JOURNAL OF THE
CHUNGCHEONG MATHEMATICAL SOCIETY
Volume **37**, No. 1, February 2024
http://dx.doi.org/10.14403/jcms.2024.37.1.41

GENERALIZED PROXIMAL ITERATIVELY REWEIGHTED ℓ_1 ALGORITHM WITH CO-COERCIVENESS FOR NONSMOOTH AND NONCONVEX MINIMIZATION PROBLEM.

Myeongmin Kang*

Abstract. The nonconvex and nonsmooth optimization problem has been widely applicable in image processing and machine learning. In this paper, we propose an extension of the proximal iteratively reweighted ℓ_1 algorithm for nonconvex and nonsmooth minmization problem. We assume the co-coerciveness of a term of objective function instead of Lipschitz gradient condition, which is generalized property of Lipschitz continuity. We prove the global convergence of the proposed algorithm. Numerical results show that the proposed algorithm converges faster than original proximal iteratively reweighed algorithm and existing algorithms.

1. Introduction

We consider the following nonconvex and nonsmooth optimization problem:

$$\min_x f_1(x) + f_2(x) + h(g(x)), \tag{1.1}$$

where $f_1 : \mathbb{R}^n \to \mathbb{R}$ is a proper, differentiable, convex function and $f_2 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a proper, lower semicontinuous (l.s.c.) and convex function, $g : \mathbb{R}^n \to \mathbb{R}^m$ is a proper and l.s.c. function, and $h : Image(g) \to \mathbb{R}$ is continuously differentiable function. Furthermore, we assume the following conditions:

- ∇f_1 is co-coercive.

Received January 11, 2024; Accepted January 29, 2024.
2020 Mathematics Subject Classification: Primary 65K10, 90C26, 56B34; Secondary 68U10.
Key words and phrases: co-coercive, iteratively ℓ_1 algorithm, linearization, Kurdyka-Łojasiewicz inequality, nonconvex optimization.
*Supported (in part) by NRF, Project No. NRF-2019R1I1A3A01055168.

- g is coordinate-wise convex, i.e. coordinate functions g_i of g are convex.
- h has a strictly continuous gradient and is coordinate-wise nondecreasing, i.e. $h(x) \leq h(x+ae_i)$ for all $i = 1, \cdots, m$, where e_i is i-th standard basis vector, $a > 0$ is constant and $x, x+ae_i \in Image(g)$.
- The objective function of the problem (1.1) is coercive, closed, bounded below and definable in an o-minimal structure.

Since studies have shown the effectiveness of nonconvex regularization in image processing problems, the problem of the form (1.1) is often used as a variational model in image processing or signal processing.

Over the past few decades, many algorithms for convex minimization problem have been developed, and the theory of convex optimization has made many advances. On the other hand, algorithms for solving nonconvex optimization have been proposed recently. The classical algorithms for nonconvex optimization are extensions of methods for convex optimization, such as gradient-based methods [9], proximal point methods [6], iterative shrinkage thresholding algorithm [4], and alternating direction method of multipliers [14, 5] . In recent, an iteratively reweighted ℓ_1 algorithm (IRL1) for nonconvex regularization model [3] to solve a compressive sensing problem without convergence analysis. Here, the nonconvex regularization based model has the form of the problem (1.1). Ochs et al. [10] proposed generalized version of the IRL1, which is an iterative convex majorization-minimization method for solving nonsmooth and nonconvex optimization problem (1.1) and proved the convergence analysis. By adopting linearization technique, the proximal linearized iteratively reweighted algorithms were proposed in [13].

A proximal linearlized iteratively reweighted ℓ_1 algorithm (PL-IRL1) is an example of proximal linearlized convex majorization-minimization methods proposed in [13]. For the convergence of PL-IRL1, it was assumed that f_1 has Lipschitz gradient to deal with the linearization of f_1. In this paper, we propose a generalized version of PL-PIRL1 under the assumption the co-coerciveness of ∇f_1. This enables the proposed method to be applied to many applications and it also leads that the proposed method has faster convergence than PL-PIRL1.

The rest of this manuscript is organized as follows. Section 2 will give the definitions of the mathematical concepts in assumptions. In Section 3 we will define the proposed algorithm and prove the global convergence. In Section 4, we will apply the proposed algorithm to image processing and signal processing examples and compare the performance of the proposed algorithm to the existing algorithms. Finally, in Section 5, we will conclude our work.

2. Mathematical Preliminary

In this section, we introduce several mathematical concepts and properties for our method. First, we define Kurdyka-Łojasiewicz property [7, 8] that is often used to prove the convergence of algorithms for solving nonconvex minimization problems.

