AIMS Mathematics, 11(6): 16414–16447.
DOI: 10.3934/math.2026674
Received: 27 February 2026
Revised: 12 May 2026
Accepted: 18 May 2026
Published: 09 June 2026

https://www.aimspress.com/journal/Math

Research article

An inertial generalized iteratively reweighted ℓ_1 algorithm for nonconvex and nonsmooth optimization

Myeongmin Kang*

Department of Mathematics, Chungnam National University, Daejeon, 34134, Korea

* **Correspondence:** Email: myeongminkang@cnu.ac.kr; Tel: +820428216532.

Abstract: We study a class of nonconvex and nonsmooth optimization problems arising in sparse recovery and related applications, which are often addressed using iteratively reweighted ℓ_1 (IRL1)-type algorithms. Classical IRL1 methods are typically developed under a Lipschitz gradient assumption, which may limit their applicability. In this paper, we propose a generalized iteratively reweighted ℓ_1 algorithm with inertial extrapolation (GIRL1E), where the generalization is based on a co-coercivity condition imposed on the smooth component, thereby allowing a broader class of problems to be treated. By integrating inertial extrapolation into the generalized IRL1 framework, we establish global convergence of the proposed algorithm to a critical point of the objective function. In particular, we prove a sufficient descent property of an associated Lyapunov function and the convergence of the entire sequence without assuming convexity. Numerical experiments on compressive sensing-based signal recovery and image deblurring demonstrated that GIRL1E consistently achieves improved practical performance compared with existing IRL1 methods.

Keywords: iteratively reweighted ℓ_1 algorithm; inertial extrapolation; co-coercivity; nonconvex optimization; Kurdyka–Łojasiewicz property
Mathematics Subject Classification: 90C26, 65K10, 49M37

1. Introduction

Nonconvex and nonsmooth optimization problems arise widely in modern applied mathematics, signal processing, machine learning, and imaging sciences. Typical applications include sparse recovery, image restoration, compressed sensing, and statistical learning, where nonsmooth regularization terms such as the ℓ_1 norm or its nonconvex variants are employed to promote sparsity or structural simplicity. These models often take the form of composite objectives combining nonsmooth regularization with smooth data-fidelity terms, giving rise to challenging optimization problems that require specialized numerical algorithms. The simultaneous presence of nonconvexity

and nonsmoothness creates considerable challenges in both algorithm design and theoretical analysis, particularly in establishing convergence guarantees while maintaining computational efficiency.

We consider the following nonconvex and nonsmooth minimization problem:

$$\min_{x\in\mathbb{R}^n} F(x) := p(x) + q(x) + h(g(x)), \tag{1.1}$$

where $p : \mathbb{R}^n \to (-\infty, +\infty]$ is proper, convex, and lower semicontinuous; $q : \mathbb{R}^n \to \mathbb{R}$ is convex and continuously differentiable; $g : \mathbb{R}^n \to \mathbb{R}^m$ is a finite-valued mapping whose components are proper, lower semicontinuous, and convex; and $h : D \to \mathbb{R}$ is continuously differentiable and concave for some convex set $D \supset \mathrm{Im}(g)$. We further assume that ∇q is co-coercive with respect to a symmetric positive definite matrix L, and that h has a Lipschitz continuous gradient and is coordinate-wise nondecreasing. Finally, we assume that the objective function is coercive, lower semicontinuous, bounded from below, and definable in an o-minimal structure [1–3]. The structure of (1.1) separates the objective into components with distinct roles: p corresponds to a possibly nonsmooth convex component handled via proximal steps, q is a smooth convex function with a co-coercive gradient, and $h \circ g$ represents a composite (possibly nonconvex) term treated via iterative reweighting. This decomposition plays a central role in the algorithm design and convergence theory, and underlies the numerical experiments. Such formulations arise in a wide range of applications, including image and signal processing.

Among existing numerical approaches, the iteratively reweighted ℓ_1 (IRL1) algorithm [4, 5] is a prominent method for handling nonconvex sparsity-promoting regularization. By iteratively solving weighted convex subproblems, IRL1 transforms a nonconvex objective into a sequence of tractable updates. This procedure can be interpreted as a local linearization of concave penalties and, more generally, as an instance of the majorization–minimization (MM) principle [6]. Despite strong empirical performance in compressive sensing and statistical learning, classical IRL1 schemes may exhibit relatively slow practical convergence when high-accuracy solutions are required, particularly in large-scale settings. Related extensions of IRL1 have also been developed through iteratively linearized reweighted ADMM-type frameworks, such as the work of Sun et al. [7], which provides an important contribution to nonconvex optimization by integrating reweighting, linearization, and alternating minimization strategies. Our work is inspired by this broader line of research, while further exploring proximal inertial extensions under co-coercive assumptions and establishing corresponding convergence guarantees.

From a broader optimization perspective, various frameworks have been developed for nonconvex problems depending on structural assumptions. For smooth nonconvex optimization, descent-based methods guarantee convergence to first-order stationary points under Lipschitz continuity of the gradient [8]. For structured nonconvex objectives, MM and difference-of-convex (DC) programming provide systematic surrogate-based treatments of nonconvexity [9]. Proximal and splitting approaches have also been extensively studied for composite nonsmooth models. Under suitable regularity conditions, global convergence can be established via the Kurdyka–Łojasiewicz (KL) property [10,11]. Representative examples include inertial proximal algorithms such as iPiano [12], where convergence is obtained by combining descent properties with the KL framework.

To accelerate first-order schemes, inertial and extrapolation techniques have been widely employed. Nesterov-type methods achieve improved convergence rates in convex settings [13,14], and extensions to smooth nonconvex problems guarantee convergence to stationary points without convexity assumptions [8]. These schemes admit interpretations as discretizations of second-order dissipative

dynamical systems [15]. Inertial proximal-type methods have also been analyzed in nonsmooth nonconvex settings under the KL framework [16–18], providing a solid theoretical foundation for acceleration within proximal paradigms.

Iteratively reweighted algorithms form another important class of methods for nonsmooth nonconvex optimization. A significant contribution in this direction is provided by [19], which established global convergence of a broad class of iteratively reweighted algorithms to critical points via variational analysis and the KL property. Subsequent works introduced linearized and extrapolated IRL1 variants to enhance efficiency and scalability [20, 21].

Despite these developments, inertial acceleration and IRL1 methodologies have largely evolved independently. Existing IRL1 schemes have primarily focused on non-inertial settings, whereas inertial techniques have mainly been developed within proximal or splitting frameworks. In particular, a unified convergence analysis for inertial IRL1 algorithms involving both nonsmooth convex components and smooth terms—beyond fully linearized formulations—has not yet been fully established. Moreover, when smooth components satisfy structural conditions such as co-coercivity [22], the interplay between reweighting and inertial extrapolation has not been systematically clarified. These observations motivate the present work.

Building upon the extrapolated IRL1 framework of [21], we consider the more general problem (1.1), in which the nonsmooth convex term is treated via a nonlinearized proximal step and the smooth component satisfies a co-coercivity condition. This structural extension enables the design of an inertial IRL1 algorithm that integrates extrapolation with a nonlinearized proximal treatment. The co-coercivity assumption allows us to control inertial effects without fully linearizing the smooth term, leading to the construction of an appropriate Lyapunov function and the establishment of whole-sequence convergence to a critical point. Numerical experiments demonstrate improved practical convergence behavior compared with non-inertial IRL1 and existing extrapolated variants based on linearized subproblems.

The main contributions of this paper are summarized as follows:

- We propose an inertial IRL1 algorithm for a class of nonsmooth nonconvex composite problems, combining inertial extrapolation with a nonlinearized proximal step.
- Under a co-coercivity assumption on the smooth component, we construct a Lyapunov function and establish whole-sequence convergence to a critical point.
- The analysis extends existing extrapolated IRL1 methods beyond linearized reweighted subproblems to a nonlinearized proximal framework.
- Numerical experiments demonstrate enhanced practical efficiency compared with non-inertial IRL1 and existing extrapolated variants based on linearized subproblems.

The remainder of this paper is organized as follows. In Section 2, we introduce the mathematical preliminaries and basic assumptions that are required for the subsequent analysis, including notation and variational analytic tools. Section 3 presents the proposed inertial iteratively reweighted ℓ_1 algorithm and establishes its convergence properties. Numerical experiments are reported in Section 4 to demonstrate the effectiveness and computational advantages of the proposed method in comparison with existing approaches. Finally, Section 5 concludes the paper with a summary of the main results and a discussion of possible future research directions. Additional technical details and proofs are provided in the Appendix.

2. Mathematical preliminaries

This section recalls several notions from variational analysis and nonconvex optimization that will be used to establish the convergence of the proposed inertial iteratively reweighted ℓ_1 algorithm.

Throughout the paper, $\mathbb{R}^n$ denotes the n-dimensional Euclidean space equipped with the standard inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|_2$. For a proper function $f:\mathbb{R}^n\to(-\infty,+\infty]$, its effective domain is defined as $\operatorname{dom} f := \{x\in\mathbb{R}^n : f(x) < +\infty\}$.

To characterize stationarity in nonsmooth and nonconvex settings, we adopt the limiting (Mordukhovich) subdifferential [10, 23]. Let $f:\mathbb{R}^n\to(-\infty,+\infty]$ be a proper lower semicontinuous function and let $x\in\operatorname{dom} f$. We first recall the Fréchet (regular) subdifferential

$$\hat{\partial} f(x) := \Big\{ v\in\mathbb{R}^n : \liminf_{y\to x,\, y\neq x} \frac{f(y)-f(x)-\langle v, y-x\rangle}{\|y-x\|_2} \geq 0 \Big\},$$

and define the limiting subdifferential at x by

$$\partial f(x) := \Big\{ v\in\mathbb{R}^n : \exists\, x^k\to x,\ f(x^k)\to f(x),\ v^k\to v,\ \text{and } v^k\in\hat{\partial} f(x^k) \Big\}.$$

A point $x\in\operatorname{dom}\partial f$ is called a *critical point* of f if $0\in\partial f(x)$, which generalizes the first-order optimality condition. We also use the notation $\operatorname{dist}(0,\partial f(x)) := \inf\{\|v\|_2 : v\in\partial f(x)\}$.

The limiting subdifferential enjoys several fundamental properties that will be used throughout the paper. In particular, it is closed in the sense that if $x^k\to x$, $f(x^k)\to f(x)$, and $v^k\in\partial f(x^k)$ with $v^k\to v$, then $v\in\partial f(x)$. Moreover, any local minimizer x of f satisfies $0\in\partial f(x)$ (Fermat rule). When f is convex, ∂f coincides with the classical convex subdifferential. We will also use a basic sum rule stating that, under standard qualification conditions,

$$\partial(f+g)(x)\subseteq\partial f(x)+\partial g(x).$$

In particular, if one of the functions is continuously differentiable, this inclusion holds automatically. These properties follow from standard results in variational analysis; see [23].

While the above sum rule provides an inclusion in general, the specific structure of the problem considered here allows us to obtain a more explicit form of the subdifferential.

Let $H(x) := \theta(\Psi(x))$, where $\theta:\mathbb{R}^m\to\mathbb{R}$ is continuously differentiable and $\Psi:\mathbb{R}^n\to(-\infty,+\infty]^m$ is defined by

$$\Psi(x) = (\Psi_1(x),\dots,\Psi_m(x)),$$

with each $\Psi_i:\mathbb{R}^n\to(-\infty,+\infty]$ being proper, lower semicontinuous, and convex. Then, for any x such that $\Psi(x)\in\mathbb{R}^m$, one has

$$\partial H(x) = \left\{ \sum_{i=1}^m \nabla_i\theta(\Psi(x))\,\xi_i : \xi_i\in\partial\Psi_i(x),\ i=1,\dots,m \right\}.$$

Moreover, for any proper, lower semicontinuous, and convex function F_1, the corresponding sum rule holds with equality:

$$\partial(F_1+H)(x) = \partial F_1(x)+\partial H(x). \tag{2.1}$$

These properties follow from standard results in variational analysis under the convexity of Ψ and the fact that θ is continuously differentiable and coordinatewise nondecreasing; see [19, Lemma 1].

While sufficient descent is necessary for convergence, it does not, in general, guarantee convergence of the whole sequence in nonconvex settings. To address this issue, we employ the Kurdyka–Łojasiewicz (KL) property, introduced in [24] and further developed in [10], which we recall below.

A proper lower semicontinuous function f is said to satisfy the KL property at $\bar{x} \in \operatorname{dom} \partial f$ if there exist $\eta > 0$, a neighborhood U of $\bar{x}$, and a continuous concave function $\varphi : [0, \eta) \to \mathbb{R}_+$ that is continuously differentiable on $(0, \eta)$ with $\varphi'(s) > 0$, such that for all $x \in U$ satisfying

$$f(\bar{x}) < f(x) < f(\bar{x}) + \eta,$$

one has

$$\varphi'(f(x) - f(\bar{x}))\ \operatorname{dist}(0, \partial f(x)) \geq 1.$$

The KL property allows one to combine a sufficient decrease condition with a relative error bound to deduce convergence of the entire sequence [10]. A broad class of functions arising in applications, including semi-algebraic and definable functions, satisfies the KL property [10], which provides the basis for verifying the KL property of the objective function and the Lyapunov function considered in this paper.

Linearization techniques are widely used in optimization to simplify subproblems arising in iterative algorithms. Typically, these methods assume that the objective function, or part of it, is continuously differentiable with a Lipschitz continuous gradient. In this work, we use the matrix co-coercivity condition introduced in [25], which extends the classical scalar co-coercivity condition to a matrix-induced metric.

A mapping $B : \mathbb{R}^n \to \mathbb{R}^n$ is said to be co-coercive with respect to a symmetric positive definite matrix L if

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

where $\|z\|^2_{L^{-1}} := \langle L^{-1}z, z\rangle$. Here and throughout the paper, for $L \succ 0$, we define $\|z\|_L := \sqrt{\langle Lz, z\rangle}$. This condition yields Lipschitz continuity as a consequence. In particular, when $L = \ell I$, it reduces to the classical scalar co-coercivity condition, which implies that B is ℓ-Lipschitz.

