

Article

Proximal Linearized Iteratively Reweighted Algorithms for Nonconvex and Nonsmooth Optimization Problem

Juyeb Yeo and Myeongmin Kang *

Department of Mathematics, Chungnam National University, Daejeon 34134, Korea; mapiano911@gmail.com
* Correspondence: myeongminkang@cnu.ac.kr

Abstract: The nonconvex and nonsmooth optimization problem has been attracting increasing attention in recent years in image processing and machine learning research. The algorithm-based reweighted step has been widely used in many applications. In this paper, we propose a new, extended version of the iterative convex majorization–minimization method (ICMM) for solving a nonconvex and nonsmooth minimization problem, which involves famous iterative reweighted methods. To prove the convergence of the proposed algorithm, we adopt the general unified framework based on the Kurdyka–Łojasiewicz inequality. Numerical experiments validate the effectiveness of the proposed algorithm compared to the existing methods.

Keywords: iterative reweighted algorithm; linearization; nonconvex optimization; nonsmooth objective function; Kurdyka–Łojasiewicz property

MSC: 65K10; 65K05; 68U10



Citation: Yeo, J.; Kang, M. Proximal Linearized Iteratively Reweighted Algorithms for Nonconvex and Nonsmooth Optimization Problem. *Axioms* **2022**, *11*, 201. <https://doi.org/10.3390/axioms11050201>

Academic Editor: Manuel Arana-Jimenez

Received: 20 March 2022

Accepted: 20 April 2022

Published: 24 April 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this paper, we consider the following nonconvex and nonsmooth optimization problem of a specific structure in a n -dimensional real vector space:

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) + p(x) + h(g(x)), \quad (1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is a proper, lower semicontinuous (l.s.c.), convex and continuously differentiable function which has the Lipschitz gradient with Lipschitz constant L_f , $p : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is a proper, l.s.c. and convex function, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a proper and l.s.c. function, and $h : \text{Image}(g) \rightarrow \mathbb{R}$ is continuously differentiable. Furthermore, we assume that the coordinate functions g_i of g are convex, and the function h has a strictly continuous gradient and is coordinate-wise nondecreasing, i.e., $h(x) \leq h(x + \lambda e_i)$, where $x, x + \lambda e_i \in \text{Image}(g)$ and $\lambda > 0$ with i -th standard basis vector e_i for $i = 1, \dots, m$. We also suppose that F is coercive, closed and definable in an σ -minimal structure. Several nonconvex optimization problems in image processing or signal processing have an objective function of the form (1). For example, a nonconvex and nonsmooth minimization problem for image denoising

$$\min_{u \in \mathbb{R}^{n \times m}} \frac{\alpha_1}{2} \|u - f\|_2^2 + \alpha_2 \|u - b\|_1 + \sum_i \log(1 + |\nabla u|_i)$$

has the form of the problem (1) whose objective function satisfies all assumptions. Here, $b \in \mathbb{R}^{n \times m}$ is an observed noisy image, u is a restored image, and α_1, α_2 are positive parameters. The following nonconvex minimization for compressive sensing problem is also an example of the proposed problem (1):

$$\min_{x \in \mathbb{R}^n} \frac{1}{2\rho} \sum_{i=1}^n \log(1 + \rho |x|_i^2) + \frac{\beta}{2} \|x\|_2^2 + \frac{\mu}{2} \|Ax - b\|_2^2,$$

where β, μ, ρ are positive parameters, A is a $m \times n$ matrix with $m \ll n$, and b is an observed signal. As we will see in the numerical experiments, we will apply the proposed method to this application.

Minimizing the sum of a finite number of given functions is an important issue of mathematical optimization research. For the minimizing of convex functions, many efficient algorithms were proposed with convergence analysis, such as gradient based method [1,2], the iterative shrinkage thresholding algorithm [3], proximal point method [4] and alternating minimization algorithm. On the other hand, it is difficult to prove the global convergence of an algorithm to solve the nonconvex minimization problem. Nevertheless, several algorithms for solving nonconvex minimization problems have been developed. Extensions in the nonconvex setting of many first-order algorithms have been proposed, such as the gradient method, proximal point method [5] and iterative shrinkage thresholding algorithm [6] for nonconvex optimization. Recently, Attouch et al. [7] extended the alternating minimization algorithm by adding a proximal term to minimize the nonconvex function. The iteratively reweighted ℓ_1 algorithm [8] was proposed for solving a nonconvex minimization problem in compressive sensing. The iteratively reweighted least square method [9] was also developed for the nonconvex ℓ_p norm-based model applied to the compressive sensing problem. Very recently, Ochs et al. [10] extended these algorithms in a framework of the iterative convex majorization–minimization method (ICMM) to solve nonsmooth and nonconvex optimization problems and gave the global convergence analysis.

The Kurdyka–Łojasiewicz (KL) inequality is key when proving the global convergence of algorithms for nonsmooth and nonconvex optimization problems. A function which satisfies the KL inequality is called a KL function. Smooth KL functions and nonsmooth KL functions were introduced in [11–14]. Almost objective functions of the minimization problem in image processing satisfy the KL inequality, and it is a very useful when we deal with nonconvex objective functions. Many methods [7,15,16] whose global convergence was given based on the KL inequality have been proposed. Recently, Ochs et al. [17] proposed an inertial proximal algorithm by combining forward–backward splitting with an inertial force. Attouch et al. [18] proposed a general framework for the global convergence of descent methods for minimizing the KL function. In this paper, we utilize this general framework for the convergence of the proposed method.

In general, a nonlinear optimization problem does not have a closed-form solution. Hence, iterative algorithms frequently are used to solve a nonlinear minimization problem. Several algorithms minimize alternatively a linear approximation of nonlinear differentiable objective function. This technique is called “linearization”. The linearization technique of continuously differentiable objective function was applied in many algorithms [3,19,20] to solve constrained or unconstrained optimization problems.

The ICMM [10] is a popular algorithm for solving the problem (1). In this article, we propose an extension of the ICMM to solve the nonconvex and nonsmooth minimization problem (1). The linearization of the convex differentiable function f is considered unlike the ICMM. This enables the proposed method to deal with more applications. Further details of applications are given in Section 3. Based on the general framework introduced in [18], we prove the global convergence of the proposed model. From numerical experiments, we demonstrate the superiority of the proposed method over the existing methods, which are presented in Section 4.

The rest of this paper is organized as follows. In Section 2, we present the mathematical preliminary for nonconvex optimization and introduce the ICMM. In Section 3, we propose an extended version of the ICMM, and the iterative reweighted algorithms, which are special examples of the ICMM, are also suggested. The global convergence of the proposed algorithm is proved. In Section 4, numerical experiments are provided for our method, with comparisons to state-of-the-art methods. Finally, Section 5 summarizes our work.

2. Background

In this section, we present the mathematical preliminary for our work, and the iterative convex majorization–minimization method [10] is introduced.

2.1. Mathematical Preliminary

In Section 2.1, we introduce basic mathematical concepts and properties. More details are given in [21,22].

The concept of the Lipschitz continuity of a function is important in mathematical optimization theory. In the problem (1), we consider a continuously differentiable function which is included in a class of functions, namely the function with the Lipschitz gradient. This class is denoted by $C^{1,1}$. A continuously differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ has a Lipschitz gradient with constant L if there exists $L \in [0, \infty)$ with

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2 \quad \text{for all } x, y \in \mathbb{R}^n.$$

There is a property for a function with the Lipschitz gradient: If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function whose gradient is Lipschitz continuous with Lipschitz constant L_f , i.e., $f \in C^{1,1}$, then, for any $L \geq L_f$, the following inequality holds:

$$f(x) \leq f(y) + \langle x - y, \nabla f(y) \rangle + \frac{L}{2}\|x - y\|_2^2, \quad \forall x, y \in \mathbb{R}^n. \tag{2}$$

Now we introduce the generalized subdifferentials for a nonsmooth and nonconvex function.

Definition 1. For a function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ and a point $\bar{x} \in \text{dom } f$,

- ω is a regular subgradient of f at \bar{x} , denoted by $\omega \in \hat{\partial}f(\bar{x})$, if

$$\liminf_{\substack{x \rightarrow \bar{x} \\ x \neq \bar{x}}} \frac{1}{\|x - \bar{x}\|_2} (f(x) - f(\bar{x}) - \langle x - \bar{x}, \omega \rangle) \geq 0.$$

- v is a (limiting) subgradient of f at \bar{x} , denoted by $v \in \partial f(\bar{x})$, if there are sequences $x^v \rightarrow \bar{x}$, with $f(x^v) \rightarrow f(\bar{x})$, $v^v \rightarrow v$, $v^v \in \hat{\partial}f(x^v)$.

A trivial property of the subdifferential is that $\hat{\partial}f(x) \subseteq \partial f(x)$. Moreover, there are many important properties of the subdifferential. These properties will be used to prove the convergence of the proposed method.

Proposition 1 (subgradient properties). Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be a function.