DEFINITION 2.1. A function $f : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ satisfies the Kurdyka-Łojasiewicz (KL) property at a point $x^* \in \mathrm{dom}(\partial f)$ if there exist $\rho \in (0, \infty]$, a neighborhood U of x^* and a continuous and convex function $\varphi : [0, \rho) \to [0, \infty)$ such that

- $\varphi(0) = 0$
- φ is differentiable in $(0, \rho)$
- $\varphi(t) > 0, \forall t \in (0, \rho)$
- $\forall x \in U \cap \{x \in \mathbb{R} : f(x^*) < f(x) < f(x^* + \rho)\}$,

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

If f satisfies the KL property at every point in $\mathrm{dom}(\partial f)$, f is called a KL function.

Simple KL functions are semialgebraic functions. It was proved in [2] that functions that are definable in some o-minimal structure satisfy the KL property. Note that definable sets and functions in some o-minimal structure share several properties of semialgebraic objects. The log-exp structure [12, 11] is a well-known example, which is a large o-minimal structure including all semialgebraic functions as well as log and exponential functions.

In order to solve minimization problems, the linearlization technique is frequently used to solve optimization problems and can make it easier to solve a subproblem of an iterative algorithm. For linearlization technique, it is usually assumed that the objective function or its part is a continuously differentiable and also has the Lipschitz continuity of its gradient. In this work, we introduce a concept more general than Lipschitz continuity, which is called co-coercive.

DEFINITION 2.2. A operator $B : \mathbb{R}^n \to \mathbb{R}^n$ is called co-coercive with a symmetric and positive definite map L if it satisfies the following inequality:

$$\langle B(x) - B(y), x - y\rangle \geq \|B(x) - B(y)\|^2_{L^{-1}}, \quad \forall x, y \in \mathbb{R}^n,$$

where $\|x\|_{L^{-1}}^2 = \langle L^{-1}x, x\rangle$.

Note that if the operator B is $\frac{1}{l}I$ co-coercive, then it is Lipschitz continuous with Lipschitz constant l.

Lastly, we introduce a general framework [1] for an iterative method for solving an unconstrained minimization problem. We consider the following general unconstrained minimization problem:

$$\min_{x\in\mathbb{R}^n} F(x), \tag{2.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ is a l.s.c., proper and KL function. It was proved in [1] that an iterative algorithm applied to solve the problem (2.1) converges when it satisfies the following three conditions. Let $\{x^k\}$ be a sequence generated by given algorithm.

C1 (Sufficient decrease condition) There exists a value $\alpha > 0$ such that

$$F(x^{k+1}) + \alpha\|x^{k+1} - x^k\|_2^2 \le F(x^k), \quad \forall k \in \mathbb{N}.$$

C2 (Relative error condition) For each $k \in \mathbb{N}$, there exist a sequence $\{u^k\} \in \partial F(x^k)$ and a constant $\beta > 0$ such that

$$\|u^{k+1}\|_2 \le \beta\|x^{k+1} - x^k\|_2.$$

C3 (Continuity condition) There exists a subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ of $\{x^k\}$ and $\bar{x}$ such that

$$x^{k_j} \to \bar{x} \quad \text{and} \quad F(x^{k_j}) \to F(\bar{x}), \quad \text{as } j \to \infty.$$

Under the above assumptions, the following convergence result is given in [1].

THEOREM 2.3. *Let $\{x^k\}$ be a sequence generated by given algorithm that satisfies the conditions C1-C3. If F has KL property at a cluster point $\bar{x}$ in C3, then the sequence $\{x^k\}$ converges to $\bar{x}$ as k goes to ∞ and $\bar{x}$ is a critical point of F. Furthermore, $\{x^k\}$ has a finite length, i.e.*

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

3. Proposed method

In this section, we propose an extension of the (proximal) iteratively reweighted ℓ_1 algorithm for solving the nonconvex and nonsmooth minimization problem (1.1). Let $F(x) := f_1(x) + f_2(x) + h(g(x))$ be the objective function of the problem (1.1). The iteratively reweighted ℓ_1 algorithm is a well-known algorithm, which uses a weight function as a convex realxation of the noncovex term in (1.1):

$$x^{k+1} = \arg\min_x f_1(x) + f_2(x) + \langle w^k, g(x)\rangle,$$

where the weight $w^k := \nabla h(g(x^k))$. This subproblem is convex, but it is not solved exactly in general. By adopting the linearization technique of f_1, the proximal linearized iteratively reweighted ℓ_1 algorithm (PL-IRL1) was proposed in [13]:

$$x^{k+1} = \arg\min_x \langle \nabla f_1(x), x - x^k\rangle + f_2(x) + \langle w^k, g(x)\rangle + \frac{\alpha}{2}\|x - x^k\|_2^2.$$

The proximal iteratively reweighted ℓ_1 algorithm converges globally under the assumptions that ∇f_1 is Lipschitz continuous with Lipschitz constant l and $\alpha > \frac{l}{2}$ holds. When l is large, it converges too slowly. Moreover, the assumption of Lipschitz continuity of ∇f_1 is too strong. To overcome this drawback, we assume more generalized concept which is co-coerciveness of ∇f_1. Under this assumption, the proposed method to solve the problem (1.1) is given in Algorithm 1.