As a direct consequence of co-coercivity, the following inequality holds:

$$\|B(x) - B(y)\|_{L^{-1}} \leq \|x - y\|_L, \quad \forall x, y \in \mathbb{R}^n. \tag{2.2}$$

Moreover, since $L \succ 0$, (2.2) implies that B is globally Lipschitz continuous with respect to the Euclidean norm. That is, there exists a constant $L_B > 0$ such that

$$\|B(x) - B(y)\|_2 \leq L_B\|x - y\|_2, \qquad \forall x, y \in \mathbb{R}^n.$$

In particular, when $B = \nabla f$, there exists a constant $L_f > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L_f\|x - y\|_2, \qquad \forall x, y \in \mathbb{R}^n. \tag{2.3}$$

We will also use the following quadratic upper bound [26, Lemma 1] in the subsequent analysis: If $B = \nabla f$ is co-coercive, then

$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\|x - y\|^2_L. \tag{2.4}$$

This property will be used below to establish sufficient descent and to control the error induced by inertial extrapolation.

3. Proposed algorithm

In this section, we present an inertial iteratively reweighted ℓ_1 (IRL1) algorithm for solving composite nonsmooth and nonconvex optimization problems of the form (1.1), which arise in a wide range of applications, including sparse recovery, image restoration, and machine learning.

For completeness, we briefly recall the classical IRL1 scheme for nonconvex sparsity-promoting problems [4, 5]. The method updates the iterate according to

$$x^{k+1} = \arg\min_x \; f(x) + \langle \omega^k, |x| \rangle, \tag{3.1}$$

where f denotes a smooth data-fidelity term and the weights ω^k are determined by the current iterate, typically as functions of $|x^k|$.

The proposed algorithm extends classical IRL1 schemes by incorporating a proximal treatment of the nonsmooth term p without full linearization, together with an inertial extrapolation step. Existing inertial or extrapolated approaches primarily address either smooth nonconvex models or fully linearized subproblems. In contrast, the present method integrates these mechanisms within a unified framework, enabling accelerated updates while accommodating a broader class of nonsmooth and nonconvex regularizers. Such sparsity-promoting penalties are widely used in statistical learning and signal processing [27, 28].

The incorporation of inertia, however, generally breaks the monotonic decrease of the objective function and significantly complicates the convergence analysis. To address this issue, we employ a Lyapunov framework that captures both the objective value and the contribution of the inertial extrapolation. This Lyapunov function plays a central role in establishing sufficient descent and in controlling the inertial error along the iterates.

For comparison, we recall the update rule of the linearized IRL1 with extrapolation (LIRL1E) introduced in [21]:

$$x^{k+1} = \arg\min_x \langle \nabla f(y^k), x - y^k \rangle + \langle \omega^k, |x| \rangle + \frac{1}{2\alpha_k} \|x - y^k\|_2^2, \tag{3.2}$$

where ω^k denotes the vector of reweighting coefficients and f denotes the smooth part of the objective.

Algorithm 1 summarizes the proposed procedure. At each iteration, the method performs a weighted proximal update for the nonsmooth component, incorporates inertial extrapolation to accelerate convergence, and handles the smooth component through a first-order approximation at the extrapolated point.

It follows from [19] that the function $\hat{h}^k$ defined by

$$\hat{h}^k(y) := h(g(x^k)) + \langle \nabla h(g(x^k)),\, y - g(x^k) \rangle$$

serves as an optimal convex majorizer of h on $\mathrm{Im}(g)$. Consequently, $\hat{h}^k \circ g$ majorizes $h \circ g$ at x^k. Specifically, it is proper, convex, and nondecreasing on $\mathrm{Im}(g)$, and it satisfies

$$\hat{h}^k(y) \geq h(y), \quad \forall y \in \mathrm{Im}(g), \qquad \hat{h}^k(g(x^k)) = h(g(x^k)).$$

Due to the inertial extrapolation step, the objective function associated with problem (1.1) is not necessarily monotonically decreasing. Consequently, classical descent-based arguments are no longer applicable.

Algorithm 1 Generalized Iteratively Reweighted ℓ_1 Algorithm with Inertial Extrapolation (GIRL1E)

Require: Initial point $x^0 \in \mathbb{R}^n$, tolerance $\varepsilon > 0$
Ensure: Approximate solution x^k

1: Set $k = 0$.
2: Initialize $x^{-1} = x^0$.
3: **repeat**
4: Compute weights

$$\omega^k = \nabla h(g(x^k)).$$

5: $\bar{x}^k = x^k + \beta_k(x^k - x^{k-1})$.
6: Update x^{k+1} by solving

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} \left\{ p(x) + \langle \nabla q(\bar{x}^k), x - \bar{x}^k\rangle + \langle \omega^k, g(x)\rangle + \frac{1}{2}\|x - \bar{x}^k\|_{\bar{L}}^2 \right\},$$

7: $k \leftarrow k + 1$
8: **until** the stopping condition is satisfied.

To address this issue and to ensure consistency with the theoretical framework introduced earlier, we introduce a Lyapunov function that incorporates both the objective value and the inertial term.

The Lyapunov function $\bar{F} : \mathbb{R}^n \times \mathbb{R}^n$ is defined as

$$\bar{F}(x, y) := F(x) + \frac{1}{2}\|x - y\|_{\bar{L}}^2.$$

The additional quadratic term penalizes large inertial steps and allows us to control the error induced by extrapolation.

We now analyze the Lyapunov function along the sequence generated by Algorithm 1. In addition to the structural assumptions on p, q, g, and h introduced in Section 1, we impose the following assumptions:

- There exists a constant $\bar{\beta} \in [0, 1)$ such that $0 \le \beta_k \le \bar{\beta}$ for all $k \ge 0$.
- The matrix L is the co-coercivity matrix associated with ∇q, and $\bar{L} - L \succ 0$.

The following lemmas establish the key properties required for the convergence analysis.

We first establish a sufficient descent property of the Lyapunov function, which serves as a fundamental ingredient in the analysis.

Lemma 3.1. *The following statements hold.*

(i) There exists a constant $C > 0$ such that

$$\bar{F}(x^k, x^{k-1}) - \bar{F}(x^{k+1}, x^k) \ge C\|x^k - x^{k-1}\|_{\bar{L}}^2. \tag{3.3}$$

Consequently, $\{\bar{F}(x^k, x^{k-1})\}$ is a nonincreasing and convergent sequence.

(ii) The sequence $\{x^k\}$ is bounded and

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

Proof. (i) Let $\tilde{F}(x)$ be the objective function of the subproblem in GIRL1E:

$$\tilde{F}(x) := p(x) + \langle \nabla q(\bar{x}^k), x - \bar{x}^k \rangle + \langle \omega^k, g(x) \rangle + \tfrac{1}{2}\|x - \bar{x}^k\|_{\bar{L}}^2.$$

By the convex majorization property, $\hat{h}^k$ is convex and nondecreasing on $\mathrm{Im}(g)$. Since h is coordinate-wise nondecreasing, we have $\omega^k = \nabla h(g(x^k)) \geq 0$. Together with the coordinate-wise convexity of g, this implies that $\langle \omega^k, g(x) \rangle$ is convex. Since p is convex and $\bar{L} \succ 0$, $\tilde{F}$ is strongly convex with respect to $\bar{L}$. For any $\xi_x \in \partial \tilde{F}(x)$, the strong convexity of $\tilde{F}$ yields

$$\tilde{F}(y) \geq \tilde{F}(x) + \langle \xi_x, y - x \rangle + \tfrac{1}{2}\|y - x\|_{\bar{L}}^2.$$

Applying this inequality with $x = x^{k+1}$ and $y = x^k$, together with $\mathbf{0} \in \partial \tilde{F}(x^{k+1})$, yields

$$\tilde{F}(x^{k+1}) - \tilde{F}(x^k) \leq -\tfrac{1}{2}\|x^{k+1} - x^k\|_{\bar{L}}^2.$$

Adding $q(\bar{x}^k) + h(g(x^k))$ to both sides, we obtain

$$\begin{aligned} & p(x^{k+1}) + q(\bar{x}^k) + \langle \nabla q(\bar{x}^k), x^{k+1} - \bar{x}^k \rangle + \langle \omega^k, g(x^{k+1}) - g(x^k) \rangle + h(g(x^k)) + \tfrac{1}{2}\|x^{k+1} - \bar{x}^k\|_{\bar{L}}^2 \\ \leq \; & p(x^k) + q(\bar{x}^k) + \langle \nabla q(\bar{x}^k), x^k - \bar{x}^k \rangle + h(g(x^k)) + \tfrac{1}{2}\|x^k - \bar{x}^k\|_{\bar{L}}^2 - \tfrac{1}{2}\|x^{k+1} - x^k\|_{\bar{L}}^2. \end{aligned} \tag{3.4}$$

By the quadratic upper bound established in (2.4), we have

$$q(x^{k+1}) \leq q(\bar{x}^k) + \langle \nabla q(\bar{x}^k), x^{k+1} - \bar{x}^k \rangle + \tfrac{1}{2}\|x^{k+1} - \bar{x}^k\|_L^2. \tag{3.5}$$

By the majorization property of $\hat{h}^k$,

$$h(g(x^{k+1})) \leq \hat{h}^k(g(x^{k+1})) = h(g(x^k)) + \langle \omega^k, g(x^{k+1}) - g(x^k) \rangle.$$

Combining (3.4), (3.5), and the majorization property of $\hat{h}^k$, we obtain

$$\begin{aligned} & F(x^{k+1}) - \tfrac{1}{2}\|x^{k+1} - \bar{x}^k\|_L^2 + \tfrac{1}{2}\|x^{k+1} - \bar{x}^k\|_{\bar{L}}^2 \\ \leq \; & p(x^k) + q(\bar{x}^k) + \langle \nabla q(\bar{x}^k), x^k - \bar{x}^k \rangle + h(g(x^k)) + \tfrac{1}{2}\|x^k - \bar{x}^k\|_{\bar{L}}^2 - \tfrac{1}{2}\|x^{k+1} - x^k\|_{\bar{L}}^2 \\ \leq \; & F(x^k) + \tfrac{1}{2}\|x^k - \bar{x}^k\|_{\bar{L}}^2 - \tfrac{1}{2}\|x^{k+1} - x^k\|_{\bar{L}}^2, \end{aligned}$$

where the second inequality follows from the convexity of q, namely,

$$q(x^k) \geq q(\bar{x}^k) + \langle \nabla q(\bar{x}^k), x^k - \bar{x}^k \rangle.$$

Using the extrapolation relation

$$\bar{x}^k = x^k + \beta_k(x^k - x^{k-1}),$$

we have

$$\|x^k - \bar{x}^k\|_{\bar{L}}^2 = \beta_k^2 \|x^k - x^{k-1}\|_{\bar{L}}^2.$$

Therefore,

$$F(x^k) + \tfrac{1}{2}\|x^k - \bar{x}^k\|_{\bar{L}}^2 = F(x^k) + \tfrac{\beta_k^2}{2}\|x^k - x^{k-1}\|_{\bar{L}}^2.$$

Since $\bar{L} - L \succ 0$, we have

$$\tfrac{1}{2}\|x^{k+1} - \bar{x}^k\|_{\bar{L}}^2 - \tfrac{1}{2}\|x^{k+1} - \bar{x}^k\|_L^2 \geq 0.$$

Therefore,

$$\begin{aligned} F(x^{k+1}) + \tfrac{1}{2}\|x^{k+1} - x^k\|_{\bar{L}}^2 &\leq F(x^k) + \tfrac{\beta_k^2}{2}\|x^k - x^{k-1}\|_{\bar{L}}^2 \\ &\leq F(x^k) + \tfrac{1}{2}\|x^k - x^{k-1}\|_{\bar{L}}^2 - \tfrac{1-\beta_k^2}{2}\|x^k - x^{k-1}\|_{\bar{L}}^2. \end{aligned}$$

Thus,

$$\bar{F}(x^{k+1}, x^k) \leq \bar{F}(x^k, x^{k-1}) - \tfrac{1-\beta_k^2}{2}\|x^k - x^{k-1}\|_{\bar{L}}^2.$$

Hence (3.3) holds with $C = (1 - \bar{\beta}^2)/2 > 0$. Since F is bounded from below, so is $\bar{F}$, and thus $\{\bar{F}(x^k, x^{k-1})\}$ is convergent.

(ii) By the definition of $\bar{F}$ and (3.3),

$$F(x^k) \leq \bar{F}(x^k, x^{k-1}) \leq \bar{F}(x^0, x^{-1}) = F(x^0),$$

where $x^{-1} := x^0$. Hence $\{x^k\}$ is bounded by the coercivity of F. Summing (3.3) over $k = 0, \ldots, K$ yields

$$\sum_{k=0}^{K} \|x^k - x^{k-1}\|_{\bar{L}}^2 \leq \frac{1}{C}(\bar{F}(x^0, x^{-1}) - \bar{F}(x^{K+1}, x^K)).$$

Letting $K \to \infty$ and using the convergence of $\{\bar{F}(x^k, x^{k-1})\}$, we obtain

$$\sum_{k=0}^{\infty} \|x^k - x^{k-1}\|_{\bar{L}}^2 < \infty.$$

Hence $\|x^k - x^{k-1}\|_{\bar{L}} \to 0$, and therefore $\|x^{k+1} - x^k\|_{\bar{L}} \to 0$ by shifting the index. Moreover, $\|x\|_{\bar{L}}^2 \geq \lambda_{\min}(\bar{L})\|x\|_2^2$, where $\lambda_{\min}(\bar{L})$ denotes the smallest eigenvalue of $\bar{L}$. This yields $\|x^{k+1} - x^k\|_2 \to 0$. □

To analyze the behavior of the Lyapunov function $\bar{F}$ in the presence of inertial extrapolation, we establish a subgradient estimate that yields the relative error condition required for the convergence analysis.