1. If f is convex, the regular and limiting subdifferentials are same sets, which is the subdifferential in the convex analysis:

$$\{\omega | f(y) - f(x) \geq \langle \omega, y - x \rangle, \forall y \in \mathbb{R}^n\}.$$

2. If f is continuously differentiable on a neighborhood of \bar{x} , then $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.
3. If g is a l.s.c. function and f is continuously differentiable on a neighborhood of \bar{x} , then we can obtain the subdifferential of $f + g$ as follows:

$$\hat{\partial}(f + g)(x) = \nabla f(x) + \hat{\partial}(g)(x), \quad \partial(f + g)(x) = \nabla f(x) + \partial(g)(x).$$

4. If a proper and l.s.c. function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ has a local minimum at \bar{x} , then $\mathbf{0} \in \hat{\partial}f(\bar{x}) \subseteq \partial f(\bar{x})$. Furthermore, if f is convex, then this condition is also sufficient for a global minimum.

For obtain the global convergence of the proposed algorithm, the objective function in the problem (1) must be a Kurdyka–Łojasiewicz function.

Definition 2. A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ satisfies the Kurdyka–Łojasiewicz (KL) property at a point $x^* \in \text{dom } \partial f$ if there exist $\eta \in (0, +\infty]$, neighborhood U of x^* , and a continuous and convex function $\phi : [0, \eta) \rightarrow \mathbb{R}_+$ which satisfy

- (i) $\phi(0) = 0$;
- (ii) ϕ is differentiable in $(0, \eta)$;
- (iii) $\phi(s) > 0$ for all $s \in (0, \eta)$;
- (iv) For any $x \in U \cap \{x \in \mathbb{R} : f(x^*) < f(x) < f(x^* + \eta)\}$,

$$\text{dist}(0, \partial f(x))\phi'(f(x) - f(x^*)) \geq 1.$$

A function $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is called Kurdyka–Łojasiewicz (KL) function if F satisfies the KL property at every point in $\text{dom } \partial f$.

Although the KL property is a strong condition, many functions are KL functions. The semialgebraic functions [23] are typical examples of KL functions. The set of semialgebraic functions involves polynomials, indicator functions of semialgebraic sets [23], and the ℓ_2 norm. The compositions, finite sums, and finite products of semialgebraic functions are also semialgebraic. However, log and exponential functions are not semialgebraic functions. Recently, a class of definable functions in the log-exp σ -minimal structure [24,25] was proposed, which involves log and exponential functions, and all semialgebraic functions. Moreover, it is proved that the functions in this class are KL functions. More details of the log-exp σ -minimal structure are given in [24,25].

Lastly, we recall the general framework for an iterative algorithm when the objective function in an unconstrained minimization problem is a KL function. We consider a nonconvex unconstrained minimization problem:

$$\min_{x \in \mathbb{R}^n} F(x), \tag{3}$$

where F is a l.s.c. and proper function. Attouch et al. [18] suggested a framework for the convergence of an iterative method to solve the nonconvex minimization problem (3). They proved that a given algorithm converges when the following three conditions hold: Let $\{x^k\}_{k \in \mathbb{N}}$ be a sequence generated by a given iterative algorithm.

Hypothesis 1 (sufficient decrease condition). *There exists a positive value $a > 0$ such that for each $k \in \mathbb{N}$,*

$$F(x^{k+1}) + a\|x^{k+1} - x^k\|_2^2 \leq F(x^k).$$

Hypothesis 2 (relative error condition). *For each $k \in \mathbb{N}$, there exists a sequence $w^{k+1} \in \partial F(x^{k+1})$ such that*

$$\|w^{k+1}\|_2 \leq b\|x^{k+1} - x^k\|_2,$$

where b is a fixed positive constant.

Hypothesis 3 (continuity condition). *There exists a subsequence $\{x^{k_j}\}_{j \in \mathbb{N}}$ and \tilde{x} such that*

$$x^{k_j} \rightarrow \tilde{x} \text{ and } F(x^{k_j}) \rightarrow F(\tilde{x}), \text{ as } j \rightarrow \infty.$$

The following theorem is the convergence result, and the proof is given in ([18], Theorem 2.9).

Theorem 1. *Let $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be a proper l.s.c. function. We consider a sequence $\{x^k\}_{k \in \mathbb{N}}$ that satisfies Hypotheses 1–3. If F has the KL property at the cluster point \tilde{x} specified in Hypothesis 3, then the sequence $\{x^k\}_{k \in \mathbb{N}}$ converges to $\bar{x} = \tilde{x}$ as k goes to infinity, and \bar{x} is a critical point of F . Moreover, the sequence $\{x^k\}_{k \in \mathbb{N}}$ has a finite length, i.e.,*

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|_2 < \infty$$

2.2. Iterative Convex Majorization–Minimization Method

This section recalls the iterative convex majorization–minimization method (ICMM) [10] for solving the following nonconvex minimization problem:

$$\min_{x \in \mathbb{R}^n} G(x) := p(x) + h(g(x)), \tag{4}$$

where $G : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is proper, l.s.c., and bounded below. The further assumptions are required; $p : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is proper, l.s.c., and convex. $g : \mathbb{R}^n \rightarrow X_2$ is coordinatewise convex. $h : g(X_2) \rightarrow \mathbb{R} \cup \{\infty\}$ is coordinatewise nondecreasing.

The ICMM to solve this nonconvex problem (4) is a famous iterative algorithm. This algorithm is an adopted majorization–minimization technique that chooses a suitable family of convex surrogate functions called majorizers, and it minimizes a convex majorizer function instead of the objective function G at each iteration. The specific algorithm is summarized in Algorithm 1.

Algorithm 1 Iterative Convex Majorization–Minimization Method (ICMM).

Initialization Choose a starting point $x^0 \in \mathbb{R}^n$ with $G(x^0) < \infty$ and define a suitable family of convex surrogate functions $(h^x)_{x \in \mathbb{R}^n}$ such that for all $x \in \mathbb{R}^n, h^x \in \mathcal{H}(x)$ holds, where

$$\mathcal{H}(x) := \left\{ q : \text{Im}(g) \rightarrow \mathbb{R} \cup \{\infty\} \mid \begin{array}{l} q \text{ is proper, convex,} \\ q \text{ is nondecreasing on } \text{Im}(g), \\ q(g(x)) = h(g(x)), \\ q(y) \geq h(y), \text{ for } y \in \text{Im}(g) \end{array} \right\}.$$

repeat

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} p(x) + h^{x^k}(g(x))$$

until The algorithm satisfies a stopping condition

The convergence of Algorithm 1 was studied in [10]. Additional conditions are required for the global convergence of the ICMM. First, h should have a locally Lipschitz continuous gradient on a compact set B containing all x^k , and majorizers h^x should have globally Lipschitz continuous gradients on B for all $x \in \mathbb{R}^n$, with a uniform Lipschitz constant. Another condition is that $p + h^x \circ g$ should be strongly convex, which is a stronger condition. To show the global convergence of the ICMM, it was proved that the objective function G satisfies the three Hypotheses 1–3 for any KL function G , which were introduced in the previous section. Then, Theorem 1 is applied. As examples of the ICMM, several iteratively reweighted convex algorithms were introduced in [10], such as the iterative reweighted ℓ_1 algorithm, iteratively reweighted tight convex algorithm, iteratively reweighted Huber algorithm and iteratively reweighted least squares algorithm.

3. Proposed Algorithm

3.1. Proximal Linearized Iteratively Convex Majorization–Minimization Method

In this section, we propose a novel algorithm for solving the nonconvex and nonsmooth minimization problem (1). The ICMM in Algorithm 1, introduced in the previous section, can be applied to the problem (1), since $f(x) + p(x)$ is also a proper, l.s.c, convex function. This yields the following iteration:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} f(x) + p(x) + h^{x^k}(g(x)). \tag{5}$$

In many cases, this problem does not have a closed form solution and we cannot compute the exact solution of (5). Since the convergence of the ICMM is only guaranteed under the assumption that the subproblem (5) is solved exactly, it is not applicable to many problems. To overcome this fatal drawback, an inexact stopping criterion of the subproblem (5) was also proposed in [26]. Specifically, solving the subproblem (5) requires an inner algorithm such as the iterative shrinkage thresholding algorithm [27], and fast iterative shrinkage thresholding algorithm [3]. However, it is often time consuming for solving a large-scale problem. Therefore, we extend the ICMM by adopting a linearization technique of f . Here, we consider the linear approximation of f at the k -th iterate x^k with an additional proximal term instead of $f(x)$:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{\alpha}{2} \|x - x^k\|^2.$$

Utilizing this technique, we propose the following minimization of a convex surrogate function:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \langle \nabla f(x^k), x - x^k \rangle + p(x^k) + h^{x^k}(g(x)) + \frac{\alpha}{2} \|x - x^k\|_2^2. \tag{6}$$

where $\alpha > \frac{L_f}{2}$ is a proximal parameter. The proposed algorithm is summarized in Algorithm 2 and it is called as proximal linearized ICMM (PL-ICMM).