Algorithm 1 Generalized proximal iteratively rewieghted ℓ_1 algorithm (GPL-IRL1)

1: **Input** : x^0 with $F(x^0) < \infty$, $\bar{L}$ is a symmetric positive definite matrix.
2: **repeat**
3: $\quad w^k = \nabla h(g(x^k))$
4: $\quad x^{k+1} = \arg\min_x \langle \nabla f_1(x), x - x^k\rangle + f_2(x) + \langle w^k, g(x)\rangle + \frac{1}{2}\|x - x^k\|_{\bar{L}}^2,$
5: **until** *a stopping criterion is satisfied.*

Let $p^k(x) := \langle \nabla h(g(x^k)), x\rangle$. Note that the function $p^k(x)$ is convex, proper, nondecreasing on $\mathrm{Im}(g)$. Moreover, $p^k(y) \geq h(y)$ for any $y \in \mathrm{Im}(g)$ and $p(g(x^k)) = h(g(x^k))$. A function that satisfies these conditions is called a convex majorizer of $h \circ g$. It follows form [10] that $p^k(g(x))$ is the optimal majorizer of $h \circ g$ at x^k.

First, we need a further assumption for the convergence of the proposed algorithm. It is assume that the functions p^k for all $k \in \mathbb{N}$ have strictly continuous gradients with common constant.

To prove the convergence of the proposed method, we need a property of co-coercive function.

Lemma 3.1. *If $\nabla f(x)$ is co-coercive, then*

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}\|x-y\|_L^2, \quad x, y \in \mathbb{R}^n. \tag{3.1}$$

Proof. Since ∇f is co-coercive, we have the following inequalities:

$$\begin{aligned}
\|\nabla f(x) - \nabla f(y)\|_{L^{-1}}^2 &\le \langle \nabla f(x) - \nabla f(y), x-y\rangle \\
&= \langle L^{-1/2}(\nabla f(x) - \nabla f(y)), L^{1/2}(x-y)\rangle \\
&\le \|L^{-1/2}(\nabla f(x) - \nabla f(y))\|_2 \|L^{1/2}(x-y)\|_2 \\
&= \sqrt{\langle L^{-1}(\nabla f(x) - \nabla f(y)), (\nabla f(x) - \nabla f(y))\rangle} \\
&\quad \cdot \sqrt{\langle L(x-y), (x-y)\rangle} \\
&= \|\nabla f(x) - \nabla f(y)\|_{L^{-1}} \|x-y\|_L,
\end{aligned}$$

where third inequality is obtained from the Cauchy-Schwarz inequality. Hence, we can obtain

$$\|\nabla f(x) - \nabla f(y)\|_{L^{-1}} \le \|x-y\|_L. \tag{3.2}$$

Let $g(x) = \frac{1}{2}\|x\|_L^2 - f(x)$. From previous inequalities, we have

$$\begin{aligned}
\langle \nabla f(x) - \nabla f(y), x-y\rangle &\le \|\nabla f(x) - \nabla f(y)\|_{L^{-1}} \|x-y\|_L \\
&\le \|x-y\|_L^2
\end{aligned} \tag{3.3}$$

From the gradient $Lx - \nabla f(x)$ of g, substituting $\nabla f(x) = Lx - \nabla g(x)$ to the equation (3.3),

$$\begin{aligned}
\|x-y\|_L &\ge \langle Lx - \nabla g(x) - Ly + \nabla g(y), x-y\rangle \\
&= \langle Lx - Ly, x-y\rangle - \langle \nabla g(x) - \nabla g(y), x-y\rangle \\
&= \|x-y\|_L^2 - \langle \nabla g(x) - \nabla g(y), x-y\rangle,
\end{aligned}$$

i.e.

$$\langle \nabla g(x) - \nabla g(y), x-y\rangle \ge 0.$$

Thus, g is convex. It follows from gradient inequality of g that

$$\frac{1}{2}\|y\|_L^2 - f(y) \ge \frac{1}{2}\|x\|_L^2 - f(x) + (Lx - \nabla f(x))^T(y-x), \quad \forall x, y \in \mathbb{R}^n.$$

Rearranging the above equation, we can obtain the final inequality: for any $x, y \in \mathbb{R}^n$,

$$\begin{aligned} f(y) &\leq f(x) + \frac{1}{2}\|y\|_L^2 - \frac{1}{2}\|x\|_L^2 - (Lx - \nabla f(x))^T(y-x) \\ &= f(x) + (-\frac{1}{2}\|x\|_L^2 + \frac{1}{2}\|y\|_L^2 - Lx^T(y-x)) + \nabla f(x)^T(y-x) \\ &= f(x) + \nabla f(x)^T(y-x) + (\frac{1}{2}\|x\|_L^2 + \frac{1}{2}\|y\|_L^2 - Lx^T y) \\ &= f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}\|x-y\|_L^2. \end{aligned}$$

□

By using Lemma 3.1, we can prove the sufficient decrease condition C1 of the proposed algorithm.