Lemma 3.2. *There exists a constant $C_1 > 0$ such that, for all $k \geq 1$,*

$$\operatorname{dist}(\mathbf{0}, \partial\bar{F}(x^k, x^{k-1})) \leq C_1(\|x^k - x^{k-1}\|_2 + \|x^{k-1} - x^{k-2}\|_2). \tag{3.6}$$

Proof. From Lemma 3.1, the sequence $\{x^k\}$ is bounded. Hence, there exists a compact set $K \subset \mathbb{R}^n$ such that $x^k \in K$ for all k. Since the components of g are finite-valued and convex on $\mathbb{R}^n$, they are locally Lipschitz continuous, and thus Lipschitz continuous on K. Therefore, there exists a constant $M_g > 0$ such that

$$\|g(x) - g(y)\|_2 \leq M_g\|x - y\|_2, \quad \text{for all } x, y \in K.$$

Moreover, since each component of g is convex and Lipschitz continuous on K, its subgradients are uniformly bounded on K. Hence, for any $\zeta_i^k \in \partial g_i(x^k)$, $i = 1, \ldots, m$, we have

$$\|\zeta_i^k\|_2 \leq M_g.$$

By (2.3), there exists $L_q > 0$ such that

$$\|\nabla q(x) - \nabla q(y)\|_2 \leq L_q \|x - y\|_2, \quad \forall x, y \in \mathbb{R}^n.$$

Since ∇h is Lipschitz continuous, there exists a constant $M_h > 0$ such that

$$\|\nabla h(u) - \nabla h(v)\|_2 \leq M_h \|u - v\|_2, \quad \text{for all } u, v \in \mathrm{Im}(g).$$

We now construct a subgradient of $\bar{F}$. Recall that $\omega^{k-1} = \nabla h(g(x^{k-1}))$. By the optimality condition for the subproblem at iteration $k-1$ and the sum rule (2.1), there exist

$$\xi^k \in \partial p(x^k), \qquad \zeta_i^k \in \partial g_i(x^k), \quad i = 1, \ldots, m,$$

such that

$$\mathbf{0} = \nabla q(\bar{x}^{k-1}) + \xi^k + \sum_{i=1}^{m} \omega_i^{k-1} \zeta_i^k + \bar{L}(x^k - \bar{x}^{k-1}).$$

Similarly, applying (2.1) to the full objective F, we obtain

$$\nabla q(x^k) + \xi^k + \sum_{i=1}^{m} \omega_i^k \zeta_i^k \in \partial F(x^k).$$

Therefore, a subgradient of $\bar{F}$ at (x^k, x^{k-1}) is given by

$$w^k := \begin{pmatrix} \nabla q(x^k) + \xi^k + \sum_{i=1}^{m} \omega_i^k \zeta_i^k + \bar{L}(x^k - x^{k-1}) \\ \bar{L}(x^{k-1} - x^k) \end{pmatrix} \in \partial \bar{F}(x^k, x^{k-1}).$$

It suffices to bound $\|w^k\|_2$. Let

$$v^k := \nabla q(x^k) + \xi^k + \sum_{i=1}^{m} \omega_i^k \zeta_i^k + \bar{L}(x^k - x^{k-1}).$$

Using the optimality condition, we rewrite

$$v^k = \nabla q(x^k) - \nabla q(\bar{x}^{k-1}) + \sum_{i=1}^{m} (\omega_i^k - \omega_i^{k-1}) \zeta_i^k + \bar{L}(\bar{x}^{k-1} - x^{k-1}).$$

We estimate each term separately.
(1) By (2.3),

$$\|\nabla q(x^k) - \nabla q(\bar{x}^{k-1})\|_2 \leq L_q \|x^k - \bar{x}^{k-1}\|_2.$$

(2) Using the Lipschitz continuity of ∇h and g, we obtain

$$\|\omega^k - \omega^{k-1}\|_2 = \|\nabla h(g(x^k)) - \nabla h(g(x^{k-1}))\|_2 \leq M_h M_g \|x^k - x^{k-1}\|_2.$$

Hence,

$$\begin{aligned} \left\| \sum_{i=1}^{m} (\omega_i^k - \omega_i^{k-1}) \zeta_i^k \right\|_2 &\leq M_g \sum_{i=1}^{m} |\omega_i^k - \omega_i^{k-1}| \\ &\leq M_g \sqrt{m}\, \|\omega^k - \omega^{k-1}\|_2 \\ &\leq M_h M_g^2 \sqrt{m}\, \|x^k - x^{k-1}\|_2. \end{aligned}$$

(3) Since

$$\bar{x}^{k-1} = x^{k-1} + \beta_{k-1}(x^{k-1} - x^{k-2}),$$

we have

$$\|\bar{L}(\bar{x}^{k-1} - x^{k-1})\|_2 \leq \|\bar{L}\|_2\beta_{k-1}\|x^{k-1} - x^{k-2}\|_2.$$

Combining the above estimates and using

$$\|x^k - \bar{x}^{k-1}\|_2 \leq \|x^k - x^{k-1}\|_2 + \beta_{k-1}\|x^{k-1} - x^{k-2}\|_2,$$

we obtain

$$\begin{aligned}\|v^k\|_2 &\leq L_q\|x^k - \bar{x}^{k-1}\|_2 + M_h M_g^2\sqrt{m}\|x^k - x^{k-1}\|_2 + \|\bar{L}\|_2\beta_{k-1}\|x^{k-1} - x^{k-2}\|_2 \\ &\leq (L_q + M_h M_g^2\sqrt{m})\|x^k - x^{k-1}\|_2 + \beta_{k-1}(L_q + \|\bar{L}\|_2)\|x^{k-1} - x^{k-2}\|_2.\end{aligned}$$

Using the uniform bound $\beta_k \leq \bar{\beta} < 1$, there exists a constant $C_2 > 0$ such that

$$\|v^k\|_2 \leq C_2(\|x^k - x^{k-1}\|_2 + \|x^{k-1} - x^{k-2}\|_2).$$

Since the second component of w^k satisfies

$$\|\bar{L}(x^{k-1} - x^k)\|_2 \leq \|\bar{L}\|_2\|x^k - x^{k-1}\|_2,$$

there exists a constant $C_1 > 0$ such that

$$\|w^k\|_2 \leq C_1(\|x^k - x^{k-1}\|_2 + \|x^{k-1} - x^{k-2}\|_2).$$

Since

$$\operatorname{dist}(\mathbf{0}, \partial\bar{F}(x^k, x^{k-1})) \leq \|w^k\|_2,$$

this yields (3.6). □

As a consequence of Lemmas 3.1 and 3.2, we obtain the following corollary.

Corollary 3.1. *The sequence $\{x^k\}$ is asymptotically critical for F, i.e.,*

$$\operatorname{dist}(0, \partial F(x^k)) \to 0, \quad as \quad k \to \infty.$$

Proof. By Lemma 3.1(ii) and Lemma 3.2, we have

$$\operatorname{dist}(0, \partial\bar{F}(x^k, x^{k-1})) \to 0 \quad \text{and} \quad \|x^k - x^{k-1}\|_2 \to 0.$$

Using the definition of $\bar{F}$, for any $(u, v) \in \partial\bar{F}(x^k, x^{k-1})$, the first component satisfies

$$u - \bar{L}(x^k - x^{k-1}) \in \partial F(x^k).$$

Thus,

$$\operatorname{dist}(0, \partial F(x^k)) \leq \|u - \bar{L}(x^k - x^{k-1})\|_2 \leq \|(u, v)\|_2 + \|\bar{L}\|_2\|x^k - x^{k-1}\|_2.$$

Taking the infimum over $(u, v) \in \partial\bar{F}(x^k, x^{k-1})$, we obtain

$$\operatorname{dist}(0, \partial F(x^k)) \leq \operatorname{dist}(0, \partial\bar{F}(x^k, x^{k-1})) + \|\bar{L}\|_2\|x^k - x^{k-1}\|_2.$$

Passing to the limit yields

$$\operatorname{dist}(0, \partial F(x^k)) \to 0.$$

□

We now establish several properties of the limit sets associated with the sequence $\{x^k\}$.

Lemma 3.3. *Let $\mathcal{A}$ be the set of limit points of $\{x^k\}$ and $\widetilde{\mathcal{A}}$ be the set of limit points of $\{(x^k, x^{k-1})\}$. Then,*

(i) $\mathcal{A}$ is a nonempty compact connected set and

$$\operatorname{dist}(x^k, \mathcal{A}) \to 0 \text{ as } k \to +\infty.$$

(i)' $\widetilde{\mathcal{A}}$ is a nonempty compact connected set and

$$\operatorname{dist}((x^k, x^{k-1}), \widetilde{\mathcal{A}}) \to 0 \text{ as } k \to +\infty.$$

(ii) $\mathcal{A} \subset \operatorname{crit}(F)$.
(ii)' $\widetilde{\mathcal{A}} \subset \operatorname{crit}(\bar{F})$.
(iii) The function F is finite and constant on $\mathcal{A}$, equal to $\lim_{k\to\infty} F(x^k)$.
(iii)' The function $\bar{F}$ is finite and constant on $\widetilde{\mathcal{A}}$, equal to $\inf_k \bar{F}(x^k, x^{k-1}) = \lim_{k\to\infty} \bar{F}(x^k, x^{k-1})$.

Proof. (i) Since $\{x^k\}$ is bounded and $\|x^{k+1} - x^k\|_2 \to 0$, a standard result on bounded sequences with vanishing successive differences implies that the set of accumulation points $\mathcal{A}$ is nonempty, compact, and connected, and

$$\operatorname{dist}(x^k, \mathcal{A}) \to 0.$$

(i)' Let $z^k = (x^k, x^{k-1})$. Since $\{z^k\}$ is bounded and $\|z^{k+1} - z^k\|_2 \to 0$, the same standard result implies that the set of accumulation points $\widetilde{\mathcal{A}}$ is nonempty, compact, and connected, and

$$\operatorname{dist}(z^k, \widetilde{\mathcal{A}}) \to 0.$$

Equivalently,

$$\operatorname{dist}((x^k, x^{k-1}), \widetilde{\mathcal{A}}) \to 0.$$

(ii) Let x^* be a point in $\mathcal{A}$. Then, there exists a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to x^*. By the optimality of the subproblem in GIRL1E, there exist $\xi^{k_j} \in \partial p(x^{k_j})$ and $\zeta_i^{k_j} \in \partial g_i(x^{k_j})$, $i = 1, \ldots, m$, such that

$$\mathbf{0} = \nabla q(\bar{x}^{k_j-1}) + \xi^{k_j} + \sum_{i=1}^{m} \omega_i^{k_j-1} \zeta_i^{k_j} + \bar{L}(x^{k_j} - \bar{x}^{k_j-1}).$$

Let $d_{k_j}(x) := \frac{1}{2}\|x - \bar{x}^{k_j-1}\|_{\bar{L}}^2$. Since $\{x^k\}$ is bounded and $\sup_k \beta_k < 1$, the auxiliary sequence $\{\bar{x}^k\}$ is also bounded. Hence, we may choose a compact set K containing both $\{x^k\}$ and $\{\bar{x}^k\}$.

Then d_{k_j} is continuously differentiable, and its gradient is uniformly bounded on K. Since the components of g are finite-valued and convex on $\mathbb{R}^n$, they are locally Lipschitz continuous, and hence g is Lipschitz continuous on K. Moreover, since $\{x^k\}$ is bounded and g is continuous on K, the sequence $\{g(x^k)\}$ is bounded. Hence, by the continuity of ∇h, the weights $\omega^k = \nabla h(g(x^k))$ are bounded. Therefore, the affine functions $\hat{h}^k$ are uniformly Lipschitz continuous on the relevant bounded range. Together with the Lipschitz continuity of g on K, this implies that $\hat{h}^{k_j-1} \circ g$ is Lipschitz continuous on K. Consequently, $d_{k_j} + \hat{h}^{k_j-1} \circ g$ is Lipschitz continuous on a neighborhood of K. Hence, the subgradients of $d_{k_j} + \hat{h}^{k_j-1} \circ g$ at the points x^{k_j} are uniformly bounded. Since $\{\bar{x}^k\}$ is bounded and ∇q is continuous,

the sequence $\{\nabla q(\bar{x}^{k_j-1})\}$ is bounded. Using the optimality condition, together with the boundedness of $\sum_{i=1}^{m} \omega_i^{k_j-1} \zeta_i^{k_j}$ and $\bar{L}(x^{k_j} - \bar{x}^{k_j-1})$, we conclude that $\{\xi^{k_j}\}$ is bounded. Consequently,

$$\lim_{j\to\infty} \langle \xi^{k_j}, x^* - x^{k_j} \rangle = 0.$$

By convexity of p,

$$p(x^*) \geq p(x^{k_j}) + \langle \xi^{k_j}, x^* - x^{k_j} \rangle.$$

Combining the lower semicontinuity of F, the continuity of q, and the local Lipschitz continuity of $h \circ g$, we obtain

$$\begin{aligned}
F(x^*) &\leq \liminf_{j\to\infty} F(x^{k_j}) \\
&\leq \limsup_{j\to\infty} F(x^{k_j}) \\
&\leq \limsup_{j\to\infty} p(x^{k_j}) + \limsup_{j\to\infty} q(x^{k_j}) + \limsup_{j\to\infty} (h \circ g)(x^{k_j}) \\
&= q(x^*) + \limsup_{j\to\infty} p(x^{k_j}) + \lim_{j\to\infty} \langle \xi^{k_j}, x^* - x^{k_j} \rangle + (h \circ g)(x^*) \\
&= q(x^*) + \limsup_{j\to\infty} \left\{ p(x^{k_j}) + \langle \xi^{k_j}, x^* - x^{k_j} \rangle \right\} + (h \circ g)(x^*) \\
&\leq q(x^*) + p(x^*) + (h \circ g)(x^*) \\
&= F(x^*).
\end{aligned}$$

Therefore, $\lim_{j\to\infty} F(x^{k_j}) = F(x^*)$. Moreover, by Corollary 3.1, $\operatorname{dist}(0, \partial F(x^{k_j})) \to 0$. By the closedness of the limiting subdifferential ∂F, it follows that $0 \in \partial F(x^*)$, that is, x^* is a critical point of F.

(ii)' Let $(x^*, y^*) \in \widetilde{\mathcal{A}}$. Then there exists a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ such that $(x^{k_j}, x^{k_j-1}) \to (x^*, y^*)$. Since $\|x^k - x^{k-1}\|_{\bar{L}} \to 0$ by the proof of Lemma 3.1(ii), we obtain $x^* = y^*$. Moreover, as established in the proof of (ii), $F(x^{k_j}) \to F(x^*)$. Therefore,

$$\bar{F}(x^{k_j}, x^{k_j-1}) = F(x^{k_j}) + \frac{1}{2}\|x^{k_j} - x^{k_j-1}\|_{\bar{L}}^2 \to F(x^*) = \bar{F}(x^*, x^*).$$

By Lemma 3.2 and Lemma 3.1(ii), $\operatorname{dist}(0, \partial \bar{F}(x^{k_j}, x^{k_j-1})) \to 0$. Using the closedness of $\partial \bar{F}$, we deduce that

$$0 \in \partial \bar{F}(x^*, x^*),$$

so that (x^*, x^*) is a critical point of $\bar{F}$.