Algorithm 2 Proximal Linearized Iteratively Convex Majorization–Minimization Method (PL-ICMM).

Conditions

- f is differentiable and has Lipschitz gradient with Lipschitz constant L_f .
- p is proper, convex, and l.s.c.
- $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is l.s.c. and convex.
- $h : \text{Im}(g) \rightarrow \mathbb{R}$ is differentiable.
- F is bounded below.

Initialization Choose a starting point $x^0 \in \mathbb{R}^n$ with $F(x^0) < \infty$ and define a suitable family of convex surrogate functions $(h^x)_{x \in \mathbb{R}^n}$ such that for all $x \in \mathbb{R}^n$, $h^x \in \mathcal{H}(x)$ holds, where

$$\mathcal{H}(x) := \left\{ q : \text{Im}(g) \rightarrow \mathbb{R} \cup \{\infty\} \mid \begin{array}{l} q \text{ is proper, convex,} \\ q \text{ is nondecreasing on } \text{Im}(g), \\ q(g(x)) = h(g(x)), \\ q(y) \geq h(y), \quad \forall y \in \text{Im}(g) \end{array} \right\}.$$

repeat
Solve

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \langle \nabla f(x^k), x - x^k \rangle + p(x) + h^{x^k}(g(x)) + \frac{\alpha}{2} \|x - x^k\|_2^2.$$

until The algorithm satisfies a stopping condition.

The proposed method can be regarded as a generalized version of the ICMM and is more applicable than the ICMM. For examples, the PL-ICMM can be directly applied to the following minimizations for a regression problem, while the ICMM cannot be applied:

$$\min_x \frac{1}{2} \|b - Ax\|_2^2 + \mu \sum_{i=1}^N (|x_i| + \epsilon)^q,$$

and

$$\min_x \frac{1}{2} \|b - Ax\|_2^2 + \mu \sum_{i=1}^n \log(1 + |x_i|),$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\epsilon, \mu > 0$, and $0 < q < 1$.

Many iteratively reweighted algorithms [28–30] are examples of the ICMM, which use a weighted function appropriately to serve a convex majorizer in the ICMM. A convex majorizer h^{x^k} has the form $h^{x^k}(x) = \langle w^{x^k}, \bar{h}(x) \rangle$ for some given $\bar{h}(x)$. The weight w^{x^k} must be selected such that the function h^{x^k} satisfies the conditions of the class of convex majorizers. Similar to the ICMM, the PL-ICMM is a general algorithm which includes lots of iteratively reweighted algorithms. We also introduce proximal linearized versions of iteratively reweighted algorithms.

Algorithm 3 Proximal linearized iteratively reweighted ℓ_1 algorithm (PL-IRL1).

Conditions

- f is differentiable and has Lipschitz gradient with Lipschitz constant L_f .
- p is proper, convex, and l.s.c.
- $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is l.s.c. and convex.
- $h : \text{Im}(g) \rightarrow \mathbb{R}$ is differentiable, **concave**.
- F is bounded below.

Initialization Choose a starting point $x^0 \in \mathbb{R}^n$ with $F(x^0) < \infty$.

repeat

$$w^{x^k} = \nabla h(g(x^k)).$$

Solve

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \langle \nabla f(x^k), x - x^k \rangle + p(x) + \langle w^{x^k}, g(x) \rangle + \frac{\alpha}{2} \|x - x^k\|_2^2.$$

until The algorithm satisfies a stopping condition

First, we propose proximal linearized iteratively reweighted ℓ_1 algorithm (PL-IRL1). We further assume that the function h is concave on $\text{Im}(g)$. For a concave function h , we can define the limiting supergradient of h as an element of $-\partial(-h)$. The set of all limiting supergradients of h is denoted by $\bar{\partial}(h)$. Since $-h$ is convex and differentiable, $\bar{\partial}(h)(x)$ on $\text{Im}(g)$ has only one element $\nabla h(x)$ from the property 2 in Proposition 1. The PI-IRL1 considers the majorizer $h^{x^k}(y) = \langle \nabla h(g(x^k)), y - g(x^k) \rangle$ and minimizes iteratively the following convex problem:

$$\min_x \langle \nabla f(x^k), x - x^k \rangle + p(x) + \langle \nabla h(g(x^k)), g(x) - g(x^k) \rangle + \frac{\alpha}{2} \|x - x^k\|_2^2.$$

In ([10], Proposition 2), it was proved that the majorizer $\langle \nabla h(g(x^k)), y - g(x^k) \rangle + h(g(x^k))$ is the optimal majorizer of $h \circ g$ at x^k . The PL-IRL1 is summarized in Algorithm 3.

Remark 1. We assume that p is also continuously differentiable and has a Lipschitz continuous gradient with Lipschitz constant L_p , and h is additively separable, i.e.,

$$h(y_1, \dots, y_m) = h_1(y_1) + \dots + h_m(y_m).$$

Then, the proximal iteratively reweighted algorithm [31] can be applied to the problem (1), which is the following iterative algorithm:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \langle \nabla f(x^k) + \nabla p(x^k), x - x^k \rangle + \langle w^{x^k}, g(x) \rangle + \frac{\alpha}{2} \|x - x^k\|_2^2,$$

where $w^{x^k} = \nabla h(g(x^k))$, and $\alpha > \frac{L_f + L_p}{2}$ is a parameter. Hence, our PL-IRL1 can be also regraded as an extension of the proximal iteratively reweighted algorithm.

Due to the optimality of the iterative reweighted ℓ_1 algorithm, it has been used frequently to solve many applications. However, it cannot be applied to solve the problem (1) when h is not concave on $Im(g)$, such as $h(|y|) = \log(1 + |y|^2)$. For nonconcave function h , the iterative reweighted least square algorithm (IRLS) is well known. For a proximal linearized version of IRLS, we need more assumptions that h is additively separable on \mathbb{R}_+^n and each separable function $h_j(y_j)$ is convex in $[0, r_j]$ and concave in $[r_j, +\infty)$ for some $r_j > 0$. The IRLS makes use of a convex majorizer

$$h^{x^k}(y) = \langle w^{x^k}, \frac{1}{2}y^2 \rangle,$$

where the weights w^{x^k} are given as $w_i^{x^k} = \frac{(\nabla h(y))_i}{y_i}$ and the the square in y^2 means the coordinatewise square operation. This yields the following iterative algorithm:

$$\min_x \langle \nabla f(x^k), x - x^k \rangle + p(x) + \langle w^{x^k}, \frac{1}{2}g(x)^2 \rangle + \frac{\alpha}{2} \|x - x^k\|_2^2. \tag{7}$$

The specific algorithm is given in Algorithm 4.

Algorithm 4 Proximal Linearized Iterative Reweighted Least Square Algorithm (PL-IRLS).

Conditions

- f is differentiable and has Lipschitz gradient with Lipschitz constant L_f .
- p be proper, convex, and l.s.c.
- $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be l.s.c. and convex.
- h is **additively separable on \mathbb{R}_+^n , i.e., $h(x_1, \dots, x_m) = h_1(x_1) + \dots + h_m(x_m)$, and each h_j is convex in $[0, r_j]$ and concave in $[r_j, \infty)$ for some $r_j > 0$.**

Initialization Choose a starting point $x^0 \in \mathbb{R}^n$ with $F(x^0) < \infty$.

repeat

$$y = g(x^k).$$

$$w_i^{x^k} = \frac{(\nabla h(y))_i}{y_i}, i = 1, \dots, m.$$

Solve

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \langle \nabla f(x^k), x - x^k \rangle + p(x) + \left\langle w^{x^k}, \frac{1}{2}(g(x))^2 \right\rangle + \frac{\alpha}{2} \|x - x^k\|_2^2.$$

until The algorithm satisfies a stopping condition

The majorization property of PL-IRLS can be also obtained from ([10], Proposition 23), when $h(y)$ has the form $h(y) = \log(1 + \rho y^2)$ for any $\rho > 0$.

3.2. Convergence Analysis of the PL-ICMM

First, we prove the partial convergence of the PL-ICMM. From the following proposition, the sequence $\{x^k\}$ generated by Algorithm 2 induces the convergence of the sequence $\{F(x^k)\}$ of objective function value at x^k .

Proposition 2. Let $\{x^k\}_{k \in \mathbb{N}}$ be generated by Algorithm 2, and let $h^x \in \mathcal{H}(x)$ for all $x \in \mathbb{R}^n$. If $\alpha > L_f$, then the sequence $\{F(x^k)\}_{k \in \mathbb{N}}$ monotonically decreases and hence converges.