THEOREM 3.2. *Let x^k be generated by GPL-IRL1. If $2\bar{L} - L$ is symmetric and positive definite, then*

$$F(x^{k+1}) + \frac{\lambda}{2}\|x^{k+1} - x^k\| \leq F(x^k), \quad \forall k \in \mathbb{N},$$

where $\lambda > 0$ is the smallest eigenvalue of $2\bar{L} - L$.

Proof. For any subgradients $q_1^{k+1} \in \partial f_2(x^{k+1}), q_2^{k+1} \in \partial(p^k \circ g)(x^{k+1})$, we can obtain

$$\begin{aligned} f_2(x^{k+1}) - f_2(x^k) &\leq (q_1^{k+1})^T(x^{k+1} - x^k), \\ p^k(g(x^{k+1})) - p^k(g(x^k)) &\leq (q_2^{k+1})^T(x^{k+1} - x^k). \end{aligned}$$

It follows from [10, Lemma 1] that

$$\partial(f_2 + p^k \circ g)(x) = \partial f_2(x) + \partial(p^k \circ g)(x), \quad \forall x \in \mathbb{R}^n.$$

By optimality of the subproblem of the proposed method, there exist subgradients $q_1^{k+1} \in \partial f_2(x^{k+1})$, $q_2^{k+1} \in \partial(p^k \circ g)(x^{k+1})$ s.t.

$$\nabla f_1(x^k) + q_1^{k+1} + q_2^{k+1} + \bar{L}(x^{k+1} - x^k) = 0. \tag{3.4}$$

Then, we have

$$\begin{aligned} F(x^{k+1}) - F(x^k) &= f_1(x^{k+1}) + f_2(x^{k+1}) + h(g(x^{k+1})) \\ &\qquad -(f_1(x^k) + f_2(x^k) + h(g(x^k))) \\ &\le f_1(x^{k+1}) + f_2(x^{k+1}) + p^k(g(x^{k+1})) \\ &\qquad -(f_1(x^k) + f_2(x^k) + p^k(g(x^k))) \\ &\le \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{1}{2}\|x^{k+1} - x^k\|_L^2 \\ &\quad +(q_1^{k+1})^T(x^{k+1} - x^k) + (q_2^{k+1})^T(x^{k+1} - x^k) \\ &= (\nabla f(x^k) + q_1^{k+1} + q_2^{k+1})^T(x^{k+1} - x^k) \\ &\qquad +\frac{1}{2}\|x^{k+1} - x^k\|_L^2 \\ &= -\|x^{k+1} - x^k\|_{\bar{L}}^2 + \frac{1}{2}\|x^{k+1} - x^k\|_L^2 \\ &= -\frac{1}{2}\|x^{k+1} - x^k\|_{2\bar{L}-L}^2 \\ &\le -\frac{\lambda}{2}\|x^{k+1} - x^k\|_2^2, \end{aligned}$$

where the second inequality is from the property of p that $p^k(g(x^{k+1})) \ge h(g(x^{k+1}))$ and $p^k(g(x^k)) = h(g(x^k))$, the third inequality is obtained from the equation (3.1) and the fifth equality can be obtained from (3.4). □

Second, we prove the relative error condition(C2) of the proposed algorithm.

Theorem 3.3. *There exist a positive constant $C > 0$ and a subgradient $q^{k+1} \in \partial F(x^{k+1})$ for fixed $k \in \mathbb{N}$ such that*

$$\|q^{k+1}\|_2 \le C\|x^{k+1} - x^k\|_2.$$

Proof. By previous theorem and coercivity of F, the sequence x^k is bounded. Thus, there exists a compact and convex set containing $\{x^k\}$ in $\mathbb{R}^n$. From the convexity of g and the strictly continuity of ∇h and ∇p^k, the functions g, ∇h, ∇p^k are Lipschitz continuous on this compact and convex set containing x^k. Let L_1, L_2, L_3 be the (common) Lipschitz constants of g, ∇h and ∇p^k, respectively.