(iii) From Lemma 3.1(i), the sequence $\{\bar{F}(x^k, x^{k-1})\}$ converges. Furthermore, as $\|x^k - x^{k-1}\|_{\bar{L}} \to 0$ by Lemma 3.1(ii), we deduce that $\{F(x^k)\}$ also converges. Let

$$F^* := \lim_{k\to\infty} F(x^k),$$

and let $x^* \in \mathcal{A}$ be a limit point of $\{x^k\}$. Then, there exists a subsequence $\{x^{k_j}\}$ such that $x^{k_j} \to x^*$. As established in the proof of (ii), $F(x^{k_j}) \to F(x^*)$. Thus,

$$F(x^*) = \lim_{j\to\infty} F(x^{k_j}) = \lim_{k\to\infty} F(x^k) = F^* < \infty.$$

Therefore, F takes the constant finite value F^* on $\mathcal{A}$.

(iii)' Let $\bar{F}^* := \lim_{k\to\infty} \bar{F}(x^k, x^{k-1})$ and let $(x^*, x^*) \in \widetilde{\mathcal{A}}$. Then there exists a subsequence (x^{k_j}, x^{k_j-1}) converging to (x^*, x^*). By (ii)', $\bar{F}(x^{k_j}, x^{k_j-1}) \to \bar{F}(x^*, x^*)$. Since $\bar{F}(x^k, x^{k-1})$ is nonincreasing, its limit coincides with its infimum. Therefore,

$$\bar{F}(x^*, x^*) = \inf_k \bar{F}(x^k, x^{k-1}) = \bar{F}^* < +\infty.$$

Consequently, $\bar{F}$ is finite and constant on $\widetilde{\mathcal{A}}$. □

Theorem 3.1. *Assume that F is a KL function. Then,*

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

and $\{x^k\}$ converges to a critical point of problem (1.1).

The proof follows standard KL convergence arguments similar to those in [21] and is deferred to Appendix A for completeness.

4. Numerical experiments

In this section, we illustrate applications of the proposed algorithms to problems arising in signal and image processing. Specifically, we consider a compressive-sensing problem in signal processing and image-deblurring problems in image processing. In Subsection 4.1, numerical experiments are conducted for the compressive-sensing problem to compare the performance of the proposed method with those of IRL1 and LIRL1, with a particular focus on examining the effect of inertial extrapolation. In Subsection 4.2, the proposed algorithm is applied to image-deblurring models, which represent a fundamental class of problems in image processing. In this setting, the experiments are designed to evaluate the combined impact of inertial acceleration and the co-coercive structure within the proposed framework. All numerical experiments were performed on a 64-bit macOS system using a MacBook Pro equipped with an Intel Core i9 processor at 2.40 GHz and 32 GB of RAM.

4.1. Compressive sensing

We consider the following nonconvex minimization problem:

$$\min_{x\in\mathbb{R}^n} \frac{1}{\rho}\sum_{i=1}^{n} \log\left(1+\rho|x_i|^2\right) + \frac{\lambda}{2}\|Ax-b\|_2^2, \tag{4.1}$$

where $A \in \mathbb{R}^{m\times n}$ with $m \ll n$, $b \in \mathbb{R}^m$, and $\rho, \lambda > 0$. This problem fits into the proposed framework by setting $p \equiv 0$,

$$q(x) = \frac{\lambda}{2}\|Ax-b\|_2^2, \qquad g(x) = x^2, \qquad h(y) = \frac{1}{\rho}\sum_{i=1}^{n} \log(1+\rho y_i),$$

where $g(x) = x^2$ is understood componentwise, so that $h(g(x)) = \frac{1}{\rho}\sum_{i=1}^{n} \log(1+\rho x_i^2)$ holds. The objective function is proper, coercive, and definable in the log-exp o-minimal structure [3, 29, 30].

Hence, it is a KL function. The function q is continuously differentiable and has a Lipschitz continuous gradient with constant $L_A = \lambda\|A\|_2^2$. In particular, in the matrix notation of Section 2, ∇q is co-coercive with respect to $L_A I$. Moreover, h is continuously differentiable, concave, and coordinate-wise nondecreasing on $\mathbb{R}^n_+$, while the components of g are convex. Under the above decomposition, the assumptions required by the proposed algorithm are satisfied. To solve model (4.1), we compare the proposed algorithm with the iteratively reweighted ℓ_1 (IRL1) method [19] and its linearized variant (LIRL1) [31].

Since h is concave, we apply an iteratively reweighted ℓ_1 (IRL1) scheme by majorizing the nonconvex term $h(g(x))$ at the current iterate x^k using its first-order Taylor expansion. Ignoring constant terms, this leads to the following weighted subproblem:

$$x^{k+1} = \arg\min_x \langle \omega^k, x^2 \rangle + \frac{\lambda}{2}\|Ax - b\|_2^2, \tag{4.2}$$

where the weight vector ω^k is defined componentwise as

$$\omega^k := \nabla h(g(x^k)) = \frac{1}{1 + \rho(x^k)^2}.$$

Although (4.2) is convex, solving it may still be computationally demanding for large-scale problems. In our implementation, the associated linear system

$$(\lambda A^\top A + 2W)x^{k+1} = \lambda A^\top b$$

is solved directly at each iteration by a linear system solver, where W is a diagonal matrix with diagonal entries ω^k.

To further reduce the computational cost, we adopt a linearized IRL1 (LIRL1) scheme, in which the smooth data-fidelity term q is linearized at the current iterate x^k. Specifically, we consider the following subproblem:

$$x^{k+1} = \arg\min_x \langle \omega^k, x^2 \rangle + \langle \nabla q(x^k), x - x^k \rangle + \frac{\gamma}{2}\|x - x^k\|_2^2, \tag{4.3}$$

where $\gamma > \frac{\lambda}{2}\|A\|_2^2$ is the proximal parameter required by the convergence analysis of LIRL1 [31]. Here, $\nabla q(x) = \lambda A^\top(Ax - b)$. The minimization problem (4.3) admits a closed-form solution given by

$$x_i^{k+1} = \frac{(\gamma x_i^k - \lambda[A^\top(Ax^k - b)]_i)}{2\omega_i^k + \gamma}. \tag{4.4}$$

As a result, each iteration of LIRL1 only requires matrix–vector multiplications with A and $A^\top$, and no linear system needs to be solved.

We now apply the proposed method to model (4.1), which yields the following iterative scheme:

$$\begin{cases} \omega^k &= \dfrac{1}{1 + \rho(x^k)^2}, \\ \bar{x}^k &= x^k + \beta_k(x^k - x^{k-1}), \\ x^{k+1} &= \arg\min\limits_x \langle \omega^k, x^2 \rangle + \langle \nabla q(\bar{x}^k), x - \bar{x}^k \rangle + \dfrac{\gamma}{2}\|x - \bar{x}^k\|_2^2. \end{cases} \tag{4.5}$$

The parameter γ is selected such that $\gamma > \lambda\|A\|_2^2$, ensuring the convergence condition established in the theoretical analysis. This choice corresponds to the theoretical condition $\bar{L} > L$, with $\bar{L} = \gamma I$ and $L = L_A I$. The resulting subproblem in (4.5) can be solved explicitly as

$$x_i^{k+1} = \frac{\gamma \bar{x}_i^k - \lambda[A^\top(A\bar{x}^k - b)]_i}{2\omega_i^k + \gamma}.$$

The extrapolation parameter β_k is selected according to the update rule proposed in [21], which follows a fast iterative shrinkage-thresholding algorithm (FISTA)-type acceleration strategy:

$$\beta_k = \theta_k(\theta_{k-1}^{-1} - 1), \qquad \theta_{k+1} = \frac{2}{1 + \sqrt{1 + 4/\theta_k^2}}, \qquad \theta_0 = \theta_{-1} = 1.$$

In addition, a restart strategy is employed by setting $\theta_{k-1} = \theta_k = 1$ every 200 iterations. In this setting, it holds that $\sup_k \beta_k < 1$. This choice preserves the acceleration structure commonly used in extrapolated first-order methods.

In the compressive-sensing experiments, we set the signal dimension to $n = 4000$ and the number of measurements to $m = 600$. The measurement matrix $A \in \mathbb{R}^{m\times n}$ is generated with independent and identically distributed (i.i.d.) standard Gaussian entries, and its columns are normalized to satisfy $\|A_{:,j}\|_2 = 1$ for all $j = 1, \ldots, n$. The original signal $x^{\text{true}} \in \mathbb{R}^n$ is assumed to be sparse with $s = 0.02n = 80$ nonzero components, whose locations are selected uniformly at random. The values of the nonzero entries are drawn from the standard Gaussian distribution. The observation vector $b \in \mathbb{R}^m$ is generated according to $b = Ax^{\text{true}} + e$, where e is a Gaussian noise vector with mean zero and standard deviation $\sigma = 0.0001$. The regularization parameter is fixed as $\lambda = 80$, which provides a good balance between data fidelity and sparsity in our experiments. The parameter ρ is set to 500, and the proximal parameter γ is chosen slightly larger than $\lambda\|A\|_2^2$ (e.g., $\gamma = \lambda\|A\|_2^2 + 10^{-5}$) to satisfy the theoretical condition $\gamma > \lambda\|A\|_2^2$. These parameter values are selected to ensure stable convergence and are kept fixed across all experiments.

The stopping criterion for GIRL1E is given by

$$(2\rho\|x^{k+1}\|_2\|x^{k+1} - x^k\|_2 + 2\gamma\|x^{k+1} - \bar{x}^k\|_2) < \text{tol}\|x^{k+1}\|_2 \quad \text{or} \quad k > \text{maxiter}. \tag{4.6}$$

Note that the first inequality in (4.6) implies

$$\text{dist}(\mathbf{0}, \partial F(x^{k+1})) \le \text{tol}\,\|x^{k+1}\|_2,$$

which provides a theoretically motivated stopping rule. For fair comparison, the competing methods employ stopping criteria with comparable accuracy thresholds. The same tolerance level is used across all methods whenever applicable. In this experiment, we set $\text{tol} = 3 \cdot 10^{-3}$ and $\text{maxiter} = 10,000$. In practice, the algorithms converge well before reaching the maximum number of iterations.

To evaluate the performance of the methods, we consider multiple quantitative metrics, including computational time, iteration count, final objective value, and relative error. The relative error is defined by

$$\frac{\|x - x^{\text{true}}\|_2}{\|x^{\text{true}}\|_2},$$

where x is the recovered signal.

To investigate the influence of the extrapolation parameter, we compare several constant choices, $\beta = 0.1, 0.5, 0.9$, with the FISTA-type rule equipped with periodic restart. As reported in Table 1, all choices of the extrapolation parameter yield nearly identical final energy values and relative errors, indicating that the resulting solution quality is largely insensitive to the specific choice of β. Despite this similarity in accuracy, the convergence efficiency differs markedly. In particular, the FISTA-type rule attains the same level of accuracy with substantially fewer iterations and reduced computational time compared to the constant extrapolation strategies. While a larger constant parameter (e.g., $\beta = 0.9$) accelerates the initial decrease, it remains less efficient overall. These results suggest that, although the extrapolation parameter has limited impact on the final solution quality, it plays a decisive role in determining convergence efficiency. Motivated by this observation, we employ the FISTA-type extrapolation rule in all subsequent experiments.

Figure 1 further illustrates these trends. All choices approach a similar final accuracy, but the FISTA-type rule shows the fastest decrease and reaches the target accuracy in substantially fewer iterations. The zoomed-in view highlights the early-stage behavior, where larger constant parameters yield a rapid initial decrease but remain less efficient than the FISTA-type rule over the full iteration process.

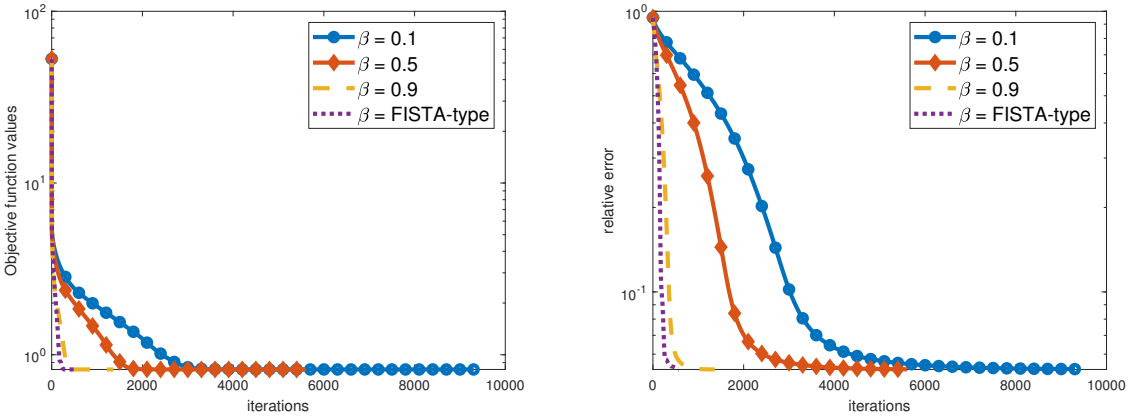
Having examined the effect of the extrapolation parameter, we now turn to the main comparison among the considered methods. In all subsequent experiments, GIRL1E employs the FISTA-type extrapolation rule described above. Table 2 reports the mean and standard deviation (over 10 independent runs) of computational time, iteration count, final objective value, and relative reconstruction error, providing a comprehensive comparison of efficiency and reconstruction accuracy. Although IRL1 requires the smallest number of outer iterations on average, its computational time is substantially higher than that of the other methods. This is due to its high per-iteration cost, since each outer iteration requires solving a linear system. In contrast, both LIRL1 and GIRL1E employ a linearization of the data-fidelity term q, thereby avoiding expensive subproblem solves. As a result, each iteration mainly involves matrix–vector multiplications, gradient evaluations, and simple element-wise operations, resulting in substantially lower per-iteration computational costs than IRL1.

Table 1. Comparison of different extrapolation parameters in compressive sensing.

β	Time (s)	Iterations	Energy	Rel. Error
0.1	10.033	9305	0.8207	0.0531
0.5	6.047	5603	0.8208	0.0530
0.9	1.506	1364	0.8207	0.0529
FISTA-type	**0.670**	**566**	0.8207	0.0529

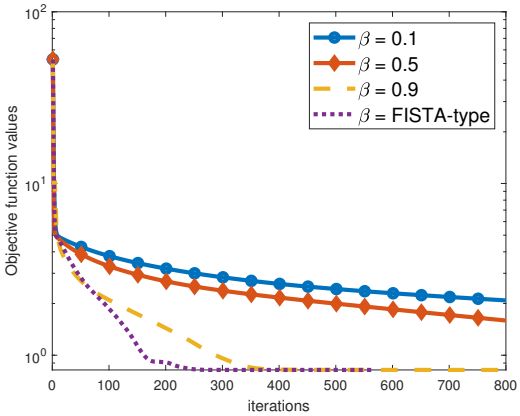
Table 2. Mean ± standard deviation (over 10 independent runs) of the computational time, iteration count, final objective value, and relative error.