Proof. Let F be bounded below by L . We can obtain

$$\begin{aligned} L &\leq F(x^{k+1}) = f(x^{k+1}) + p(x^{k+1}) + h(g(x^{k+1})), \\ &\leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{\alpha}{2} \|x^{k+1} - x^k\|_2^2 + p(x^{k+1}) + h^{x^k}(g(x^{k+1})), \\ &\leq f(x^k) + p(x^k) + h^{x^k}(g(x^k)), \\ &= f(x^k) + p(x^k) + h(g(x^k)) = F(x^k), \end{aligned}$$

where the second inequality is obtained from the property (2) and the majorization property of h^{x^k} , the third inequality is obtained from the optimality of the subproblem in Algorithm 2 and the last inequality is obtained from the majorization property of h^{x^k} . The sequence $\{F(x^k)\}_{k \in \mathbb{N}}$ decreases and is bounded from below. Hence, it converges. \square

We can also obtain the local convergence of the PL-ICMM, given in the following proposition.

Proposition 3. *Let F be coercive. Then the sequence $\{x^k\}_{k \in \mathbb{N}}$ is bounded and has at least one accumulation point.*

Proof. By Proposition 2, the sequence $\{F(x^k)\}$ is monotonically decreasing, and therefore the sequence $\{x^k\}$ is contained in the level set

$$\mathcal{L}(x^0) := \{x \in \mathbb{R}^n : F(x) \leq F(x^0)\}.$$

From the coercivity of F , we conclude the boundedness of the set $\mathcal{L}(x^0)$. By the Bolzano–Weierstrass theorem, $\{x^k\}_{k \in \mathbb{N}}$ has at least one accumulation point. \square

Now, we prove the global convergence of the proposed algorithm. We utilize the general framework for the convergence of an iterative method, introduced in [18]. Specifically, we prove the three Hypothesis 1–3 with $F(x) = f(x) + p(x) + h(g(x))$ in the problem (1) and the sequence $\{x^k\}$ generated by the PL-ICMM. As a result, we obtain the global convergence of the PL-ICMM by using Theorem 1. We further assume the following:

- h has a locally Lipschitz gradient on a compact set containing the sequence $\{g(x^k)\}$, and majorizers h^x have a Lipschitz gradient on a compact set containing the sequence $\{g(x^k)\}$ with a common Lipschitz constant.

To prove the global convergence of the PL-ICMM, we need the following lemma which shows the subdifferential calculus about the composition and summation of two functions. The proof of this lemma is provided in ([10], Lemma 1).

Lemma 1. *Under the given conditions for PL-ICMM, the following holds for all $x^* \in \mathbb{R}^n$:*

1. For all $x \in \mathbb{R}^n$,

$$\begin{aligned} \partial(h^{x^*} \circ g)(x) &= \partial \langle \nabla h^{x^*}, g \rangle(x), \\ \partial(h \circ g)(x) &= \partial \langle \nabla h, g \rangle(x). \end{aligned}$$

2. For all $x \in \mathbb{R}^n$,

$$\begin{aligned} \partial(p + h^{x^*} \circ g)(x) &= \partial p(x) + \partial(h^{x^*} \circ g)(x), \\ \partial(p + h \circ g)(x) &= \partial p(x) + \partial(h \circ g)(x). \end{aligned}$$

First we prove the sufficient decreasing condition for the proposed method in the following proposition.

Proposition 4 (Sufficient decreasing conditions). Let $\alpha > \frac{L_f}{2}$. There exists a $a > 0$ such that for all $k \in \mathbb{N}$

$$F(x^{k+1}) \leq F(x^k) - \frac{a}{2} \|x^k - x^{k+1}\|_2^2.$$

Proof. From the property (2) of $C^{1,1}$ functions, we have

$$f(x^{k+1}) - f(x^k) \leq \langle x^{k+1} - x^k, \nabla f(x^k) \rangle + \frac{L_f}{2} \|x^{k+1} - x^k\|_2^2. \tag{8}$$

By the definition of the subdifferential of a convex function, we can obtain

$$p(x^{k+1}) - p(x^k) \leq \langle \zeta_p, x^{k+1} - x^k \rangle, \tag{9}$$

$$h^{x^k}(g(x^{k+1})) - h^{x^k}(g(x^k)) \leq \langle \zeta_h^k, x^{k+1} - x^k \rangle, \tag{10}$$

where $\zeta_p \in \partial p(x^{k+1})$ and $\zeta_h^k \in \partial(h^{x^k} \circ g(x^{k+1}))$ are subderivatives of p and $h^{x^k} \circ g$ at x^{k+1} , respectively. Since x^{k+1} is the minimizer of the problem

$$\arg \min_x \langle \nabla f(x^k), x - x^k \rangle + p(x) + h^{x^k} \circ g(x) + \frac{\alpha}{2} \|x - x^k\|_2^2,$$

there exist the subderivatives $\zeta_p \in \partial p(x^{k+1})$ and $\zeta_h^k \in \partial(h^{x^k} \circ g(x^{k+1}))$ by Lemma 1 such that

$$\nabla f(x^k) + \zeta_p + \zeta_h^k + \alpha(x^{k+1} - x^k) = 0. \tag{11}$$

From the facts that $h \circ g(x^{k+1}) \leq h^{x^k} \circ g(x^{k+1})$ and $h \circ g(x^k) = h^{x^k} \circ g(x^k)$, we obtain that for all $k \in \mathbb{N}$,

$$\begin{aligned} F(x^{k+1}) - F(x^k) &= (f(x^{k+1}) + p(x^{k+1}) + h(g(x^{k+1}))) \\ &\quad - (f(x^k) + p(x^k) + h(g(x^k))), \\ &\leq (f(x^{k+1}) + p(x^{k+1}) + h^{x^k}(g(x^{k+1}))) \\ &\quad - (f(x^k) + p(x^k) + h^{x^k}(g(x^k))), \\ &\leq \langle \zeta_p, x^{k+1} - x^k \rangle + \langle \zeta_h^k, x^{k+1} - x^k \rangle \\ &\quad + \langle x^{k+1} - x^k, \nabla f(x^k) \rangle + \frac{L_f}{2} \|x^{k+1} - x^k\|_2^2, \\ &= \langle \zeta_p + \zeta_h^k + \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L_f}{2} \|x^{k+1} - x^k\|_2^2, \\ &= -\frac{2\alpha - L_f}{2} \|x^{k+1} - x^k\|_2^2, \end{aligned}$$

where the third inequality is obtained from Equations (8)–(10) and the last equality is obtained from the property (11). Let $a = \frac{2\alpha - L_f}{2}$. Since $\alpha > \frac{L_f}{2}$, $a > 0$. Therefore, we can obtain the following result:

$$F(x^{k+1}) \leq F(x^k) - \frac{a}{2} \|x^k - x^{k+1}\|_2^2.$$

□

The relative error condition (Hypothesis 2) for the PL-ICMM is proved in Proposition 5.

Proposition 5 (relative error condition). For all $k \in \mathbb{N}$, there exists a positive constant $C > 0$ (independent of k) and $\zeta^{k+1} \in \partial F(x^{k+1})$ such that

$$\|\zeta^{k+1}\|_2 \leq C \|x^{k+1} - x^k\|_2$$

Proof. By the optimality of the subproblem of PL-ICMM and Lemma 1, there exist $\xi_p \in \partial p(x^{k+1})$ and $\xi_h^k \in \partial(h^{x^k} \circ g(x^{k+1}))$, satisfying

$$\nabla f(x^k) + \xi_p + \xi_h^k + \alpha(x^{k+1} - x^k) = \mathbf{0}.$$

Let $y^k = \nabla h^{x^k}(g(x^{k+1}))$. Then, by Lemma 1 and the property of the subdifferential,

$$\partial(h^{x^k} \circ g)(x^{k+1}) = \partial\langle y^k, g \rangle(x^{k+1}) = \sum_i \partial(y_i^k g_i)(x^{k+1}) = \sum_i y_i^k \partial g_i(x^{k+1}).$$

We can decompose

$$\xi_h^k = \partial(h^{x^k} \circ g)(x^{k+1}) = \sum_i y_i^k \eta_i$$

for some $\eta_i \in \partial g_i(x^{k+1})$. Similarly, for any subderivative $\xi_h \in \partial(h \circ g)(x^{k+1})$, ξ_h can also be decomposed as

$$\xi_h = \partial(h \circ g)(x^{k+1}) = \partial\langle y, g \rangle(x^{k+1}) = \sum_i \partial(y_i g_i)(x^{k+1}) = \sum_i y_i \partial g_i(x^{k+1}),$$

where $y = \nabla h(g(x^{k+1}))$. Hence, it can be obtained from Lemma 1 that $\xi^{k+1} := \nabla f(x^{k+1}) + \xi_p + \xi_h$ is a subderivative of F at x^{k+1} , i.e., $\xi^{k+1} \in \partial F(x^{k+1})$, where $\xi_h := \sum_i y_i \eta_i \in \partial(h \circ g)(x^{k+1})$.