We consider the subgradients $q_1^{k+1} \in \partial f_2(x^{k+1})$, $q_2^{k+1} \in \partial(p^k \circ g)(x^{k+1})$ satisfying (3.4). With setting $y^{k+1} = \nabla p^k(g(x^{k+1}))$ and $y := \nabla h(g(x^{k+1}))$, we can decompose $q_2^{k+1} = \sum_i y_i^{k+1}\xi_i$ for some $\xi_i \in \partial g_i(x^{k+1})$ and

$\partial(h \circ g)(x^{k+1}) = \sum_i y_i \partial g_i(x^{k+1})$ by [10, Lemma 1]. We set $q_2 := \sum_i y_i \xi_i \in \partial(h \circ g)(x^{k+1})$. By Lipschitz continuity of g, we have

$$\|q_2 - q_2^{k+1}\|_2 \le \|\sum_i (y_i - y_i^{k+1})\xi_i\|_2 \le L_1\|y - y^{k+1}\|_2. \tag{3.5}$$

On the other hands, note that for any symmetric positive definite matrix P, there exist an orthogonal matrix U and a diagonal positive definite matrix D s.t. $P = UDU^T$ by Spectral theorem. Here, diagonal components of D are eigenvalues of P. For any $x \in \mathbb{R}^n$, we have

$$\begin{aligned} \|x\|_P^2 = x^T P x = x^T(UDU^T)x = (U^Tx)^T D(U^Tx) \\ \ge \lambda_n (U^Tx)(U^Tx) = \lambda_n \|x\|_2^2, \end{aligned}$$

where λ_n is the smallest eigenvalue of P. That is,

$$\|x\|_2 \le \frac{1}{\sqrt{\lambda_n}}\|x\|_P, \quad \forall x \in \mathbb{R}^n. \tag{3.6}$$

Similarly, we also obtain

$$\|x\|_P \le \sqrt{\lambda_1}\|x\|_2, \quad \forall x \in \mathbb{R}^n, \tag{3.7}$$

where λ_1 is the largest eigenvalue of P. Let $q^{k+1} := \nabla f(x^{k+1}) + q_1^{k+1} + q_2$. Then, $q^{k+1} \in \partial F(x^{k+1})$ from [10, Lemma 1]. From (3.4),

$$\begin{aligned} \|q^{k+1}\|_2 &= \|\nabla f(x^{k+1}) + q_1^{k+1} + q_2 \\ &\quad -(\nabla f(x^k) + q_1^{k+1} + q_2^{k+1} + \bar{L}(x^{k+1} - x^k))\|_2 \\ &\le \|\nabla f(x^{k+1}) - \nabla f(x^k)\|_2 + \|q_2 - q_2^{k+1}\|_2 + \|x^{k+1} - x^k\|_{\bar{L}^2} \\ &\le \sqrt{\lambda_1}\|\nabla f(x^{k+1}) - \nabla f(x^k)\|_{L^{-1}} + L_1\|y - y^{k+1}\|_2 \\ &\quad +\lambda_1\|x^{k+1} - x^k\|_2 \\ &\le \sqrt{\lambda_1}\|x^{k+1} - x^k\|_L + \lambda_1\|x^{k+1} - x^k\|_2 \\ &\quad +L_1\|y - \nabla h(g(x^k)) + \nabla p^k(g(x^k)) - y^{k+1}\|_2 \\ &\le (2\lambda_1 + L_1^2(L_2 + L_3))\|x^{k+1} - x^k\|_2, \end{aligned}$$

where the third inequality is obtained from (3.6), (3.5) and (3.7), the fourth inequality is obtained from the fact $\nabla h(g(x^k)) = \nabla p^k(g(x^k))$ and (3.2), and the last inequality is from the equation (3.7). Since $2\lambda_1 + L_1^2(L_2 + L_3)$ is positive constant, the desired results is obtained. □

THEOREM 3.4. *There exist a convergent subsequence $\{x^{k_j}\}$ of $\{x^k\}$ and a limit $\bar{x}$ such that*

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

Proof. Since $\frac{1}{2}\|x^{k+1} - x^k\|_L^2$ is convex, we can obtain the desired results by similar steps in [13, Proposition 6]. □

From Theorems 3.2-3.4, the three conditions C1-C3 are satisfied. Finally, we can obtain the following global convergence of our algorithm.

THEOREM 3.5. *Let F be a proper, l.s.c function. Let $\{x^k\}_{k\in\mathbb{R}^n}$ be generated by GPL-IRL1. If F has the KL property at a cluster point $\bar{x} := \lim_{j\to\infty} x^{k_j}$, then $\{x^k\}$ converges to $\bar{x}$ as $k \to \infty$ and $\bar{x}$ is a critical point of F. Moreover, it has finite length:*