Method	Time (s)	Iteration	Energy	Rel. Error
LIRL1	4.204 ± 0.679	3874 ± 614	$(78.21 \pm 5.22) \times 10^{-2}$	0.058 ± 0.012
IRL1	26.058 ± 9.007	62 ± 21	$(78.26 \pm 5.23) \times 10^{-2}$	0.058 ± 0.012
GIRL1E	0.613 ± 0.075	529 ± 66	$(78.21 \pm 5.22) \times 10^{-2}$	0.058 ± 0.012



(a) Objective value

(b) Relative error



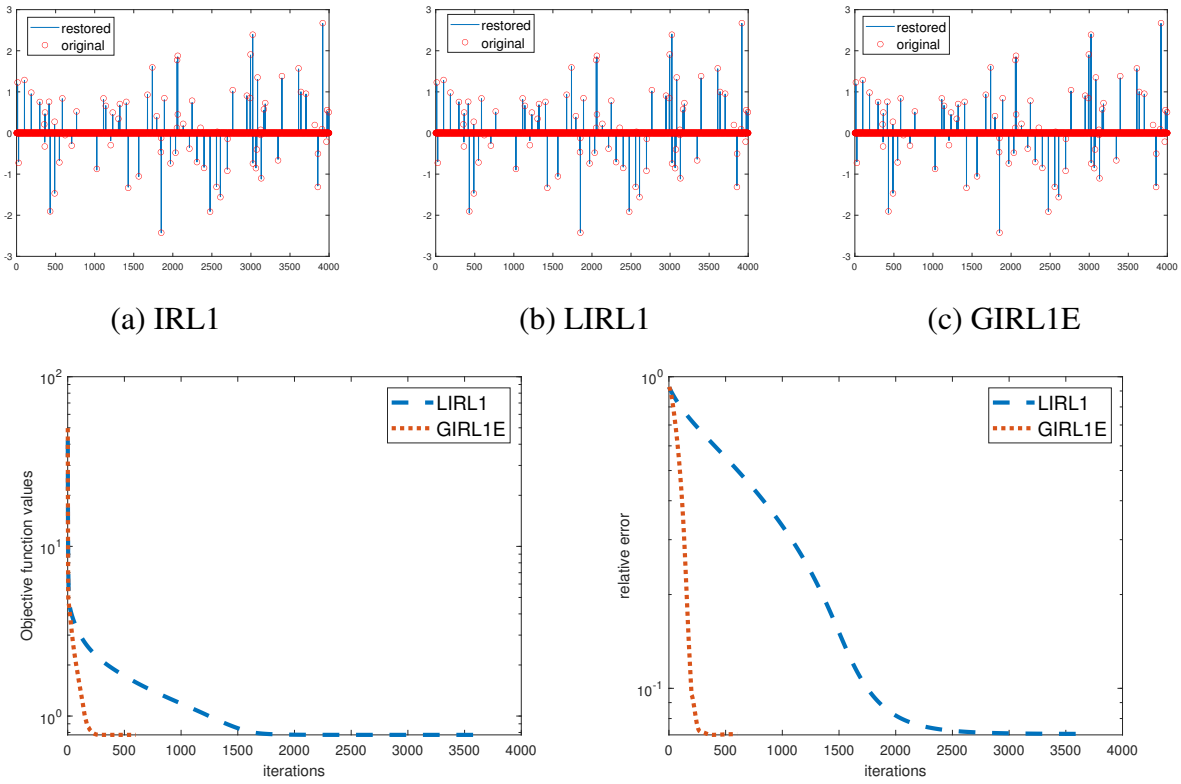
(c) Objective value (zoom)

Figure 1. Convergence behavior for different extrapolation parameters in terms of the objective value (a, c) and the relative error (b), all plotted on a logarithmic scale. All methods attain a similar level of accuracy, while the FISTA-type rule converges significantly faster. The zoomed-in view highlights the early-stage behavior.

The main difference between LIRL1 and GIRL1E lies in the use of an extrapolation step. While LIRL1 follows a standard proximal gradient-type update, GIRL1E incorporates extrapolation using previous iterates, which accelerates practical convergence. This trend is clearly reflected in Table 2: Although LIRL1 requires a large number of iterations, GIRL1E markedly reduces both the iteration count and the overall runtime. Since LIRL1 and GIRL1E involve similarly simple updates without solving large-scale subproblems, their per-iteration computational costs are comparable. Therefore, the runtime reduction from LIRL1 to GIRL1E mainly reflects the improved convergence behavior induced by extrapolation. Finally, while the iteration counts exhibit moderate variability across runs, the final objective values and relative reconstruction errors remain highly consistent, indicating stable and comparable reconstruction quality among all methods.

To provide a clearer visualization of these observations, Figure 2 shows the reconstructed sparse signals obtained by the three methods, together with the iteration progress of the objective value and the relative error, both plotted on a logarithmic scale. The visual results are nearly indistinguishable,

confirming comparable reconstruction quality among the methods. The objective curves show a clear decreasing trend, and GIRL1E reduces the objective more rapidly than LIRL1. This faster convergence is further supported by the relative error curves, which show that GIRL1E reaches a comparable level of reconstruction accuracy in considerably fewer iterations, particularly during the early stages. Since IRL1 operates through outer iterations that involve solving a linear system, its iteration count is not directly comparable to the lightweight first-order updates used in LIRL1 and GIRL1E. These observations are consistent with the quantitative results in Table 2, demonstrating that GIRL1E achieves efficient convergence while maintaining competitive reconstruction accuracy. This behavior also suggests an improved practical convergence rate, as evidenced by the faster decay of both the objective value and the relative error.



(a) IRL1 (b) LIRL1 (c) GIRL1E

(d) Objective value (e) Relative error

Figure 2. Top row: Reconstructed sparse signals obtained by IRL1, LIRL1, and the proposed method (GIRL1E). Bottom row: Convergence behavior in terms of the objective value (left) and the relative error (right), both plotted on a logarithmic scale. All methods achieve comparable reconstruction quality, while the proposed method exhibits significantly faster convergence.

To further examine scalability, we consider larger compressive-sensing problems with $n = 10^4$, 5×10^4, and 10^5. As the problem dimension increases, we keep the sampling ratio m/n and the sparsity level consistent with the previous setting. For these experiments, we employ a partial discrete cosine transform (DCT) measurement operator, which enables efficient transform-based matrix-vector

multiplications. Since IRL1 requires solving a linear system at each iteration, its computational cost becomes prohibitive in large-scale settings. Hence, we restrict the comparison to first-order methods.

As reported in Table 3, the computational time naturally increases as the problem dimension grows. Notably, the number of iterations required by the proposed method remains nearly constant across different problem dimensions. This suggests that the scalability of GIRL1E is mainly influenced by the per-iteration computational cost rather than the number of iterations. Nevertheless, GIRL1E consistently requires substantially fewer iterations and shorter runtime than LIRL1.

Table 3. Scalability comparison for large-scale compressive-sensing problems with a partial DCT measurement operator.

n	Method	Time (s)	Iteration	Energy	Rel. Error
10^4	LIRL1	7.484	6692	1.8579	0.0674
	GIRL1E	**0.677**	**601**	1.8579	0.0670
5×10^4	LIRL1	20.681	5416	9.8710	0.0591
	GIRL1E	**3.026**	**801**	9.8710	0.0591
10^5	LIRL1	57.818	6622	19.7427	0.0568
	GIRL1E	**5.474**	**601**	19.7427	0.0568

In particular, for the largest case $n = 10^5$, our method reduces runtime from 57.818 seconds to 5.474 seconds while achieving nearly identical final energy and reconstruction accuracy. These findings demonstrate that the computational advantage of the proposed method remains robust in larger-scale settings while preserving stable reconstruction quality.

4.2. Image deblurring

We next examine the proposed method on an image-deblurring problem as a representative imaging application. We describe the experimental setup used in all experiments. Five standard test images are used, as shown in Figure 3. These images exhibit diverse textures and structural features, providing a suitable benchmark for evaluating the performance of the algorithms.



(a) Bird (b) Cameraman (c) Clock (d) Fields (e) House

Figure 3. Test images used in the experiments. Most images are of size 256×256, while Bird and Fields have sizes 271×271 and 271×266, respectively.

The observed images are generated by applying a blur operator followed by additive Gaussian noise. We consider two types of blur: a Gaussian blur with kernel size 9 and standard deviation 2, and a motion blur with direction 60° and length 5. The noise level is fixed with standard deviation $\sigma = 5$. To provide a comprehensive evaluation, we report multiple quantitative metrics, including

runtime, iteration count, final energy, peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). While PSNR and SSIM assess restoration quality, our primary focus is on convergence behavior and computational efficiency rather than solely on reconstruction quality. Unless otherwise stated, the above settings are used throughout the following experiments.

Under these settings, a log-type regularized image-deblurring model with total variation regularization is considered. A standard nonconvex total variation (TV)-based formulation is commonly used for image deblurring problems. However, due to the structure of the regularization term, GIRL1E cannot be applied directly to this formulation. To overcome this difficulty, a reformulated model is adopted by introducing auxiliary variables together with a quadratic penalty term. This approach enables the application of the proposed method while preserving the essential characteristics of the original variational model. Accordingly, the following two formulations are considered, corresponding to anisotropic and isotropic TV regularizations. First, the anisotropic TV formulation is given by

$$\min_{u,d_x,d_y} \frac{\lambda}{2}\|Au-b\|_2^2+\frac{\mu}{2}\|\nabla_x u-d_x\|_2^2+\frac{\mu}{2}\|\nabla_y u-d_y\|_2^2+\frac{1}{\rho}\sum_i\Big\{\log(1+\rho|d_{x,i}|)+\log(1+\rho|d_{y,i}|)\Big\}, \quad (4.7)$$

where A denotes the blurring operator, b is the observed image, Ω denotes the pixel index set, and μ is a penalty parameter. The isotropic TV formulation is defined similarly as follows:

$$\min_{u,d_x,d_y} \frac{\lambda}{2}\|Au-b\|_2^2+\frac{\mu}{2}\|\nabla_x u-d_x\|_2^2+\frac{\mu}{2}\|\nabla_y u-d_y\|_2^2+\frac{1}{\rho}\sum_i\log(1+\rho\sqrt{d_{x,i}^2+d_{y,i}^2}). \quad (4.8)$$

These variants are included to illustrate the behavior of the algorithms under different regularization structures. The remaining components of the model are identical, and similar numerical behavior is observed for both formulations. The focus of the experiments is therefore on comparing algorithmic performance rather than the choice of TV variant. The model parameters are fixed as follows throughout the experiments: $\lambda = 1$, $\mu = 20$, and $\rho = 0.01$. These values are empirically selected and kept fixed across all methods to ensure a fair comparison and stable, consistent performance across different test images.

The proposed method is applied as described below, while the competing algorithms are implemented according to their respective formulations in the literature. To apply the proposed framework, the model components are identified according to the general formulation (1.1). Both models fit into this framework with the same choices of p and q:

$$p(u,d_x,d_y)=\frac{\lambda}{2}\|Au-b\|_2^2, \quad q(u,d_x,d_y)=\frac{\mu}{2}\|\nabla_x u-d_x\|_2^2+\frac{\mu}{2}\|\nabla_y u-d_y\|_2^2.$$

Here, the data-fidelity term is treated as the convex proximable component, while the quadratic penalty term is incorporated into the smooth co-coercive part. The difference between the anisotropic and isotropic cases lies only in the definitions of the mappings g and h, which reflect the corresponding structures of the TV regularization. For the anisotropic formulation, $g:\mathbb{R}^{m\times n}\times\mathbb{R}^{m\times n}\to\mathbb{R}_+^{m\times n}\times\mathbb{R}_+^{m\times n}$ is defined componentwise according to the directional gradients, and $h:\mathbb{R}_+^{m\times n}\times\mathbb{R}_+^{m\times n}\to\mathbb{R}$ corresponds to a separable log-type penalty.

$$h(y_x,y_y)=\frac{1}{\rho}\sum_{i\in\Omega}\Big\{\log(1+\rho y_{x,i})+\log(1+\rho y_{y,i})\Big\}, \quad g(d_x,d_y)=(|d_x|,|d_y|) \quad \text{(applied componentwise)},$$

where $\mathbb{R}_+^{m\times n}$ denotes the set of nonnegative real matrices. For the isotropic formulation, $g : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}_+^{m\times n}$ is defined through grouped gradient magnitudes, and $h : \mathbb{R}_+^{m\times n} \to \mathbb{R}$ is chosen accordingly to capture the coupled structure.

$$h(y) = \frac{1}{\rho}\sum_{i\in\Omega}\log(1+\rho y_i), \quad g(d_x, d_y) = \sqrt{d_x^2 + d_y^2}.$$

Under these definitions, the assumptions required by the proposed framework are satisfied in both cases. It follows from [26] that ∇q is co-coercive with respect to a positive definite matrix L, which is defined below. Consider the symmetric positive semidefinite matrix

$$B = \begin{bmatrix} \mu\nabla^\top\nabla & -\mu\nabla^\top \\ -\mu\nabla & \mu I \end{bmatrix}.$$