From Proposition 4 and the coercivity of F , the sequence $\{x^k\}$ is bounded. Hence, we can find a compact set containing this sequence in \mathbb{R}^n . The convexity of g implies the Lipschitz continuity on a compact, convex subset of \mathbb{R}^n involving $\{x^k\}$ for all k . From the further assumption, ∇h and ∇h^x are Lipschitz continuous on a compact, convex subset B of \mathbb{R}^m containing $g(x^k)$ for all k . Let L_1 and L_2 be the Lipschitz constants of g and ∇h , respectively. The common Lipschitz constant of ∇h^x on B is set to be L_h . By the property of the local Lipschitz continuity of g , we can obtain

$$\|\xi_h - \xi_h^k\|_2 = \|\sum_i (y_i - y_i^k) \eta_i\|_2 = \|\sum_i (y - y^k)_i \eta_i\|_2 \leq L_1 \|y - y^k\|_2. \tag{12}$$

Since $\nabla f(x^k) + \xi_p + \xi_h^k + \alpha(x^{k+1} - x^k) = \mathbf{0}$, the following identities hold:

$$\begin{aligned} \|\xi^{k+1}\|_2 &= \|\nabla f(x^{k+1}) + \xi_p + \xi_h\|_2, \\ &= \|\nabla f(x^{k+1}) + \xi_p + \xi_h - (\nabla f(x^k) + \xi_p + \xi_h^k + \alpha(x^{k+1} - x^k))\|_2. \end{aligned} \tag{13}$$

From the fact $\nabla h(g(x^k)) = \nabla h^{x^k}(g(x^k))$ and the Lipschitz continuity of ∇h and ∇h^{x^k} , we have

$$\begin{aligned} \|\xi^{k+1}\|_2 &\leq \|\xi_h - \xi_h^k\|_2 + \|\nabla f(x^{k+1}) - \nabla f(x^k)\|_2 + \alpha \|x^{k+1} - x^k\|_2, \\ &\leq L_1 \|y - y^k\|_2 + \|\nabla f(x^{k+1}) - \nabla f(x^k)\|_2 + \alpha \|x^{k+1} - x^k\|_2, \\ &= L_1 \|y - \nabla h(g(x^k)) + \nabla h^{x^k}(g(x^k)) - y^k\|_2 + \|\nabla f(x^{k+1}) - \nabla f(x^k)\|_2 \\ &\quad + \alpha \|x^{k+1} - x^k\|_2, \\ &\leq (L_h + L_2)L_1 \|g(x^{k+1}) - g(x^k)\|_2 + L_f \|x^{k+1} - x^k\|_2 + \alpha \|x^{k+1} - x^k\|_2, \\ &\leq ((L_h + L_2)L_1^2 + L_f + \alpha) \|x^{k+1} - x^k\|_2, \end{aligned}$$

where the first inequality is obtained from the Equation (13), and the second inequality is obtained from the Equation (12). Letting $C := (L_h + L_2)L_1^2 + L_f + \alpha$, the final result can be obtained. \square

Proposition 6 (continuity condition). *There exists a convergent subsequence $\{x^{k_j}\}$ of $\{x^k\}$ and its limit \bar{x} satisfying*

$$\lim_{j \rightarrow \infty} x^{k_j} = \bar{x}, F(x^{k_j}) \rightarrow F(\bar{x}) \text{ as } j \rightarrow \infty.$$

Proof. The boundedness of $\{x^k\}$ implies the existence of a convergent subsequence. Let $\{x^{k_j}\}$ be a convergent subsequence of $\{x^k\}$ such that $x^{k_j} \rightarrow \bar{x}$ as $j \rightarrow \infty$. We define sequences $\zeta_1^{k_j} \in \partial p(x^{k_j})$ and $\zeta_2^{k_j} \in \partial(h \circ g)(x^{k_j})$ such that

$$0 = \nabla f(x^{k_j-1}) + \zeta_1^{k_j} + \zeta_2^{k_j} + \alpha(x^{k_j} - x^{k_j-1}).$$

Let $q^{k_j}(x) = \frac{\alpha}{2} \|x - x^{k_j-1}\|_2^2$. Clearly, q^{k_j} is a convex function. Due to the strict continuity of $h^{x^{k_j-1}} \circ g + q^{k_j}$, $\partial(h^{x^{k_j-1}} \circ g + q^{k_j})(x)$ is bounded. Clearly, $\{\nabla f(x^{k_j})\}$ is bounded from the continuity of ∇f . Therefore, $\{\zeta_1^{k_j}\}$ is bounded and $\lim_{j \rightarrow \infty} \langle \zeta_1^{k_j}, \bar{x} - x^{k_j} \rangle = 0$. Using the facts of the lower semicontinuity of F , the continuity of f and the convexity of p , we obtain

$$\begin{aligned} F(\bar{x}) &\leq \liminf_{j \rightarrow \infty} F(x^{k_j}), \\ &\leq \limsup_{j \rightarrow \infty} F(x^{k_j}), \\ &\leq \limsup_{j \rightarrow \infty} f(x^{k_j}) + \limsup_{j \rightarrow \infty} p(x^{k_j}) + \limsup_{j \rightarrow \infty} (h \circ g)(x^{k_j}), \\ &= f(\bar{x}) + \limsup_{j \rightarrow \infty} p(x^{k_j}) + \lim_{j \rightarrow \infty} \langle \zeta_1^{k_j}, \bar{x} - x^{k_j} \rangle + h \circ g(\bar{x}), \\ &= f(\bar{x}) + \limsup_{j \rightarrow \infty} (p(x^{k_j}) + \langle \zeta_1^{k_j}, \bar{x} - x^{k_j} \rangle) + h \circ g(\bar{x}), \\ &\leq f(\bar{x}) + p(\bar{x}) + h \circ g(\bar{x}) = F(\bar{x}). \end{aligned}$$

Thus, $F(x^{k_j}) \rightarrow F(\bar{x})$ as $j \rightarrow \infty$. \square

In Propositions 4–6, we show that PL-ICMM satisfies the three alternative Hypotheses 1–3. Finally, we can obtain the global convergence of the proposed algorithm.

Theorem 2. *Let $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be a proper l.s.c. function. Let the sequence $\{x^k\}$ be generated by PL-ICMM. If F has the KL property at the cluster point $x^* := \lim_{j \rightarrow \infty} x^{k_j}$, then the sequence $\{x^k\}_{k \in \mathbb{N}}$ converges to $x^* \in X$ as $k \rightarrow \infty$ and x^* is a critical point of F . Moreover, the sequence $\{x^k\}_{k \in \mathbb{N}}$ has a finite length, i.e.,*

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|_2 < \infty.$$

Proof. Propositions 4–6 yield all of the requirements for Theorem 1. According to Theorem 1, we can obtain the final results. \square

4. Numerical Experiments and Discussion

In this section, we present the numerical results of the proposed methods, and applications of the proposed algorithms are provided. We consider compressive sensing in signal processing. All numerical experiments are implemented with MATLAB R2020b on a 64-bit Windows 10 desktop with an Intel Xeon(R) 2.40 GHz CPU, with 64 GB RAM.

4.1. Numerical Results for PL-IRL1

First, we show the performance of PL-IRL1. The main concept of compressive sensing is that a sparse signal can be recovered from incomplete information, i.e., underdetermined system $Ax = b$ where $m \ll n$. We say that x is k -sparse if x has only k -nonzero elements. The compressive sensing problem is generally an ill-posed problem, and there exist many solutions mathematically. To obtain the sparse solution, the basic model for compressive sensing is called lasso, which has the following form:

$$\min_{x \in \mathbb{R}^n} \|x\|_1 + \frac{\mu}{2} \|Ax - b\|_2^2,$$

where $A \in \mathbb{R}^{m \times n}$ with $m \ll n$, $b \in \mathbb{R}^m$, $\mu > 0$ is a positive regularization parameter. This problem is a convex relaxation of the following nonconvex ℓ_0 minimization problem:

$$\min_{x \in \mathbb{R}^n} \|x\|_0 + \frac{\mu}{2} \|Ax - b\|_2^2,$$

where $\|x\|_0$ is defined as the number of nonzero elements of input x . Recently, sparse signal recovery from an observed signal corrupted impulsive noise was interested in lots of works [32–35]. For the sparse recovery with impulsive noise, the following ℓ_1 -fidelity based convex problem can be often applied:

$$\min_{x \in \mathbb{R}^n} \|x\|_1 + \frac{\mu}{2} \|Ax - b\|_1.$$

We also consider nonconvex variations of this model for a compressive sensing problem with impulsive noise as follows:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n \log(1 + \rho|x_i|) + \frac{\mu}{2} \|Ax - b\|_1,$$

and

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n \frac{|x_i|}{1 + \rho|x_i|} + \frac{\mu}{2} \|Ax - b\|_1,$$

where $\rho > 0$ is a positive parameter. Unfortunately the PL-IRL1 can be directly applied to these nonconvex problems. Hence, we add the auxiliary variable $z \in \mathbb{R}^m$ and adopt the penalty technique, leading us to obtain the following nonconvex and nonsmooth minimization problems:

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} F_1(x, z) = \sum_{i=1}^n \log(1 + \rho|x_i|) + \frac{\beta}{2} \|z - Ax + b\|_2^2 + \mu \|z\|_1, \tag{14}$$

and

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} F_2(x, z) = \sum_{i=1}^n \frac{|x_i|}{1 + \rho|x_i|} + \frac{\beta}{2} \|z - Ax + b\|_2^2 + \mu \|z\|_1, \tag{15}$$

where β is a positive penalty constant. We set

$$g(x) = |x|, \quad h(y) = \mu \sum_{i=1}^n \log(1 + \rho y_i) \text{ or } \sum_{i=1}^n \frac{|x_i|}{1 + \rho|x_i|},$$

$$f(x) = \frac{\beta}{2} \|z - Ax + b\|_2^2, \quad p(x) = \frac{\mu}{2} \|z\|_1.$$

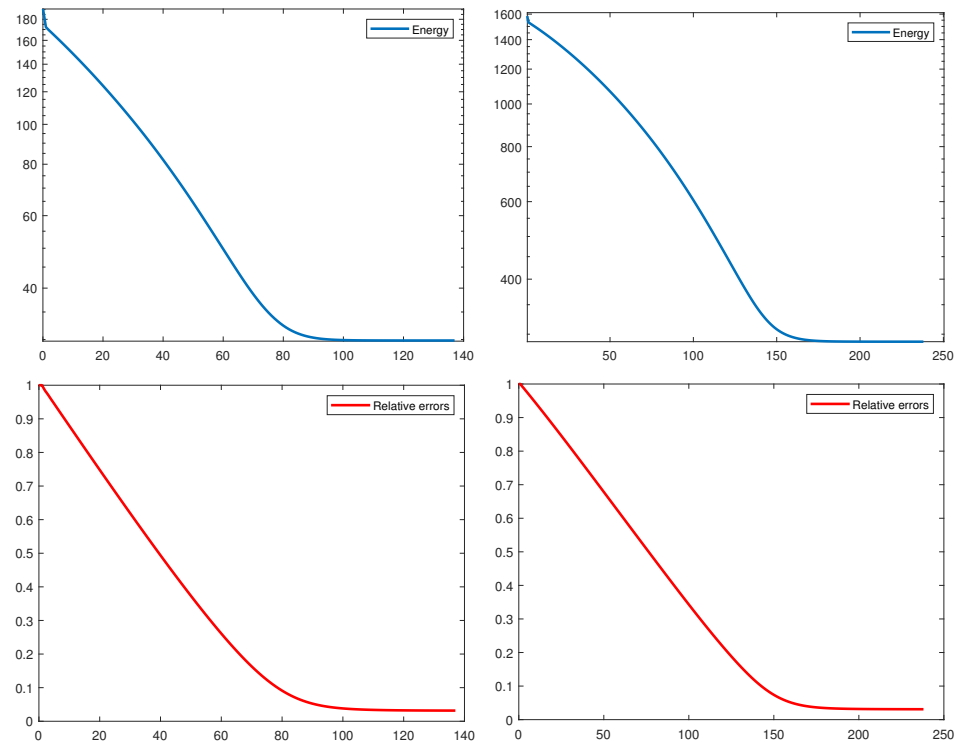


Figure 1. Comparison results of the energy and relative error for PL-IRL1 algorithm applied to problems (14) and (15). **Left:** Result for problem (14). **Right:** Result for problem (15).

Table 1. Numerical results for PL-IRL1 algorithm applied to models (14) and (15).

Model		Computing Time (s)	Iterations	Energy	Relative Error
Problem (14)	mean	5.437452	142	34.1832	3.5502×10^{-2}
	std	0.293790	8	2.1721	4.5748×10^{-3}
Problem (15)	mean	8.960141	236	282.1852	3.1598×10^{-2}
	std	0.563042	15	13.0279	4.2781×10^{-3}

In this setting, the minimization problems (14) and (15) have the form of problem (1). Since the objective function of the problem (14) is definable in the log-exp σ -minimal structure, it is a KL function. The objective function of the minimization problem (15) is a semialgebraic function and it is also a KL function. Moreover, these are closed and coercive. The function f is a convex, proper and continuously differentiable function, p is a convex, proper and l.s.c function, and g is a proper and l.s.c function. The function $h : \mathbb{R}_+^n \rightarrow \mathbb{R}$ is coordinatewise nondecreasing, continuously differentiable and concave. Since the objective function involves nondifferentiable term $\mu \|z\|_1$, the proximal iterative reweighted algorithm proposed in [31] cannot be applied. Hence, we can apply IRL1 [10] or PL-IRL1 to solve the given problem (14). The IRL1 or PL-IRL1 applied to the problem (14) is given as follows:

$$\begin{cases} w^k &= \begin{cases} \frac{\rho}{1 + \rho|x^k|} & \text{for (14)} \\ \frac{1}{(1 + \rho|x^k|)^2} & \text{for (15)} \end{cases} \\ (x^{k+1}, z^{k+1}) &= \arg \min_{x,z} \langle w^k, |x| \rangle + \frac{\beta}{2} \|z - Ax + b\|_2^2 + \mu \|z\|_1, \end{cases} \quad (16)$$

and

$$\begin{cases} w^k & = \begin{cases} \frac{\rho}{1 + \rho|x^k|} & \text{for (14)} \\ \frac{\rho}{(1 + \rho|x^k|)^2} & \text{for (15)} \end{cases} \\ (x^{k+1}, z^{k+1}) & = \arg \min_{x,z} \langle w^k, |x| \rangle + \beta \langle z^k - Ax^k + b, z - Ax + b \rangle + \mu \|z\|_1 \\ & \quad + \frac{\delta}{2} \|x - x^k\|_2^2 + \frac{\delta}{2} \|z - z^k\|_2^2, \end{cases} \tag{17}$$

respectively. For solving the convex subproblem of IRL1, the optimality conditions are given as follows:

$$\begin{cases} 0 \in \partial(\langle w^k, |x| \rangle)(x^{k+1}) + \beta A^T(Ax^{k+1} - b - z^{k+1}), \\ 0 \in \beta(z^{k+1} - Ax^{k+1} + b) + \mu \partial(\|z\|_1)(z^{k+1}). \end{cases} \tag{18}$$

This system (18) of equations does not have a closed form solution. Hence, the IRL1 cannot be employed to solve the problem (14). On the other hand, the optimality conditions of the convex subproblem of our method are given as

$$\begin{cases} 0 \in \partial(\langle w^k, |x| \rangle)(x^{k+1}) + \beta A^T(Ax^k - b - z^k) + \delta(x^{k+1} - x^k), \\ 0 \in \beta(z^k - Ax^k + b) + \mu(\|z\|_1)(z^{k+1}) + \delta(z^{k+1} - z^k). \end{cases} \tag{19}$$

These equations in (19) are separable, and each problem has a closed form solution:

$$\begin{aligned} x^{k+1} &= \text{shrink}\left(x^k - \frac{\beta}{\delta} A^T(Ax^k - b - z^k), \frac{w^k}{\delta}\right), \\ z^{k+1} &= \text{shrink}\left(z^k - \frac{\beta}{\delta}(z^k - Ax^k + b), \frac{\mu}{\delta}\right), \end{aligned}$$

where shrink function is defined as

$$\text{shrink}(a, b) = \text{sign}(a) \cdot \max(|a| - b, 0).$$

To show the convergence of the PL-IRL1 to solve problems (14) and (15), we perform the numerical experiments in the following setting. The size of A is $n = 5000$ and $m = 2500$. We use an orthonormal Gaussian measurement matrix A whose entries are randomly chosen by standard Gaussian distribution, and then each column of A is divided by its ℓ_2 norm. The number l of nonzero elements of the original sparse signal x_0 is fixed at 50, the locations of nonzero elements are selected randomly, and the values of nonzero elements are chosen by Gaussian distribution $\mathcal{N}(0, 10^2)$. The observed data b can be calculated by

$$b = Ax_0 + n,$$

where n is the Gaussian mixture noise that consists of two Gaussian components, which is one of impulsive noise in signal processing [36]. The measured equation of n is given by

$$n = vn_1 + (1 - v)n_2,$$

where n_1 and n_2 are Gaussian noise with mean 0 and standard deviation η and $q\eta$. The n_1 denotes the background noise, and n_1 represents the influence of outliers. The parameter $v \in (0, 1)$ controls the proportion of the large outliers and $q > 1$ controls the strength of the outliers. Here, we fixed these parameters at $(v, q, \eta) = (0.1, \sqrt{10}, 0.02)$.