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

4. Numerical Results

In this section, we present the numerical results comparing the proposed algorithm with the PL-IRL1 [13]. We consider an image deblurring problem as an application of the proposed method. The mathematical degraded model of image deblurring problem is given as follows,

$$f = Au + n,$$

where $f \in \mathbb{R}^{m\times n}$ or $\mathbb{R}^{mn}$ is an observed blurred and noisy image, $u \in \mathbb{R}^{m\times n}$ is an original clean image, A is a blurring linear operator and n is additive Gaussian white noise with mean 0 and standard deviation σ. A famous nonconvex total variation based model for image deblurring is given as follows,

$$\min_u \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{1}{\rho}\sum_i \log(1 + \rho(\nabla u)_i^2),$$

where λ is a regularization parameter and ρ is a positive constant that controls nonconvexity of total variation regularization. Unfortunately, the GPL-IRL1 cannot be directly applied to the above problem. Here, we adopt the axillary variable d and penalty technique. We consider the following nonconvex minimization problem:

$$\min_{u,d} F(u,d) = \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\mu}{2}\|d - \nabla u\|_2^2 + \frac{1}{\rho}\sum_i \log(1 + \rho(d)_i^2), \tag{4.1}$$

where $\mu > 0$ is a penalty parameter. We set

$$f_1(u,d) = \frac{\mu}{2}\|d - \nabla u\|_2^2,\ f_2(u,d) = \frac{\lambda}{2}\|Au - f\|_2^2,$$
$$g(u,d) = d^2,\ h(y) = \frac{1}{\rho}\sum_i \log(1+\rho(y)_i).$$

Since the objective function of the problem (4.1) is definable in the log-exp o-minimal structure, it is a KL function. Trivially, it is coercive and closed. Moreover, f_1 is convex, proper and continuously differentiable function with co-coercive gradient. Specifically,

$$\nabla f_1 = \begin{bmatrix} \mu\nabla^T(\nabla u - d) \\ \mu(d - \nabla u) \end{bmatrix} = \begin{bmatrix} \mu\nabla^T\nabla & -\mu\nabla^T \\ -\mu\nabla & \mu I \end{bmatrix}\begin{bmatrix} u \\ d \end{bmatrix}.$$

Let $B = \begin{bmatrix} \mu\nabla^T\nabla & -\mu\nabla^T \\ -\mu\nabla & \mu I \end{bmatrix}$. Then, we have

$$\begin{aligned} & \langle \nabla f_1(u_1,d_1) - \nabla f_1(u_2,d_2), (u_1,d_1)^T - (u_2,d_2)^T \rangle \\ = \ & \langle B(u_1,d_1)^T - B(u_2,d_2)^T, (u_1,d_1) - (u_2,d_1) \rangle. \end{aligned}$$

Since B is a symmetric and positive semidefinite matrix, there exist an orthogonal matrix U and diagonal matrix D such that $B = U^TDU$ by spectral theorem. Let $D' = \max(D,t)$ for small positive value $t > 0$ and $L = U^TD'U$. Then,

$$B - BL^{-1}B = U^T(D - D^2D'^{-1})U,$$

and all digaonal elements of $D - D^2D'^{-1}$ are nonnegative. So, $B - BL^{-1}B$ is postivie semidefinite.

$$\begin{aligned} & \begin{bmatrix} (u_1-u_2)^T & (d_1-d_2)^T \end{bmatrix}(B - BL^{-1}B)\begin{bmatrix} u_1-u_2 \\ d_1-d_2 \end{bmatrix} \geq 0 \\ \Leftrightarrow\ & \begin{bmatrix} (u_1-u_2)^T & (d_1-d_2)^T \end{bmatrix} B \begin{bmatrix} u_1-u_2 \\ d_1-d_2 \end{bmatrix} \\ & \geq \begin{bmatrix} (u_1-u_2)^T & (d_1-d_2)^T \end{bmatrix} BL^{-1}B \begin{bmatrix} u_1-u_2 \\ d_1-d_2 \end{bmatrix} \\ \Leftrightarrow\ & \begin{bmatrix} (u_1-u_2)^T & (d_1-d_2)^T \end{bmatrix} B \begin{bmatrix} u_1-u_2 \\ d_1-d_2 \end{bmatrix} \\ & \geq \|B(u_1,d_1)^T - B(u_2,d_2)^T\|_{L^{-1}}^2. \end{aligned}$$

Hence, ∇f_1 is co-coercive with L. Here we set $t = eps$.

Clearly, f_2 is a proper and convex, l.s.c. function. $h : \mathbb{R}_+^{2mn} \to \mathbb{R}$ is coordinatewise nondecreasing, continuously differentiable and concave.