Here, $\nabla = (\nabla_x, \nabla_y)^\top$ denotes the discrete gradient operator, and $\nabla^\top$ denotes its adjoint. By the spectral theorem, there exist an orthogonal matrix U and a diagonal matrix D such that $B = U^\top DU$. Define $L := U^\top D'U$, where $D' := \max(D, t)$ for a small positive constant $t > 0$. Hence, both formulations fall within the scope of the proposed framework. For the anisotropic case, the updates follow the separable structure of the regularization term:

$$\begin{cases} \omega_1^k = \dfrac{1}{1+\rho|d_x^k|}, \quad \omega_2^k = \dfrac{1}{1+\rho|d_y^k|}, \\ \bar{u}^k = u^k + \beta_k(u^k - u^{k-1}), \\ \bar{d}_x^k = d_x^k + \beta_k(d_x^k - d_x^{k-1}), \\ \bar{d}_y^k = d_y^k + \beta_k(d_y^k - d_y^{k-1}), \\ (u^{k+1}, d_x^{k+1}, d_y^{k+1}) = \arg\min\limits_{u,d_x,d_y} \langle\omega_1^k, |d_x|\rangle + \langle\omega_2^k, |d_y|\rangle + \dfrac{\lambda}{2}\|Au-b\|_2^2 \\ \qquad\qquad + \langle\mu(\bar{d}_x^k - \nabla_x\bar{u}^k), d_x\rangle + \langle\mu(\bar{d}_y^k - \nabla_y\bar{u}^k), d_y\rangle \\ \qquad\qquad + \langle\mu\nabla^\top(\nabla\bar{u}^k - \bar{\mathbf{d}}^k), u\rangle + \dfrac{1}{2}\|(u, d_x, d_y)^\top - (\bar{u}^k, \bar{d}_x^k, \bar{d}_y^k)^\top\|_{\bar{L}}^2, \end{cases} \tag{4.9}$$

where

$$\bar{L} = 2\begin{bmatrix} \mu\nabla^\top\nabla + \delta I & \mathbf{0} \\ \mathbf{0} & \mu I \end{bmatrix},$$

with a small positive value of $\delta > t$ and $\bar{\mathbf{d}}^k = (\bar{d}_x^k, \bar{d}_y^k)$. Similarly, for the isotropic case, the updates are adapted to account for the coupled form of the regularization:

$$\begin{cases} \omega^k = \dfrac{1}{1+\rho\sqrt{(d_x^k)^2 + (d_y^k)^2}}, \\ \bar{u}^k = u^k + \beta_k(u^k - u^{k-1}), \\ \bar{d}_x^k = d_x^k + \beta_k(d_x^k - d_x^{k-1}), \\ \bar{d}_y^k = d_y^k + \beta_k(d_y^k - d_y^{k-1}), \\ (u^{k+1}, d_x^{k+1}, d_y^{k+1}) = \arg\min\limits_{u,d_x,d_y} \langle\omega^k, \sqrt{d_x^2 + d_y^2}\rangle + \dfrac{\lambda}{2}\|Au-b\|_2^2 \\ \qquad\qquad + \langle\mu(\bar{d}_x^k - \nabla_x\bar{u}^k), d_x\rangle + \langle\mu(\bar{d}_y^k - \nabla_y\bar{u}^k), d_y\rangle \\ \qquad\qquad + \langle\mu\nabla^\top(\nabla\bar{u}^k - \bar{\mathbf{d}}^k), u\rangle + \dfrac{1}{2}\|(u, d_x, d_y)^\top - (\bar{u}^k, \bar{d}_x^k, \bar{d}_y^k)^\top\|_{\bar{L}}^2, \end{cases} \tag{4.10}$$

$\bar{L}$ is symmetric and positive definite under periodic boundary conditions, and $\bar{L} - L \succ 0$, which is consistent with the theoretical assumptions required for convergence. Here, $\delta = 10^{-5}$ and $t = \varepsilon$ (machine precision) are chosen as small positive constants to ensure numerical stability and the positive definiteness of $\bar{L}$.

The resulting convex subproblem is separable with respect to (d_x, d_y) and u, and can be solved by minimizing over each block independently. For the anisotropic formulation, the updates of d_x and d_y are given componentwise by the soft-thresholding operator:

$$d_x^{k+1} = \mathcal{S}_{\omega_1^k/(2\mu)}\left(\frac{-\mu(\bar{d}_x^k - \nabla_x \bar{u}^k) + 2\mu \bar{d}_x^k}{2\mu}\right), \qquad d_y^{k+1} = \mathcal{S}_{\omega_2^k/(2\mu)}\left(\frac{-\mu(\bar{d}_y^k - \nabla_y \bar{u}^k) + 2\mu \bar{d}_y^k}{2\mu}\right),$$

where $\mathcal{S}_\omega(t) = \text{sign}(t)\max(|t| - \omega, 0)$. For the isotropic formulation, the updates are obtained via vector soft-thresholding:

$$(d_x^{k+1}, d_y^{k+1}) = \max\left(1 - \frac{\omega^k}{2\mu\sqrt{(z_x^k)^2 + (z_y^k)^2}}, 0\right)(z_x^k, z_y^k),$$

where

$$z_x^k = \frac{-\mu(\bar{d}_x^k - \nabla_x \bar{u}^k) + 2\mu \bar{d}_x^k}{2\mu}, \quad z_y^k = \frac{-\mu(\bar{d}_y^k - \nabla_y \bar{u}^k) + 2\mu \bar{d}_y^k}{2\mu}.$$

The update of u is obtained by solving the corresponding quadratic subproblem:

$$(2\mu\nabla^\top\nabla + 2\delta I + \lambda A^\top A)u = \mu\nabla^\top(\bar{\mathbf{d}}^k + \nabla\bar{u}^k) + \lambda A^\top b + 2\delta\bar{u}^k.$$

Since $2\mu\nabla^\top\nabla + 2\delta I + \lambda A^\top A$ can be diagonalized by the two-dimensional fast Fourier transform (FFT) under periodic boundary conditions, the solution admits the following closed-form expression:

$$u^{k+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mu\nabla^\top(\bar{\mathbf{d}}^k + \nabla\bar{u}^k) + \lambda A^\top b + 2\delta\bar{u}^k)}{\mathcal{F}(2\mu\nabla^\top\nabla + 2\delta I + \lambda A^\top A)}\right).$$

The parameter β_k is selected in the same manner as in Section 4.1. The stopping criterion follows the same principle as in Section 4.1, given by

$$\frac{(9\mu + 2\delta)\|u^{k+1} - \bar{u}^k\|_2 + (9\mu + \rho)(\|d_x^{k+1} - \bar{d}_x^k\|_2 + \|d_y^{k+1} - \bar{d}_y^k\|_2)}{\max(1, \|u^{k+1}\|_2 + \|d_x^{k+1}\|_2 + \|d_y^{k+1}\|_2)} < \text{tol} \quad \text{or} \quad k > \text{maxiter}.$$

The constants in this criterion arise from the corresponding relative-error bound for this deblurring formulation. The stopping parameters are chosen as in Section 4.1.

We compare the proposed method with several related algorithms, namely, LIRL1 [31], GPL-IRL1 [26], LIRL1E1 [21], LIRL1E2 [21], and LIRL1E3 [21]. The competing algorithms are implemented following the numerical protocol of [26], with appropriate modifications to accommodate the different regularization terms. For the competing algorithms, stopping criteria of comparable accuracy are employed, defined according to their respective formulations. The detailed expressions are omitted for brevity.

For the isotropic formulation, GIRL1E, LIRL1, and GPL-IRL1 can be applied directly without modification. In contrast, the extrapolated variants LIRL1E1, LIRL1E2, and LIRL1E3 are derived

under structural assumptions specific to the original formulation, which do not directly extend to the isotropic setting. LIRL1E1 is a special case of the proposed framework when a standard ℓ_2-based co-coercive structure is employed. However, the proposed formulation is designed to preserve a more general co-coercive structure, and thus a different implementation is adopted. For LIRL1E2 and LIRL1E3, applicability to the isotropic formulation has not been theoretically established, and their direct use would require further justification. Therefore, these extrapolated variants are excluded from the isotropic comparison.

We now briefly comment on the role of the model parameters. Their influence is broadly consistent with standard variational formulations. While the convergence behavior and restoration quality depend on the parameters λ and μ, GIRL1E avoids linearization of the λ-weighted data fidelity term, and is therefore less sensitive to variations in the associated Lipschitz-type constant than fully linearized schemes. The parameter μ affects the conditioning of the subproblems, but no abnormal sensitivity is observed in the experiments.

Table 4 provides a comprehensive comparison of the quantitative performance of LIRL1, GPL-IRL1, LIRL1E1, LIRL1E2, LIRL1E3, and GIRL1E in terms of computational time, iteration count, final energy, PSNR, and SSIM on five test images under Gaussian and motion blur.

From the perspective of computational efficiency, the non-inertial method LIRL1 requires substantially more iterations and longer runtime, demonstrating the benefit of incorporating extrapolation strategies. All inertial variants (LIRL1E1, LIRL1E2, and LIRL1E3) exhibit clear improvements over LIRL1, confirming the effectiveness of inertial acceleration in reducing iteration complexity. However, the comparison among inertial variants indicates that the choice of extrapolation strategy plays a critical role. LIRL1E3 often attains fewer iterations than other extrapolated variants, suggesting improved convergence efficiency in terms of iteration count. Nevertheless, this advantage does not always translate into the shortest runtime, due to its relatively higher per-iteration computational cost associated with the two-step update structure. LIRL1E2 generally exhibits less competitive performance in terms of runtime and iteration efficiency, indicating that its cyclic extrapolation strategy is comparatively less effective for the considered problems.

A direct comparison between LIRL1E1 and GIRL1E provides further insight. Both methods employ a similar inertial extrapolation mechanism, while LIRL1E1 does not incorporate the co-coercive structure nor the generalized nonlinear modeling. GIRL1E consistently achieves lower objective values and comparable or higher PSNR and SSIM across most test images. These results indicate that the observed performance gain is not solely due to inertial acceleration, but primarily arises from the proposed generalized reweighted framework and the incorporation of the co-coercive condition. A comparison with GPL-IRL1 further supports this observation. Although GPL-IRL1 incorporates the co-coercive structure, it does not employ inertial extrapolation and therefore requires more iterations. GIRL1E combines both mechanisms, leading to improved convergence behavior and reconstruction quality.

Table 4. Quantitative results of (4.7) on five test images under two different types of blur.

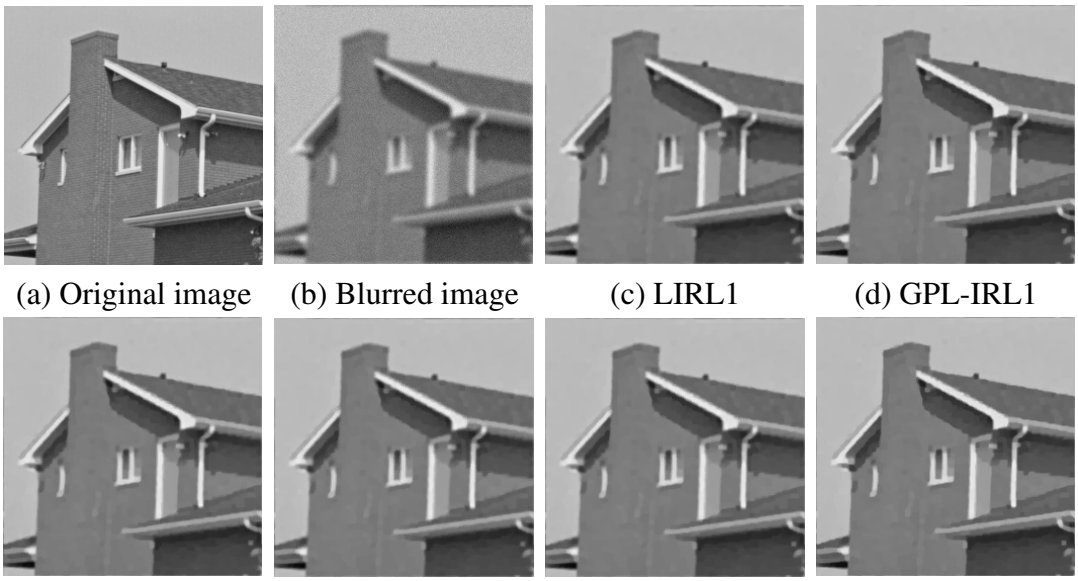
		Gaussian blur					Motion blur				
Image	Method	Time (s)	Iter	Energy	PSNR (dB)	SSIM	Time (s)	Iter	Energy	PSNR (dB)	SSIM
Bird	LIRL1	20.11	2471	1.26e6	30.94	0.9139	20.52	2629	1.20e6	33.77	0.9288
	GPL-IRL1	6.98	827	1.24e6	31.58	0.9149	6.79	847	1.18e6	34.23	0.9294
	LIRL1E1	3.95	403	1.24e6	31.47	**0.9158**	3.93	401	1.18e6	34.17	0.9295
	LIRL1E2	3.78	400	1.26e6	31.05	0.9146	3.63	400	1.20e6	33.82	0.9296
	LIRL1E3	3.33	298	1.25e6	31.26	0.9154	3.36	302	1.19e6	34.03	**0.9307**
	GIRL1E	**2.31**	**248**	**1.23e6**	**31.72**	0.9138	**2.31**	**232**	**1.17e6**	**34.28**	0.9272
Cameraman	LIRL1	13.69	3305	1.18e6	23.97	0.7615	16.26	3962	1.27e6	27.67	0.8460
	GPL-IRL1	5.52	1326	1.15e6	24.34	0.7723	6.44	1468	1.23e6	28.23	0.8509
	LIRL1E1	**2.07**	414	1.16e6	24.14	0.7672	2.24	413	1.25e6	27.91	0.8490
	LIRL1E2	3.25	500	1.18e6	23.99	0.7623	3.51	600	1.26e6	27.72	0.8466
	LIRL1E3	2.84	379	1.17e6	24.12	0.7662	2.97	397	1.25e6	27.87	0.8484
	GIRL1E	2.39	**408**	**1.13e6**	**24.45**	**0.7750**	**2.16**	**401**	**1.21e6**	**28.44**	**0.8519**
Clock	LIRL1	9.24	2227	1.16e6	26.30	0.8625	9.67	2383	1.19e6	29.36	0.9076
	GPL-IRL1	3.07	711	1.14e6	26.60	0.8693	3.46	809	1.16e6	29.89	0.9146
	LIRL1E1	2.10	401	1.13e6	26.57	0.8693	2.06	385	1.16e6	29.86	0.9148
	LIRL1E2	2.46	400	1.15e6	26.39	0.8651	2.37	400	1.18e6	29.48	0.9104
	LIRL1E3	1.95	280	1.15e6	26.46	0.8670	1.95	286	1.17e6	29.65	0.9133
	GIRL1E	**1.63**	**236**	**1.12e6**	**26.71**	**0.8711**	**1.36**	**229**	**1.14e6**	**30.16**	**0.9159**
Fields	LIRL1	15.59	2525	1.20e6	29.40	**0.7605**	18.37	2869	1.19e6	31.19	0.8239
	GPL-IRL1	5.28	784	1.19e6	**29.52**	0.7593	5.77	924	1.17e6	31.36	0.8243
	LIRL1E1	3.23	405	1.19e6	29.52	**0.7605**	3.12	404	1.17e6	31.35	0.8247
	LIRL1E2	3.35	400	1.20e6	29.43	0.7607	4.15	500	1.19e6	31.25	0.8245
	LIRL1E3	2.92	307	1.19e6	29.48	**0.7605**	3.23	325	1.18e6	31.30	**0.8248**
	GIRL1E	**1.84**	**241**	**1.18e6**	**29.52**	0.7575	**1.97**	**239**	**1.17e6**	**31.37**	0.8229
House	LIRL1	12.19	2815	1.12e6	29.46	0.8254	13.70	3028	1.10e6	31.57	0.8564
	GPL-IRL1	4.01	945	1.10e6	30.01	0.8296	4.60	1045	1.07e6	32.11	0.8595
	LIRL1E1	2.18	407	1.10e6	29.87	0.8295	2.19	404	1.08e6	31.96	0.8593
	LIRL1E2	3.64	500	1.11e6	29.61	0.8269	3.74	500	1.09e6	31.69	0.8576
	LIRL1E3	2.46	335	1.11e6	29.74	0.8282	2.41	331	1.09e6	31.82	0.8588
	GIRL1E	**1.51**	**262**	**1.09e6**	**30.14**	**0.8304**	**1.41**	**250**	**1.07e6**	**32.21**	**0.8597**

Although an explicit FLOP count is not provided, the per-iteration computational costs differ across the compared methods due to their distinct update structures. For GIRL1E and GPL-IRL1, the dominant cost arises from solving the quadratic u-subproblem, which is efficiently handled via FFT under periodic boundary conditions. In contrast, LIRL1 and the extrapolated LIRL1E variants mainly rely on gradient, divergence, and shrinkage-type operations, resulting in relatively simpler per-iteration updates. Among these variants, LIRL1E3 incurs the highest per-iteration overhead because it solves

the corresponding subproblem twice within each iteration, while LIRL1E2 also involves additional auxiliary computations associated with its extrapolation mechanism. Consequently, a smaller iteration count does not always translate directly into lower runtime. These observations indicate that overall computational performance depends not only on iteration complexity but also on method-specific per-iteration cost. Despite its FFT-based subproblem structure, GIRL1E substantially reduces the total number of iterations, resulting in highly competitive overall runtime while maintaining lower objective values and competitive or improved restoration quality.

