Corollary 1 reveals that the approximate LS solution is a good choice as the initial solution, and the iterative solution of APGG will converge to the global optimal solution of (3). The parameter ρ reveals how non-convex the penalty could be. Large $(-\rho)$ implies more non-convexity of $J(\cdot)$, which results in better recovery performance but more difficulties in the initial solution selection. When $(-\rho)$ approaches infinity, the constraint (14) is so severe that the initial solution is almost impossible to be selected. According to Theorem 3 and Corollary 1, one would expect that there exists a positive parameter ρ^* such that the performance of APGG improves as $(-\rho) \in (0, -\rho^*)$ increases, and degenerates rapidly as $(-\rho) \in (-\rho^*, +\infty)$ continues to grow. As $(-\rho)$ approaches zero, the recovery performance tends to the case of $J(\cdot) = \|\cdot\|_1$. These results are further verified by the simulations in the next section.

4. NUMERICAL SIMULATIONS

In this section, three experiments are conducted to test the recovery performance of the APGG method and verify the theoretical analysis. The sensing matrix $\mathbf{A}$ is of size $N = 1000$ and $M = 200$, whose entries are independently and identically distributed Gaussian with zero mean and variance $1/M$. The locations of non-zero entries of the sparse signal $\mathbf{x}^*$ are randomly chosen among all possible choices. These non-zero entries are independently Gaussian distributed with zero mean and the same variance. The sparse signal is finally normalized to have unit energy. In all simulations, the approximate $\mathbf{A}^\dagger$ is calculated using the method in [29].

The first experiment tests the recovery performance of the PGG method in the noiseless scenario with different sparsity-inducing

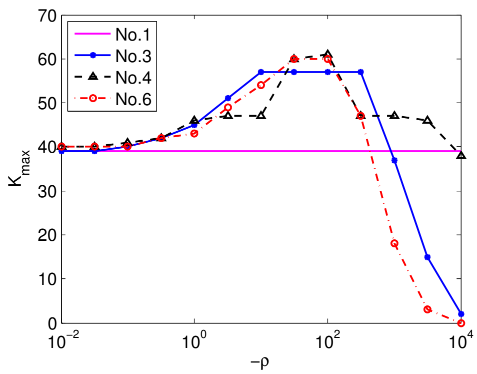

Fig. 1. The figure shows the recovery performance of the PGG method with sparsity-inducing penalties from Table 1.

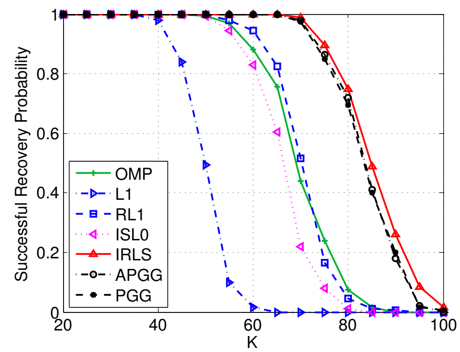

Fig. 2. The figure compares the successful recovery probability of different algorithms versus sparsity K. The approximate precision of inexact $\mathbf{A}^\dagger$ is $\zeta = 0.91$.

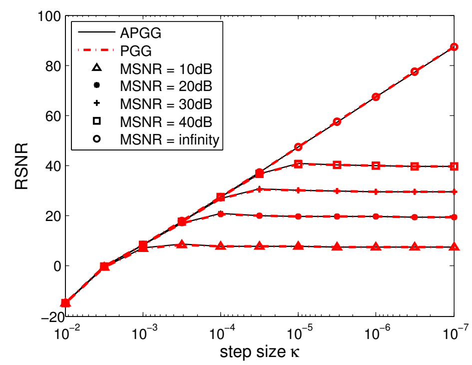

Fig. 3. The figure demonstrates the recovery precision of the APGG and PGG methods under different measurement noise and step size. The approximate precision of inexact $\mathbf{A}^\dagger$ is $\zeta = 0.22$.

penalties from Table 1. For each penalty with some certain ρ, the sparsity level K varies from 1 to 100 with increment of one. If the recovery SNR (RSNR) is higher than 40dB, this recovery is regarded as a success. The simulation is repeated 100 trials to calculate the successful recovery probability versus sparsity K. Then the crucial sparsity $K_{\max}$, which is the largest integer which guarantees 100% successful recovery, is recorded. The results are presented in Fig. 1. As is revealed, for the non-convex sparsity-inducing penalties, as $(-\rho)$ increases, the performance of PGG increases at first, and degenerates rapidly when $(-\rho)$ continues to grow. When $(-\rho)$ approaches zero, the performances of these penalties are close to that of the ℓ_1 norm. These results are consistent in the theoretical analysis of Theorem 3 and discussions after Corollary 1.

In the second experiment, the recovery performance of APGG is compared in the noiseless scenario with some typical sparse recovery algorithms, including OMP [32], the solution to BP [33], the solution to reweighted ℓ_1 minimization [13], ISL0 [15], and IRLS [12]. In the simulation, K varies from 20 to 100. The APGG method adopts the No. 6 sparseness measure in Table 1 with $\sigma = 10$, and the step size is set to 10^{-6}. The iteration number for calculating inexact pseudo-inverse matrices is 0, which means that $\varsigma\mathbf{A}^{\mathrm{T}}$ is adopted with precision $\zeta = 0.91$ [29]. For comparison, the performance of PGG is also plotted. The simulation is repeated 200 trials to calculate the successful recovery probability versus sparsity K, and the results are demonstrated in Fig. 2. As can be seen, APGG, PGG, and IRLS guarantee successful recovery for larger sparsity K than the other reference algorithms. In the noiseless scenario with sufficiently small step size, the inexact projection has little affect on the recovery performance of APGG.