Since $\|A\|_2 = 1$, the Lipschitz constant of ∇f is 2β . So, δ is set to be $\beta + 0.001$. The regularization parameter μ is fixed at 0.2 for (14) or 1.5 for (15); the penalty parameter β is set to be 2 for (14) or 28 for (15); and the controlling parameter ρ of nonconvexity is set

to be 0.1. For the stopping condition of our algorithm, we use relative errors over energy function values, whose specific formulation is given as

$$\frac{F_i(x^k) - F_i(x^{k+1})}{F_i(x^0)} < 10^{-7}, \quad i = 1, 2.$$

For this setting, 100 different numerical tests are conducted.

In Table 1, we present the computing time, number of iterations, final energy value, and relative error. From the relative errors, we can observe that the proposed algorithm finds an approximated sparse solution with small error for all cases. In Figure 1, we illustrate the relative errors and energy values $F_i(x^k)$ over k iterations. Since the penalty parameter β for (14) is smaller than that for (15), the computing time of our method for solving (14) is faster than that for solving (15). We can also see similar results in terms of the number of iterations. Ultimately, we can show in Figure 1 and Table 1 the fast convergence of the PL-IRL1, and the final energy value has enough small value.

4.2. Numerical Results for PL-IRLS

Second, we present numerical results of PL-IRLS compared with the iPiano [17] method for the compressive sensing problem in signal processing. Specifically, the restoration of the sparse signal corrupted by additive Gaussian noise is considered. We apply the algorithms PL-IRLS and iPiano to the following unconstrained problems:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2\rho} \sum_{i=1}^n \log(1 + \rho|x_i|^2) + \frac{\beta}{2} \|x\|_2^2 + \frac{\mu}{2} \|Ax - b\|_2^2, \tag{20}$$

and

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^n \frac{|x_i|^2}{1 + \rho(|x_i| + c)^2} + \frac{\beta}{2} \|x\|_2^2 + \frac{\mu}{2} \|Ax - b\|_2^2, \tag{21}$$

where β, μ are positive parameters, c is a positive constant, and $\rho > 0$ is the parameter that controls the nonconvexity of regularizing term. These problems are a nonconvex variations of lasso, which is a well-known model for compressive sensing. The objective functions of problems (20) and (21) are definable in the log-exp σ -minimal structure, and they are also close, coercive KL functions. With setting

$$g(x) = |x|, \quad h(y) = \frac{1}{2\rho} \sum_{i=1}^n \log(1 + \rho y_i^2) \text{ or } \frac{1}{2} \sum_{i=1}^n \frac{y_i^2}{1 + \rho(y_i + c)^2}$$

$$f(x) = \frac{\mu}{2} \|Ax - b\|_2^2, \quad p(x) = \frac{\beta}{2} \|x\|_2^2,$$

all assumptions in Algorithm 4 are satisfied. Since the norm of the Hessian matrix of h is bounded on \mathbb{R}_+^n , it has also a strictly continuous gradient. The PL-IRLS applied to the problems (20) and (21) can be obtained by

$$\begin{cases} w^k &= \begin{cases} \frac{1}{1 + \rho|x^k|^2}, & \text{for (20)} \\ \frac{c\rho(|x^k| + c) + 1}{(1 + \rho(|x^k| + c))^2} & \text{for (21)} \end{cases}, \\ x^{k+1} &= \arg \min_x \frac{1}{2} \langle w^k, x^2 \rangle + \frac{\delta}{2} \|x - x^k\|_2^2 + \mu \langle A^T(Ax^k - b), x - x^k \rangle. \end{cases} \tag{22}$$

The subproblem in (22) is a quadratic problem, and its normal equation is given as follows:

$$w^k x + \delta(x - x^k) + \mu A^T(Ax^k - b) = 0.$$

It can be rewritten as the following linear equation,

$$(diag(w^k) + \delta I)x = \delta x^k - \mu A^T(Ax^k - b),$$

where $diag(w^k)$ is a diagonal matrix whose diagonal entries consist of w^k . Since $diag(w^k) + \delta I$ is a diagonal matrix, this linear equation can be solved exactly and easily. The majorization property of PL-IRLS is obtained from ([10], Proposition 23) for any $\rho > 0$ and $c > 0$.

In this experiment, we use partial discrete cosine transform (DCT) matrices A whose i -th rows are selected from the $n \times n$ DCT matrix. We note that the partial DCT matrices are implicitly stored, i.e., matrix-vector multiplications in Ax or $A^T x$ are computed by the DCT or inverse of DCT. Hence, we can use partial DCT matrices of very large sizes. Here, the size n, m is fixed at $(n, m) = (100,000, 30,000)$. Then, the original IRLS can be actually applied to the nonsmooth and nonconvex problem (20), which is given by

$$\begin{cases} w^k &= \begin{cases} \frac{1}{1 + \rho|x^k|^2}, & \text{for (20)} \\ \frac{c\rho(|x^k| + c) + 1}{(1 + \rho(|x^k| + c))^2} & \text{for (21)} \end{cases} \\ x^{k+1} &= \arg \min_x \frac{1}{2} \langle w^k, x^2 \rangle + \frac{\mu}{2} \|Ax - b\|_2^2. \end{cases} \tag{23}$$

The optimality condition of the subproblem in the method (23) is

$$w^k x^{k+1} + \mu A^T(Ax^{k+1} - b) = 0,$$

and it is equivalent to

$$(diag(w^k) + \mu A^T A)x^{k+1} = \mu A^T b.$$

Since the size of our measurement matrix is very large, finding the exact solution of this linear equation is time consuming and it seem to be impossible in many cases.

The number l of nonzero elements of the sparse signal x_0 is fixed at 2000, and the locations of nonzero elements are randomly chosen. The values of nonzero elements are selected from standard Gaussian distribution. The observed data b is measured by the formula

$$b = Ax_0 + n,$$

where n is the Gaussian white noise with mean 0 and standard deviation 0.02. The regularization parameters (β, μ) are fixed at (0.001, 1.5) for (20) or (0.001, 0.9) for (21) in all tests. The values of (ρ, c) are also fixed at (250, 0.001). The proximity parameter δ in our method is set to be $\frac{\mu}{2} + 0.0001$ because the ℓ_2 norm of a partial DCT matrix is less than or equal to 1.

We present the mean values and standard deviations over 100 trials of computing time, number of iterations, energy value and relative error in Table 2. In Figure 2, we plot the relative errors and energy values of PL-IRLS and iPiano over iterations. Figure 2 shows the convergence of PL-IRLS and iPiano. For the tests, average values of the energy value and relative error for PL-IRLS are almost same with those for the iPiano. This shows that PL-IRLS and iPiano recover almost the same sparse solutions. Hence, it demonstrates the similar performance between PL-IRLS and iPiano in terms of the accuracy and optimization of energy functional. On the other hand, it can be observed in Table 2 and Figure 2 that PL-IRLS is faster than iPiano. Therefore, PL-IRLS is superior to iPiano for solving the nonconvex and nonsmooth problems (20) and (21). In conclusion, PL-IRLS gives better performance over iPiano.

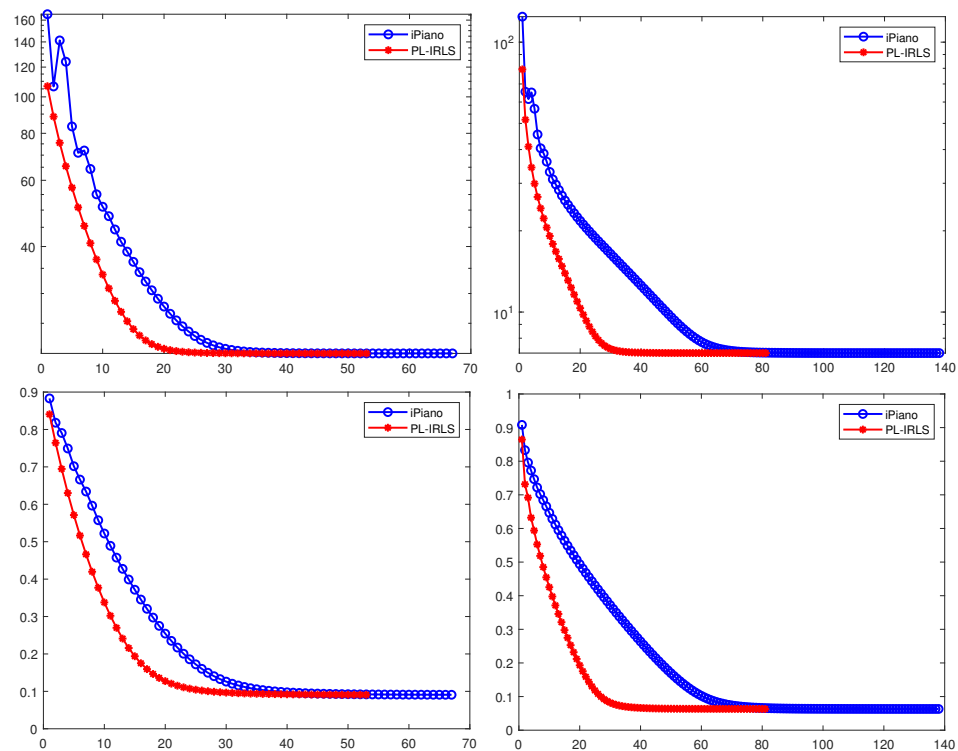


Figure 2. Comparison results of the energy and relative errors for PL-IRLS and iPiano applied to problem (20) and (21). **Left:** Result for problem (20). **Right:** Result for problem (21).