In terms of reconstruction quality, all methods yield visually similar results, as reflected by the comparable PSNR and SSIM values. Nevertheless, GIRL1E consistently achieves lower final energy values and generally competitive or improved PSNR and SSIM, indicating more stable convergence to improved solutions. From a convergence standpoint, although a rigorous rate analysis is beyond the scope of this work, the consistent reduction in iteration counts compared to non-inertial methods, together with the observed decay of the objective function, suggests empirically accelerated convergence.

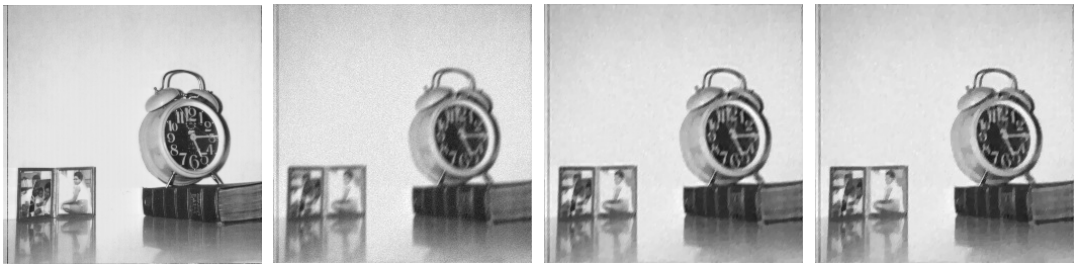
Figures 4 and 5 present representative reconstruction results for the anisotropic model (4.7). Although the visual differences are subtle, GIRL1E achieves comparable or improved reconstruction quality with fewer iterations and competitive runtime. Overall, the results show that GIRL1E achieves an effective balance between computational performance and reconstruction accuracy, with its performance gains largely attributable to the generalized modeling framework and co-coercive structure.



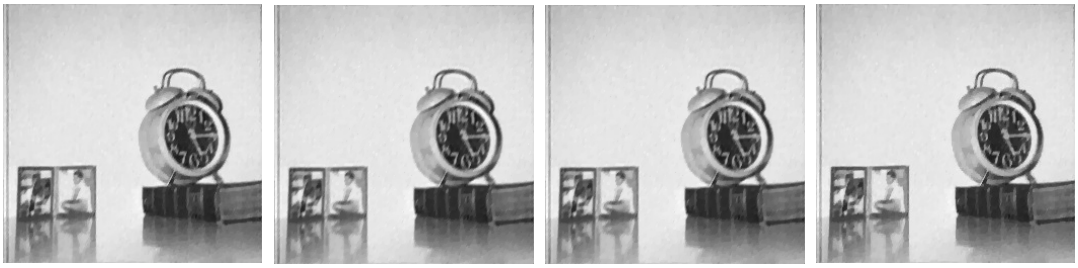
(a) Original image (b) Blurred image (c) LIRL1 (d) GPL-IRL1

(e) LIRL1E1 (f) LIRL1E2 (g) LIRL1E3 (h) GIRL1E

Figure 4. Visual comparison of deblurring results for problem (4.7) under Gaussian blur. The images correspond to the original image, the blurred observation, and the results obtained by LIRL1, GPL-IRL1, LIRL1E1, LIRL1E2, LIRL1E3, and the proposed GIRL1E.



(a) Original image (b) Blurred image (c) LIRL1 (d) GPL-IRL1



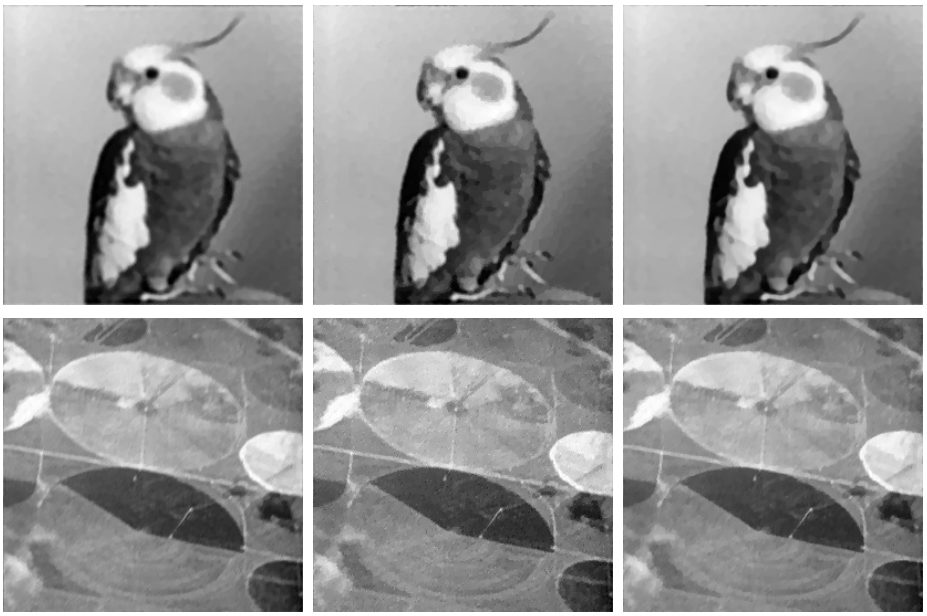
(e) LIRL1E1 (f) LIRL1E2 (g) LIRL1E3 (h) GIRL1E

Figure 5. Visual comparison of deblurring results for problem (4.7) under motion blur. The images correspond to the original image, the blurred observation, and the results obtained by LIRL1, GPL-IRL1, LIRL1E1, LIRL1E2, LIRL1E3, and the proposed GIRL1E.

Table 5 reports the results for the isotropic formulation (4.8), providing further evidence of the applicability of GIRL1E under the same experimental setting with a different reconstruction model. While the PSNR and SSIM values vary across test images, all methods achieve comparable overall reconstruction quality. Notably, GPL-IRL1 exhibits different behavior depending on the blur type: under Gaussian blur, it attains relatively low energy values and competitive PSNR and SSIM, whereas under motion blur, it yields higher energy levels and relatively lower PSNR and SSIM, as reported in Table 5. In contrast, GIRL1E generally achieves low energy values and competitive or higher PSNR and SSIM while requiring fewer iterations and less runtime, indicating reliable numerical performance across different settings. Figure 6 presents representative reconstructions for the isotropic case (4.8). Consistent with the quantitative results, visual differences among the methods can be observed, particularly under motion blur, where GIRL1E produces cleaner and more reliable reconstructions. Overall, these observations show that GIRL1E maintains stable computational efficiency and restoration quality across different blur conditions and TV-regularization structures.

Table 5. Quantitative results of (4.8) on five test images under two different types of blur.

		Gaussian blur					Motion blur				
Image	Method	Time (s)	Iter	Energy	PSNR (dB)	SSIM	Time (s)	Iter	Energy	PSNR (dB)	SSIM
Bird	LIRL1	20.25	2546	1.21e6	30.98	0.9151	21.40	2615	1.14e6	33.56	0.9179
	GPL-IRL1	9.37	1189	1.21e6	31.26	0.9055	11.31	1416	1.16e6	31.83	0.8292
	GIRL1E	**2.08**	**229**	**1.19e6**	**31.95**	**0.9198**	**2.19**	**221**	**1.11e6**	**34.17**	**0.9213**
Cameraman	LIRL1	14.42	3149	1.13e6	23.96	0.7616	16.34	3674	1.19e6	27.62	0.8416
	GPL-IRL1	7.11	1596	1.13e6	24.39	0.7723	8.96	2127	1.23e6	27.66	0.7713
	GIRL1E	**1.53**	**254**	**1.10e6**	**24.39**	**0.7752**	**1.45**	**244**	**1.16e6**	**28.31**	**0.8501**
Clock	LIRL1	10.05	2293	1.12e6	26.32	0.8613	10.35	2354	1.12e6	29.28	0.8950
	GPL-IRL1	4.55	1018	1.12e6	26.74	0.8649	5.26	1238	1.14e6	29.42	0.8362
	GIRL1E	**1.33**	**223**	**1.09e6**	**26.77**	**0.8735**	**1.33**	**216**	**1.08e6**	**30.16**	**0.9099**
Fields	LIRL1	16.71	2607	1.16e6	29.50	0.7668	18.66	2752	1.12e6	31.19	0.8238
	GPL-IRL1	8.88	1394	1.16e6	29.27	0.7494	11.04	1668	1.16e6	29.92	0.7538
	GIRL1E	**1.94**	**228**	**1.14e6**	**29.71**	**0.7667**	**1.86**	**224**	**1.10e6**	**31.46**	**0.8258**
House	LIRL1	12.44	2813	1.09e6	29.41	0.8254	13.49	2956	1.04e6	31.39	0.8491
	GPL-IRL1	6.45	1388	1.08e6	29.78	0.8248	7.10	1653	1.07e6	30.31	0.7746
	GIRL1E	**1.45**	**240**	**1.06e6**	**30.17**	**0.8329**	**1.28**	**230**	**1.01e6**	**32.05**	**0.8553**



(a) LIRL1 (b) GPL-IRL1 (c) GIRL1E

Figure 6. Visual comparison of reconstruction results for problem (4.8) under two blur settings. Top row: Gaussian blur; bottom row: motion blur.

To further investigate scalability in image deblurring, we consider the same Kodak image at three progressively higher resolutions while preserving the aspect ratio. We use a representative image (image 03) from the Kodak dataset [32]. For this experiment, we employ the anisotropic TV model. We compare GIRL1E with LIRL1E1, which serves as a representative extrapolated first-order baseline with a comparable linearized structure. In particular, LIRL1E1 is selected because it demonstrated competitive overall performance in the main experiments while maintaining a relatively simple and robust extrapolation structure, making it well-suited for scalability analysis. Other variants, including LIRL1, GPL-IRL1, LIRL1E2, and LIRL1E3, are not considered here in order to maintain clarity of comparison and to focus on representative first-order scalability analysis.

As reported in Table 6, the computational time naturally increases with image resolution. Nevertheless, GIRL1E consistently requires fewer iterations and achieves shorter runtime than LIRL1E1 across all tested resolutions. The number of iterations remains relatively stable across resolutions, indicating favorable scalability behavior. For example, at the highest tested resolution 683×1024, GIRL1E reduces the runtime from 31.91 seconds to 18.34 seconds, while decreasing the number of iterations from 405 to 242. In terms of restoration accuracy, GIRL1E achieves slightly higher PSNR values across all resolutions, while the SSIM values of the two methods remain comparable. The final energy values obtained by GIRL1E are also consistently lower than those of LIRL1E1. These results demonstrate that the proposed acceleration strategy remains effective for high-resolution image-deblurring problems, providing improved practical efficiency while preserving reconstruction quality.

Table 6. Scalability comparison for image deblurring at multiple resolutions.

Resolution	Method	Time (s)	Iteration	PSNR	SSIM	Energy
342×512	LIRL1E1	8.12	409	31.43	0.8690	2.58e6
	GIRL1E	**5.24**	**255**	**31.53**	**0.8693**	**2.57e6**
512×768	LIRL1E1	16.39	406	31.18	**0.8564**	5.65e6
	GIRL1E	**10.36**	**246**	**31.24**	**0.8564**	**5.63e6**
683×1024	LIRL1E1	31.91	405	33.02	**0.8931**	9.82e6
	GIRL1E	**18.34**	**242**	**33.04**	0.8923	**9.78e6**

5. Conclusions

In this paper, we proposed a generalized iteratively reweighted ℓ_1 algorithm with inertial extrapolation (GIRL1E) for a class of nonconvex and nonsmooth optimization problems. The proposed method integrates inertial extrapolation into generalized IRL1 approaches under co-coercive smoothness assumptions, thereby extending applicability to a broader range of structured optimization problems. Under these assumptions, we established global convergence of the generated sequence to a critical point through Lyapunov analysis. Numerical experiments in compressive sensing and image deblurring further demonstrated the practical effectiveness of GIRL1E, showing consistently improved performance over existing IRL1 methods. Future work includes inexact, stochastic, and adaptive extensions of the proposed method, as well as broader applications to structured nonconvex optimization problems.

Use of Generative-AI tools declaration

The author declares that ChatGPT was used during the preparation of this article. The tool was used for language editing and improving readability throughout the manuscript. The author reviewed and edited all AI-assisted outputs and takes full responsibility for the content of the article.

Acknowledgments

The author is grateful to the anonymous reviewers for their careful reading and valuable comments, which helped improve the quality and clarity of this manuscript. Myeongmin Kang was supported by Chungnam National University.

Conflict of interest

The author declares that there is no conflict of interest.