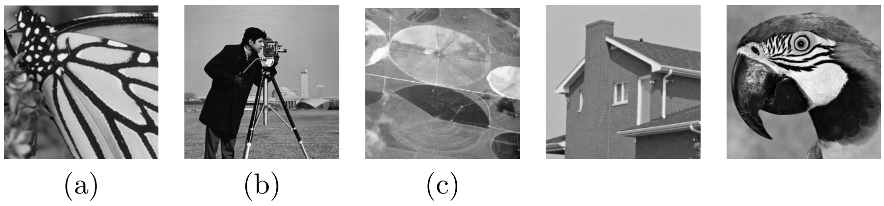

FIGURE 1. Test images

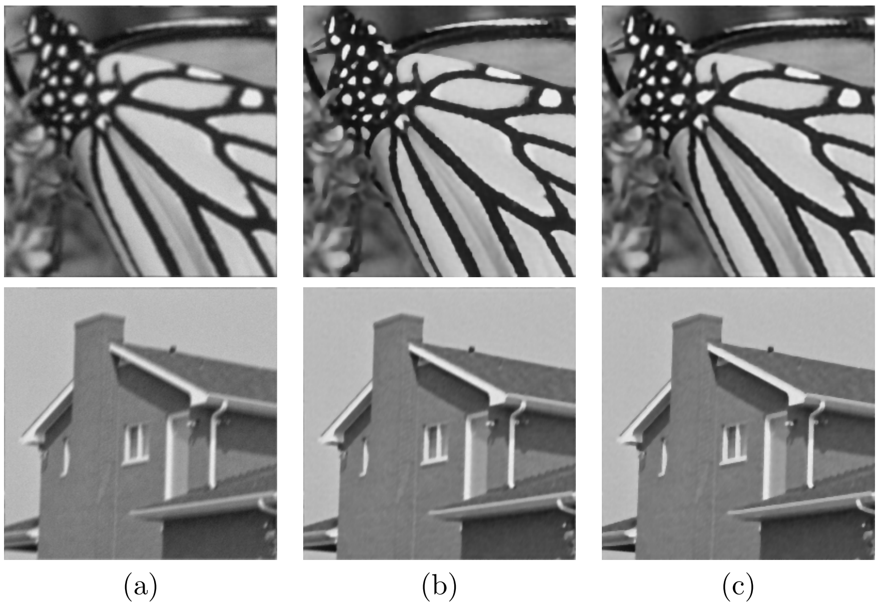

FIGURE 2. Deblurring results. Top : uniform blur, Bottom : motion blur, (a) Blurry and Noisy images, (b) PL-IRL1, (c) GPL-IRL1.

g is coordinatewise convex, proper, l.s.c. Thus, we can apply GPL-IRL1 to the problem (4.1), which is given as follows:

$$\begin{cases} w^k = \dfrac{1}{1+\rho(d^k)^2} \\ (u^{k+1}, d^{k+1}) = \arg\min\limits_{u,d} \langle w^k, d^2\rangle + \dfrac{\lambda}{2}\|Au - f\|_2^2 + \langle \mu(d^k - \nabla u^k), d\rangle \\ \qquad\qquad + \langle \mu \nabla^T(\nabla u^k - d^k), u\rangle + \|(u,d)^T - (u^k, d^k)^T\|_{\bar{L}}^2. \end{cases}$$

TABLE 1. Comparison results with noise level $\sigma = 0.01$.

Model	PL-IRL1	GPL-IRL1
Blur Type	Uniform Blur	
Image	time / Iterations / Energy / PSNR	time / Iterations / Energy / PSNR
Butterfly	3.52 / 381 / 182 / 29.03	2.07 / 223 / 164 / 30.14
Cameraman	2.96 / 408 / 105 / 25.22	1.51 / 218 / 96 / 26.08
Fields	2.01 / 212 / 77 / 30.60	1.10 / 119 / 74 / 31.26
House	2.51 / 351 / 83 / 31.21	1.22 / 174 / 78 / 32.22
Parrot	2.47 / 355 / 123 / 25.44	1.53 / 221 / 111 / 26.34
Blur Type	Motion Blur	
Image	time / Iterations / Energy / PSNR	time / Iterations / Energy / PSNR
Butterfly	3.70 / 393 / 182 / 31.27	1.89 / 202 / 168 / 31.88
Cameraman	2.89 / 409 / 114 / 27.97	1.37 / 201 / 107 / 28.73
Fields	1.99 / 210 / 80 / 31.75	1.00 / 107 / 78 / 32.21
House	2.35 / 325 / 83 / 32.34	1.19 / 169 / 78 / 33.19
Parrot	2.53 / 346 / 133 / 27.95	1.52 / 217 / 123 / 28.87

To solve the subproblem exactly, $\bar{L}$ is set to be

$$\bar{L} = \begin{bmatrix} \mu\nabla^T\nabla + \delta I & \mathbf{0} \\ \mathbf{0} & \mu I \end{bmatrix},$$

with small positive value of $\delta > t$. Then, $\bar{L}$ is symmetric and positive definite under periodic boundary condition and $\bar{L} - \frac{L}{2}$ is also positive definite. The optimality conditions of the convex subproblem in GPL-IRL1 are given as follows:

$$2w^k d + \mu(d^k - \nabla u^k) + \mu(d - d^k) = 0 \tag{4.2}$$

$$\mu\nabla^T(\nabla u^k - d^k) + \lambda A^T(Au - f) + (\mu\nabla^T\nabla + \delta I)(u - u^k) = 0. \tag{4.3}$$