Table 2. Numerical results for PL-IRL1 and iPiano applied to the model (20) and (21).

Method		Computing Time (s)	Iterations	Energy	Relative Error
Model (20)					
PL-IRLS	mean	3.255264	53	21.2322	8.7822×10^{-2}
	std	0.215043	1	0.2123	2.0949×10^{-3}
iPiano	mean	4.165273	68	21.2323	8.7787×10^{-2}
	std	0.226674	1	0.2123	2.0945×10^{-3}
Model (21)					
PL-IRLS	mean	5.588428	90	7.1838	6.1408×10^{-2}
	std	0.775879	11	0.0483	1.4041×10^{-3}
iPiano	mean	8.849843	144	7.1840	6.1389×10^{-2}
	std	0.915589	12	0.0483	1.4009×10^{-3}

5. Conclusions

In this paper, we proposed proximal linearized reweighted algorithms to solve nonconvex and nonsmooth unconstrained minimization problem (1). Based on the general unified framework, we suggested an extension of the iterative convex majorization–minimization method for solving (1). Moreover, extended versions of the iteratively reweighted ℓ_1 algorithm and iterative least square algorithm were also introduced. The global convergence of the proposed algorithm was also proved under uncertain assumptions. Lastly, the numerical results related to compressive sensing demonstrated that the proposed methods provides the outstanding performance compared with state-of-the-art methods. Recently, several algorithms were extended by imposing an additional inertial term, resulting in a faster convergence rate. In future, we will study a proximal linearized reweighted algorithm with an inertial force.

Author Contributions: Conceptualization, M.K.; Formal analysis, J.Y. and M.K.; Funding acquisition, M.K.; Methodology, J.Y. and M.K.; Software, M.K.; Validation, J.Y. and M.K.; Writing original draft, J.Y. and M.K.; Writing review and editing, J.Y. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by research fund of Chungnam National University.

Acknowledgments: The authors would like to thank the referees for their helpful comments. The authors would like to thank Chungnam National University for funding support.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

l.s.c.	Lower semicontinuous
ICMM	Iterative convex majorization–minimization method
KL	Kurdyka–Łojasiewicz
PL-ICMM	Proximal linearized iterative convex majorization–minimization method
PL-IRL1	Proximal linearized iteratively reweighted ℓ_1 algorithm
PL-IRLS	Proximal linearized iteratively reweighted least square algorithm
IRL1	Iteratively reweighted ℓ_1 algorithm
DCT	discrete cosine transform

References

- Curry, H.B. The method of steepest descent for non-linear minimization problems. *Q. Appl. Math.* **1944**, *2*, 258–261. [[CrossRef](#)]
- Nesterov, Y.E. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk Sssr* **1983**, *269*, 543–547.
- Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [[CrossRef](#)]
- Rockafellar, R.T. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **1976**, *14*, 877–898. [[CrossRef](#)]
- Kaplan, A.; Tichatschke, R. Proximal point methods and nonconvex optimization. *J. Glob. Optim.* **1998**, *13*, 389–406. [[CrossRef](#)]
- Gong, P.; Zhang, C.; Lu, Z.; Huang, J.; Ye, J. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In Proceedings of the 30th International Conference on Machine Learning, PMLR **2013**, *28*, 37–45.
- Attouch, H.; Bolte, J.; Redont, P.; Soubeyran, A. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality. *Math. Oper. Res.* **2010**, *35*, 438–457. [[CrossRef](#)]
- Candes, E.J.; Wakin, M.B.; Boyd, S.P. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.* **2008**, *14*, 877–905. [[CrossRef](#)]
- Lai, M.J.; Xu, Y.; Yin, W. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization. *SIAM J. Numer. Anal.* **2013**, *51*, 927–957. [[CrossRef](#)]
- Ochs, P.; Dosovitskiy, A.; Brox, T.; Pock, T. On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision. *SIAM J. Imaging Sci.* **2015**, *8*, 331–372. [[CrossRef](#)]
- Łojasiewicz, S. Sur la géométrie semi-et sous-analytique. *Ann. l’Institut Fourier* **1993**, *43*, 1575–1595. [[CrossRef](#)]
- Łojasiewicz, S. Une propriété topologique des sous-ensembles analytiques réels. *Les Équations aux Dérivées Partielles* **1963**, *117*, 87–89.
- Bolte, J.; Daniilidis, A.; Lewis, A.; Shiota, M. Clarke subgradients of stratifiable functions. *SIAM J. Optim.* **2007**, *18*, 556–572. [[CrossRef](#)]
- Bolte, J.; Daniilidis, A.; Lewis, A. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* **2007**, *17*, 1205–1223. [[CrossRef](#)]
- Bolte, J.; Combettes, P.L.; Pesquet, J.C. Alternating proximal algorithm for blind image recovery. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 1673–1676. [[CrossRef](#)]
- Attouch, H.; Bolte, J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program.* **2009**, *116*, 5–16. [[CrossRef](#)]
- Ochs, P.; Chen, Y.; Brox, T.; Pock, T. iPiano: Inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sci.* **2014**, *7*, 1388–1419. [[CrossRef](#)]
- Attouch, H.; Bolte, J.; Svaiter, B.F. Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Math. Program.* **2013**, *137*, 91–129. [[CrossRef](#)]
- Osher, S.; Mao, Y.; Dong, B.; Yin, W. Fast linearized Bregman iteration for compressive sensing and sparse denoising. *Commun. Math. Sci.* **2010**, *8*, 93–111.

20. Bolte, J.; Sabach, S.; Teboulle, M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **2014**, *146*, 459–494. [[CrossRef](#)]
21. Rockafellar, R.T.; Wets, R.J.B. *Variational Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; Volume 317.
22. Mordukhovich, B.S.; Nam, N.M.; Yen, N. Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming. *Optimization* **2006**, *55*, 685–708. [[CrossRef](#)]
23. Bochnak, J.; Coste, M.; Roy, M.F. *Real Algebraic Geometry*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 36.
24. Wilkie, A.J. Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function. *J. Am. Math. Soc.* **1996**, *9*, 1051–1094. [[CrossRef](#)]
25. Dries, L.P.D.v.d. *Tame Topology and O-Minimal Structures*; Cambridge University Press: Cambridge, UK, 1998; Volume 248.
26. Kang, M. Approximate versions of proximal iteratively reweighted algorithms including an extended IP-ICMM for signal and image processing problems. *J. Comput. Appl. Math.* **2020**, *376*, 112837. [[CrossRef](#)]
27. Combettes, P.L.; Wajs, V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168–1200. [[CrossRef](#)]
28. Chartrand, R.; Yin, W. Iteratively reweighted algorithms for compressive sensing. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 3869–3872.
29. Daubechies, I.; DeVore, R.; Fornasier, M.; Güntürk, C.S. Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math.* **2010**, *63*, 1–38. [[CrossRef](#)]
30. Needell, D. Noisy signal recovery via iterative reweighted L1-minimization. In Proceedings of the 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1–4 November 2009; pp. 113–117.
31. Sun, T.; Jiang, H.; Cheng, L. Global convergence of proximal iteratively reweighted algorithm. *J. Glob. Optim.* **2017**, *68*, 815–826. [[CrossRef](#)]
32. Ji, Y.; Yang, Z.; Li, W. Bayesian sparse reconstruction method of compressed sensing in the presence of impulsive noise. *Circuits Syst. Signal Process.* **2013**, *32*, 2971–2998. [[CrossRef](#)]
33. Javaheri, A.; Zayyani, H.; Figueiredo, M.A.; Marvasti, F. Robust sparse recovery in impulsive noise via continuous mixed norm. *IEEE Signal Process. Lett.* **2018**, *25*, 1146–1150. [[CrossRef](#)]
34. Wen, F.; Liu, P.; Liu, Y.; Qiu, R.C.; Yu, W. Robust Sparse Recovery in Impulsive Noise via ℓ_p - ℓ_1 Optimization. *IEEE Trans. Signal Process.* **2016**, *65*, 105–118. [[CrossRef](#)]
35. Wen, J.; Weng, J.; Tong, C.; Ren, C.; Zhou, Z. Sparse signal recovery with minimization of 1-norm minus 2-norm. *IEEE Trans. Veh. Technol.* **2019**, *68*, 6847–6854. [[CrossRef](#)]
36. Wen, F.; Pei, L.; Yang, Y.; Yu, W.; Liu, P. Efficient and robust recovery of sparse signal and image using generalized nonconvex regularization. *IEEE Trans. Comput. Imaging* **2017**, *3*, 566–579. [[CrossRef](#)]