In the third experiment, the recovery precision of the APGG method is simulated under different settings of noise bound and step size. The performance of PGG is also compared. In the simulation $K = 30$. The sparseness measure is the same as that in the previous experiment, and the iteration number for calculating inexact $\mathbf{A}^\dagger$ is 4 such that $\zeta = 0.22$. The simulation is repeated 100 trials to calculate the mean squared error (MSE), and the results are shown in Fig. 3. As can be seen, there is almost no difference between the performance of APGG and that of PGG. In the noisy scenario, the recovery SNR (RSNR) is dependent on both the measurement SNR (MSNR) and the step size. For fixed MSNR, as the step size decreases, the RSNR improves at first, and remains the same when the step size is sufficiently small. Larger MSNR results in larger RSNR limit. In the noiseless scenario, the RSNR improves as the step size decreases, and can be arbitrarily large by adopting sufficiently small step size. These results are accordant with Corollary 1, which implies that the recovery error is linear in both the noise bound and the step size.

5. CONCLUSION

This paper considers the guarantees that the non-convex optimization for sparse recovery finds the global optimal solution. Theoretical analysis reveals that there exists a neighborhood of $\mathbf{x}^*$ that contains no other local minima, while the commonly used initial LS solution lies in this neighborhood provided that the non-convexity of penalty $J(\cdot)$ is below a threshold. The APGG method is proposed to solve this optimization problem. It reveals that as long as the parameter $(-\rho)$ is below a threshold, the iterative solution will get into the neighborhood of $\mathbf{x}^*$ with radius linear in both the noise bound ε and the step size κ. Thus it is guaranteed that APGG converges from the LS solution to the global minimum. Simulation results verify the theoretical analysis, and the recovery performance of APGG is not much influenced by the approximate projection.

6. REFERENCES

[1] Laming Chen and Yuantao Gu, "A Non-convex approach for sparse recovery with convergence guarantee," submitted to *IEEE Trans. Signal Processing* for possible publication, available at http://arxiv.org/abs/1211.7089.

[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] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[5] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp 1182-1195, Dec. 2007.

[6] 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.

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

[8] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589-592, May 2008.

[9] S. Boyd and L. Vandenberghe, Convex Optimization, 2004: Cambridge Univ. Press.

[10] R. Chartrand, "Exact reconstructiono of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp.707-710, Oct. 2007.

[11] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 600-616, Mar. 1997.

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

[13] 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.

[14] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Trans. Signal Processing*, vol. 57, no. 1, pp. 289-301, Jan. 2009.

[15] M. Hyder and K. Mahata, "An improved smoothed ℓ^0 approximation algorithm for sparse representation," *IEEE Trans. Signal Processing*, vol. 58, no. 4, pp. 2194-2205, April 2010.

[16] 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, April 2010.

[17] 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.

[18] 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.

[19] R. Saab and Ö Yilmaz, "Sparse recovery by non-convex optimization - instance optimality," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 30-48, July 2010.

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

[21] A. Goldstein, "Convex programming in Hilbert space," *Bull. Amer. Math. Soc.*, vol. 70, no. 5, pp. 709-710, May 1964.

[22] Y. Alber, A. Iusem, and M. Solodov, "On the projected subgradient method for nonsmooth convex optimization in a Hilbert space," *Mathematical Programming*, vol. 81, no. 1, pp.23-35, Mar. 1998.

[23] D. Lorenz, M. Pfetsch, and A. Tillmann, "An infeasible-point subgradient method using adaptive approximate projections," avaiable online: http://arxiv.org/abs/1104.5351.

[24] X. Wang, Y. Gu, and L. Chen, "Proof of convergence and performance analysis for sparse recovery via zero-Point attracting projection," *IEEE Trans. Signal Processing*, vol. 60, no. 8, pp. 4081-4093, Aug. 2012.

[25] H. Mohimani, M. Babaie-Zadeh, I. Gorodnitsky, and C. Jutten, "Sparse recovery using smoothed ℓ^0 (SL0): convergence analysis," avaiable online: http://arxiv.org/abs/1001.5073.

[26] I. Daubechies, R. DeVore, M. Fornasier, and C. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1-38, Jan. 2010.

[27] N. Shinozaki, M. Sibuya, and K. Tanabe, "Numerical algorithms for the Moore-Penrose inverse of a matrix: direct methods," *Annals of the Institute of Statistical Mathematics*, vol. 24, no. 1, pp. 193-203, Dec. 1972.

[28] N. Shinozaki, M. Sibuya, and K. Tanabe, "Numerical algorithms for the moore-penrose inverse of a matrix: iterative methods," *Annals of the Institute of Statistical Mathematics*, vol. 24, no. 1, pp. 621-629, Dec. 1972.

[29] A. Ben-Israel and D. Cohen, "On iterative computation of generalized inverses and associated projections," *SIAM Journal on Numerical Analysis*, vol. 3, no. 3, pp. 410-419, Sept. 1966.

[30] A. Ben-Israel and T. Greville, Generalized Inverses, 2003: Springer, New York, NY, 2nd edition.

[31] 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.

[32] 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.

[33] CVX Research, Inc. CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx, September 2012.