The problem (4.3) is a linear equation. Note that $\mu\nabla^T\nabla + \delta I + \lambda A^T A$ is diagonalizable by 2D-fast Fourier transform (FFT) under periodic boundary condition. Hence, each problem has a closed form solution:

$$\begin{aligned} d^{k+1} &= \frac{\mu\nabla u^k}{(2w^k + \mu)}, \\ u^{k+1} &= \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mu\nabla^T d^k + \lambda A^T f + \delta u^k)}{\mathcal{F}(\mu\nabla^T\nabla + \delta I + \lambda A^T A)}\right), \end{aligned}$$

where $\mathcal{F}$ is 2D FFT.

In this experiment, we test on 5 images in Figure 1. We use two types of blur kernels: a uniform blur kernel of size 5×5 and a motion blur with size 5×5 and a $60°$ degree angle. The standard deviation of additive

Gaussian noise is set to 0.01. The regularization parameter λ is fixed as $\lambda = 10$ and penalty parameter μ is set to be 20 for all cases. We also set $\delta = 10^{-5}$. In Tables 1, we present the computing time, number of iterations, final energy and PSNR values. In this experiment, we assume that the image intensity range is [0,1]. Figure 2 illustrates the degraded images corrupted by blurring and Gaussian noise and resulting images of the proposed method and PL-IRL1.

First, it can be observed that the GPL-IRL1 is faster and uses a smaller number of iterations than the PL-IRL1 for all tests. Furthermore, the final energy value of the proposed algorithm is smaller than that of PL-IRL1 in all cases This shows that the proposed method minimizes the energy function more effectively. Furthermore, the GPL-IRL1 provides slightly higher PSNR values than PL-IRL1. Thus, the proposed algorithm finds better solutions than PL-IRL1. In Figure 2, we can also observe that the GPL-IRL1 produces slightly better restored images visually than the PL-IRL1. In conclusion, our algorithm gives better performance than PL-IRL1 in terms of accuracy, speed, and optimization of the energy function.

5. Conclusion

In this article, we proposed generalized proximal linearized iteratively reweighted ℓ_1 algorithm for solving the nonconvex and nonsmooth minimization problem (1.1). We extended the existing proximal linearized iteratively reweighted ℓ_1 algorithm by assuming that the co-coerciveness of the gradient of the continuously differentiable term of the objective function in (1.1). Based on unified framework, we proved the global convergence of the proposed method. The numerical results related to image deblurring problem showed that the proposed method has outstanding performance of restoration compared with proximal linearized iteratively reweighted ℓ_1 algorithm.

References

[1] H. Attouch, J. Bolte, B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods,* Math. Program., **137**(2013), 91-129.

[2] J. Bolte, and A. Daniilidis, and A. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems,* SIAM J. Optim., **17**(2007), 1205-1223.

[3] E. J. Candes, M. B. Wakin, and S. P. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization,* J. Fourier Anal. Appl., **14**(2008), 877-905.

[4] P. Gong, C. Zhang, Z. Lu, J. Huang, and J. Ye, *A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems,* international conference on machine learning, 2013.

[5] H. Mingyi, Z.Q. Luo, and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems,* SIAM J. Optim., **26**(2016), 337-364.

[6] A. Kaplan, and R. Tichatschke, *Proximal point methods and nonconvex optimization,* J. Glob. Optim., **13**(1998), 389-406.

[7] K. Kurdyka, *On gradients of functions definable in o-minimal structures,* Annales. Institut. Fourier, **48**(1998), 769-783.

[8] S. Łojasiewicz, *Sur la géométrie semi-et sous-analytique*, Annales. Institut. Fourier, **43**(1993), 1575-1595.

[9] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course,* Springer, 2004.

[10] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, *On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision,* SIAM J. Imaging Sci., **8**(2015), 331-372.

[11] L. P. D. Van den Dries, *Tame topology and o-minimal structures,* Cambridge university press, 1998.

[12] A. Wilkie, *Model Completeness Results for Expansions of the Ordered Field of Real Numbers by Restricted Pfaffian Functions and the Exponential Function,* J. Amer. Math. Soc., **9**(1996), 1051-1094.

[13] J. Yeo, M. Kang, *Proximal Linearized Iteratively Reweighted Algorithms for Nonconvex and Nonsmooth Optimization Problem,* Axioms, **11**(2022), 201.

[14] W. Yu, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization,* J. Sci. Comput., **78**(2019), 29-63.

Myeongmin Kang
Department of Mathematics
Chungnam National University
Daejon 305-764, Republic of Korea
E-mail: myeongminkang@cnu.ac.kr