References

1. L. van den Dries, *Tame Topology and o-Minimal Structures*, vol. 248 of London Mathematical Society Lecture Note Series, Cambridge Univ. Press, 1998. https://doi.org/10.1017/CBO9780511525919
2. A. Seidenberg, A new decision method for elementary algebra, *Ann. of Math. (2)*, **60** (1954), 365–374. https://doi.org/10.2307/1969640
3. L. van den Dries, C. Miller, Geometric categories and o-minimal structures, *Duke Math. J.*, **84** (1996), 497–540. https://doi.org/10.1215/S0012-7094-96-08416-1
4. E. J. Candès, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, **14** (2008), 877–905. https://doi.org/10.1007/s00041-008-9045-x
5. I. Daubechies, R. DeVore, M. Fornasier, C. S. Güntürk, Iteratively reweighted least squares minimization for sparse recovery, *Comm. Pure Appl. Math.*, **63** (2010), 1–38. https://doi.org/10.1002/cpa.20303
6. D. R. Hunter, R. Li, Variable selection using MM algorithms, *Ann. Statist.*, **33** (2005), 1617–1642. https://doi.org/10.1214/009053605000000200
7. T. Sun, H. Jiang, L. Cheng, W. Zhu, Iteratively linearized reweighted alternating direction method of multipliers for a class of nonconvex problems, *IEEE Trans. Signal Process.*, **66** (2018), 5380–5391. https://doi.org/10.1109/TSP.2018.2868269
8. S. Ghadimi, G. Lan, Accelerated gradient methods for nonconvex nonlinear and stochastic programming, *Math. Program.*, **156** (2016), 59–99. https://doi.org/10.1007/s10107-015-0871-8
9. T. Pham Dinh, H. A. Le Thi, Convex analysis approach to DC programming: Theory, algorithms and applications, *Acta Math. Vietnam.*, **22** (1997), 289–355.
10. 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. https://doi.org/10.1007/s10107-011-0484-9

11. J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.*, **146** (2014), 459–494. https://doi.org/10.1007/s10107-013-0701-9
12. P. Ochs, Y. Chen, T. Brox, T. Pock, iPiano: Inertial proximal algorithm for nonconvex optimization, *SIAM J. Imaging Sci.*, **7** (2014), 1388–1419. https://doi.org/10.1137/130942954
13. Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, *Sov. Math. Dokl.*, **27** (1983), 372–376.
14. Y. Nesterov, *Introductory lectures on convex optimization: a basic course*, Springer, Boston, 2004. https://doi.org/10.1007/978-1-4419-8853-9
15. F. Alvarez, On the minimizing property of a second order dissipative system in Hilbert spaces, *SIAM J. Control Optim.*, **38** (2000), 1102–1119. https://doi.org/10.1137/S0363012998335802
16. P. Frankel, G. Garrigos, J. Peypouquet, Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates, *J. Optim. Theory Appl.*, **165** (2015), 874–900. https://doi.org/10.1007/s10957-014-0642-3
17. H. Attouch, A. Cabot, Convergence of a relaxed inertial proximal algorithm for maximally monotone operators, *Math. Program.*, **184** (2020), 243–287. https://doi.org/10.1007/s10107-019-01412-0
18. H. Attouch, A. Cabot, Convergence of a relaxed inertial forward–backward algorithm for structured monotone inclusions, *Appl. Math. Optim.*, **80** (2019), 547–598. https://doi.org/10.1007/s00245-019-09584-z
19. P. Ochs, A. Dosovitskiy, T. Brox, T. Pock, On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision, *SIAM J. Imaging Sci.*, **8** (2015), 331–372. https://doi.org/10.1137/140971518
20. Z. Lu, Iterative reweighted minimization methods for ℓ_p-regularized unconstrained nonlinear programming, *Math. Program.*, **147** (2014), 277–307. https://doi.org/10.1007/s10107-013-0722-4
21. P. Yu, T. K. Pong, Iteratively reweighted ℓ_1 algorithms with extrapolation, *Comput. Optim. Appl.*, **73** (2019), 353–386. https://doi.org/10.1007/s10589-019-00081-1
22. H. H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011. https://doi.org/10.1007/978-1-4419-9467-7
23. R. T. Rockafellar, R. J. B. Wets, *Variational Analysis*, Springer, 1998. https://doi.org/10.1007/978-3-642-02431-3
24. J. Bolte, A. Daniilidis, A. Lewis, M. Shiota, Clarke subgradients of stratifiable functions, *SIAM J. Optim.*, **18** (2007), 556–572. https://doi.org/10.1137/060670080
25. D. A. Lorenz, T. Pock, An inertial forward–backward algorithm for monotone inclusions, *J. Math. Imaging Vis.*, **51** (2015), 311–325. https://doi.org/10.1007/s10851-014-0523-2
26. M. Kang, Generalized proximal iteratively reweighted ℓ_1 algorithm with co-coerciveness for nonsmooth and nonconvex minimization problem, *J. Chungcheong Math. Soc.*, **37** (2024), 41–55. https://doi.org/10.14403/jcms.2024.37.1.41

27. J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.*, **96** (2001), 1348–1360. https://doi.org/10.1198/016214501753382273

28. C. H. Zhang, Nearly unbiased variable selection under minimax concave penalty, *Ann. Statist.*, **38** (2010), 894–942. https://doi.org/10.1214/09-AOS729

29. A. J. Wilkie, A theorem of the complement and some new o-minimal structures, *Selecta Math. (N.S.)*, **5** (1999), 397–421. https://doi.org/10.1007/s000290050052

30. A. J. 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. https://doi.org/10.1090/S0894-0347-96-00216-0

31. J. Yeo, M. Kang, Proximal linearized iteratively reweighted algorithms for nonconvex and nonsmooth optimization problem, *Axioms*, **11** (2022), 201. https://doi.org/10.3390/axioms11050201

32. Kodak lossless true color image suite. Available from: `https://r0k.us/graphics/kodak/`, Accessed: 2026.

Appendix: Proof of Theorem 3.1

Proof. To establish the convergence of $\{x^k\}$, it suffices to prove the finite length property

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

From Lemma 3.1(i), the sequence $\{\bar{F}(x^{k+1}, x^k)\}$ is nonincreasing and convergent. Define $\bar{F}^* := \lim_{k\to\infty} \bar{F}(x^{k+1}, x^k)$. We first consider the case where there exists $k' \in \mathbb{N}$ such that $\bar{F}(x^{k'+1}, x^{k'}) = \bar{F}^*$. By monotonicity, $\bar{F}(x^{k+1}, x^k) = \bar{F}^*$ for all $k \geq k'$. Hence, by the sufficient descent property (3.3), $\|x^k - x^{k-1}\|_{\bar{L}} = 0$ for all $k \geq k' + 1$. Since $\bar{L} \succ 0$, this implies $x^k = x^{k-1}$ for all $k \geq k' + 1$. Thus, $x^k = x^{k'}$ for all $k \geq k'$, and the sequence converges in finitely many steps. Consequently,

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

It remains to consider the nontrivial case where $\bar{F}(x^{k+1}, x^k) > \bar{F}^*$ for all $k \geq 0$. By Lemma 3.3(iii)', the function $\bar{F}$ is constant on the compact set $\widetilde{\mathcal{A}}$. Since F is assumed to be a KL function, the Lyapunov function $\bar{F}$ also satisfies the KL property by standard stability results for the KL property under the addition of a smooth quadratic term. Applying the uniformized KL property [10], there exist constants $\epsilon, \eta > 0$ and a desingularizing function φ such that

$$\varphi'(\bar{F}(x, y) - \bar{F}^*) \operatorname{dist}(0, \partial \bar{F}(x, y)) \geq 1,$$

whenever $\operatorname{dist}((x, y), \widetilde{\mathcal{A}}) < \epsilon$ and $\bar{F}^* < \bar{F}(x, y) < \bar{F}^* + \eta$. By Lemma 3.3(i)', there exists $k_0 \in \mathbb{N}$ such that $\operatorname{dist}((x^k, x^{k-1}), \widetilde{\mathcal{A}}) < \epsilon$ for all $k \geq k_0$. Moreover, by the definition of $\bar{F}^*$ and assumption that $\bar{F}(x^k, x^{k-1}) > \bar{F}^*$ for all k, there exists $k_1 \in \mathbb{N}$ such that $\bar{F}^* < \bar{F}(x^k, x^{k-1}) < \bar{F}^* + \eta$ for all $k \geq k_1$. Let $N_0 = \max\{k_0, k_1, 1\}$. Then, for all $k \geq N_0$, we have

$$\varphi'(\bar{F}(x^k, x^{k-1}) - \bar{F}^*) \operatorname{dist}(0, \partial \bar{F}(x^k, x^{k-1})) \geq 1.$$

Since φ is concave, we have

$$\begin{aligned}&\varphi(\bar{F}(x^k, x^{k-1}) - \bar{F}^*) - \varphi(\bar{F}(x^{k+1}, x^k) - \bar{F}^*)\\ &\quad \geq \varphi'(\bar{F}(x^k, x^{k-1}) - \bar{F}^*)(\bar{F}(x^k, x^{k-1}) - \bar{F}(x^{k+1}, x^k)).\end{aligned} \tag{A.1}$$

Since $\bar{L} \succ 0$, we have

$$\|x\|_{\bar{L}}^2 \geq \lambda_{\min}(\bar{L})\|x\|_2^2 \qquad \forall x \in \mathbb{R}^n.$$

Therefore, by (3.3),

$$\bar{F}(x^k, x^{k-1}) - \bar{F}(x^{k+1}, x^k) \geq \widetilde{C}\|x^k - x^{k-1}\|_2^2, \tag{A.2}$$

where $\widetilde{C} := C\lambda_{\min}(\bar{L}) > 0$. Define

$$\Delta_k := \varphi(\bar{F}(x^k, x^{k-1}) - \bar{F}^*) - \varphi(\bar{F}(x^{k+1}, x^k) - \bar{F}^*).$$

Since $\bar{F}(x^k, x^{k-1}) > \bar{F}^*$ for all $k \geq N_0$, the KL inequality ensures that $\text{dist}(0, \partial\bar{F}(x^k, x^{k-1})) > 0$. Therefore, division by this quantity is well defined. Combining (A.1) with the KL inequality, we obtain

$$\begin{aligned}\Delta_k &\geq \frac{\bar{F}(x^k, x^{k-1}) - \bar{F}(x^{k+1}, x^k)}{\text{dist}(0, \partial\bar{F}(x^k, x^{k-1}))}\\ &\geq \frac{\widetilde{C}\|x^k - x^{k-1}\|_2^2}{C_1(\|x^{k-1} - x^{k-2}\|_2 + \|x^k - x^{k-1}\|_2)}.\end{aligned}$$

Here, the second inequality follows from (A.2) and Lemma 3.2. Equivalently,

$$\widetilde{C}\|x^k - x^{k-1}\|_2^2 \leq C_1\Delta_k(\|x^{k-1} - x^{k-2}\|_2 + \|x^k - x^{k-1}\|_2).$$

Dividing both sides by $\widetilde{C}$ and applying the elementary inequality $4ab \leq (a+b)^2$, we obtain

$$\begin{aligned}\|x^k - x^{k-1}\|_2^2 &\leq \frac{C_1}{\widetilde{C}}\Delta_k(\|x^{k-1} - x^{k-2}\|_2 + \|x^k - x^{k-1}\|_2)\\ &= 4\left(\frac{C_1}{\widetilde{C}}\Delta_k\right)\frac{1}{4}(\|x^{k-1} - x^{k-2}\|_2 + \|x^k - x^{k-1}\|_2)\\ &\leq \left(\frac{C_1}{\widetilde{C}}\Delta_k + \frac{1}{4}(\|x^{k-1} - x^{k-2}\|_2 + \|x^k - x^{k-1}\|_2)\right)^2.\end{aligned}$$

Then,

$$\|x^k - x^{k-1}\|_2 \leq \frac{C_1}{\widetilde{C}}\Delta_k + \frac{1}{4}(\|x^{k-1} - x^{k-2}\|_2 + \|x^k - x^{k-1}\|_2).$$

By rearranging terms, we obtain

$$\frac{3}{4}\|x^k - x^{k-1}\|_2 \leq \frac{C_1}{\widetilde{C}}\Delta_k + \frac{1}{4}\|x^{k-1} - x^{k-2}\|_2.$$

Rearranging further, we obtain

$$\frac{1}{2}\|x^k - x^{k-1}\|_2 \leq \frac{C_1}{\widetilde{C}}\Delta_k + \frac{1}{4}\left(\|x^{k-1} - x^{k-2}\|_2 - \|x^k - x^{k-1}\|_2\right).$$

Summing from $k = N_0 + 1$ to N, we obtain

$$\frac{1}{2}\sum_{k=N_0+1}^{N}\|x^k - x^{k-1}\|_2 \;\leq\; \frac{C_1}{\widetilde{C}}\sum_{k=N_0+1}^{N}\Delta_k + \frac{1}{4}\sum_{k=N_0+1}^{N}\left(\|x^{k-1} - x^{k-2}\|_2 - \|x^k - x^{k-1}\|_2\right).$$

Since both sums are telescoping series, it follows that

$$\sum_{k=N_0+1}^{N}\|x^k - x^{k-1}\|_2 \leq 2\frac{C_1}{\widetilde{C}}(\varphi(\bar{F}(x^{N_0+1}, x^{N_0}) - \bar{F}^*) - \varphi(\bar{F}(x^{N+1}, x^N) - \bar{F}^*)) + \frac{1}{2}\|x^{N_0} - x^{N_0-1}\|_2 - \frac{1}{2}\|x^N - x^{N-1}\|_2.$$

Discarding the nonpositive last term yields

$$\sum_{k=N_0+1}^{N}\|x^k - x^{k-1}\|_2 \leq 2\frac{C_1}{\widetilde{C}}(\varphi(\bar{F}(x^{N_0+1}, x^{N_0}) - \bar{F}^*) - \varphi(\bar{F}(x^{N+1}, x^N) - \bar{F}^*)) + \frac{1}{2}\|x^{N_0} - x^{N_0-1}\|_2.$$

Passing to the limit as $N \to \infty$, we conclude that

$$\sum_{k=N_0+1}^{\infty}\|x^k - x^{k-1}\|_2 \leq 2\frac{C_1}{\widetilde{C}}\varphi(\bar{F}(x^{N_0+1}, x^{N_0}) - \bar{F}^*) + \frac{1}{2}\|x^{N_0} - x^{N_0-1}\|_2 < \infty.$$

Hence,

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

Therefore, the sequence $\{x^k\}$ has finite length and converges to a critical point of F, equivalently, to a critical point of problem (1.1). □



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)