A Complexity Analysis Framework for Active Manifold Identification with Applications to L_0 and L_p Regularization Models

Min Tao[1*] and Xiao-Ping Zhang[2]

[1*]Department of Mathematics, National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China.
[2]Shenzhen Key Laboratory of Ubiquitous Data Enabling, Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China.

*Corresponding author(s). E-mail(s): taom@nju.edu.cn;
Contributing authors: xpzhang@ieee.org;

Abstract

Many applications involve nonsmooth optimization problems that often exhibit a low-dimensional structure in their optimal solutions. The Projection Gradient method (PG), the Alternating Direction Method of Multipliers (ADMM), and the Accelerated Projection Gradient method (APG) are particularly effective for solving nonconvex composite programming problems and are known to determine the optimal sparsity pattern after a finite number of iterations. However, the exact number of iterations required to identify the final sparsity pattern remains an open problem. In this work, we develop a novel analytical framework to characterize the complexity of determining the active manifold and provide a rigorous proof. Using this framework, we show that PG, ADMM, and APG satisfy the necessary assumptions, enabling us to characterize the complexity of identifying the final active manifold for composite programs with nonsmooth, nonconvex regularizers, such as the $\boldsymbol{L_0}$ and $\boldsymbol{L_p}$ norms, without requiring nondegeneracy conditions.

Keywords: Partly smooth, Active set, Nonsmooth analysis, Projection gradient method, Alternating direction method of multipliers, Accelerated projection gradient method

MSC Classification: 90C26 , 90C30 , 65K05

1 Introduction

Consider the following nonconvex, nonsmooth composite programming problem:

$$\min_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x}) := r(\mathbf{x}) + \varphi(\mathbf{x}), \tag{1}$$

where $\mathcal{X} = \mathbb{R}^n$. The regularization function $r : \mathbb{R}^n \to \mathbb{R}$ is proper, lower semicontinuous (possibly nonsmooth), and $\varphi : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. We assume that

$$\underline{F} := \min_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x}) > -\infty. \tag{2}$$

This type of problem arises frequently in signal processing, machine learning, and sparse optimization [1]. In particular, regularization functions like $r(\mathbf{x}) = \|\mathbf{x}\|_0$ or $r(\mathbf{x}) = \|\mathbf{x}\|_p^p$ (with $0 < p < 1$) are commonly used to enforce sparsity in the solution. These models have been widely studied in the literature [2], [3], [4], [5], [6], [7].

With the rapid growth of datasets, there is increasing interest in identifying the sparsity pattern of the solution during the early stages of iterative optimization. Active set-type methods, frequently used in structured optimization problems, have shown effectiveness in identifying the sparsity patterns of optimal solutions. A key concept in this context is the "finite identification" property, where algorithms determine the active manifold in a finite number of iterations. Proximal algorithms [8], projected gradient methods [9], projected Newton methods [10], forward-backward-type methods [6], stochastic methods [11], and second-order methods, such as accelerating inexact successive quadratic approximation [12], have demonstrated this property under non-degenerate conditions [13–16].

In addition to finite identification, the concept of *partly smoothness* [17] has become an important model property, unifying earlier notions such as *identifiable surfaces* [18] and the $\mathcal{UV}$ *theory* [19]. Partly smoothness has been extensively studied in both convex and nonconvex settings under non-degenerate conditions [20], [21], [22], and [23]. However, determining the exact timing at which the iterates enter the final active manifold remains an open problem for nonconvex composite problems [5]. Identifying this timing is crucial because, once the active manifold is identified, we can transition to higher-order algorithms to accelerate performance [24], [14], [12].

Recent advancements have been made in identifying the active manifold under various conditions. Liang et al. [6] studied the iteration complexity for finite identification using the forward-backward algorithm (including inertial/FISTA variants) applied to (1), under the assumptions that r is proper, convex, and lower semicontinuous, φ is $L_{\nabla\varphi}$-smooth, and that $\partial r(\mathbf{x}^k) \subset \mathrm{rbd}(\partial r(\mathbf{x}^\infty))$ whenever $\mathbf{x}^k \notin \mathcal{M}_{\mathbf{x}^\infty}$. Additionally, if φ is μ-strongly convex with an $L_{\nabla\varphi}$-Lipschitz continuous gradient, and r is separable, convex, and lower semicontinuous, the complexity of identifying the active manifold has been analyzed in [25] under non-degeneracy conditions. Furthermore, the identification complexity of various constant step size algorithms, e.g., accelerated proximal gradient decent algorithm (APG) [26], Douglas-Rachford operator splitting method (DRS) [27], alternating direction method of multipliers (ADMM) [28], proximal versions of stochastic algorithms [29] has been explored in terms of the algorithm's convergence rate and a problem-dependent "wiggle room" constant [11]. Similar analyses have been conducted for block coordinate descent algorithms [30].

From the above discussion, it is evident that these results primarily apply to convex problems and do not extend to nonconvex composite programs such as (1). These limitations motivate the following questions:

Q1: *Can we develop an abstract framework to characterize the complexity of manifold identification for commonly used algorithms applied to nonconvex composite programs like (1)?*

Q2: *Can we further derive the iteration complexity for identifying the final active manifold when applying these commonly used algorithms to solve some typical nonconvex cases of (1)?*

Contributions In this paper, we develop a novel framework for analyzing the complexity of manifold identification, based on five abstract assumptions: (i) Convergence to a critical point; (ii) Descent property of the merit function; (iii) Any nonzero entry of any solution of the proximal operator of r in (1) has a positive lower bound; (iv) Linear convergence of the concerned algorithm when restricted to a fixed active manifold; (v) Finite identification. We observe that properties (i)-(ii) and (iv)-(v) are inherent to the algorithm under consideration, while property (iii) pertains to the model itself. The assumptions imposed on the algorithm are satisfied by most of the algorithms mentioned above. Additionally, based on recent seminal work [4, 31, 32], if $r(\mathbf{x}) = \|\mathbf{x}\|_0$ or $r(\mathbf{x}) = \|\mathbf{x}\|_p^p$ with $0 < p < 1$, assumption (iii) also holds. One of our main contributions is to provide a rigorous analysis for characterizing the complexity of reaching the final active manifold. This significantly advances our understanding of iteration complexities for nonconvex, nonsmooth composite problems like (1), providing a definitive answer to Q1.

Equipped with this framework, we further analyze the complexity of identifying the active manifold for the Projection Gradient (PG), Alternating Direction Method of Multipliers (ADMM), and Accelerated Projection Gradient (APG) algorithms when applied to (1). Notably, our analysis does not rely on the non-degeneracy assumptions commonly used in previous works, offering a novel contribution to the study of manifold identification complexity for typical nonconvex, nonsmooth problems. This, in turn, provides a clear answer to Q2. We emphasize that the techniques developed in this paper–particularly the proof of the newly developed framework for analyzing manifold identification complexity are novel in the context of active manifold identification for nonconvex problems.

Organization: The remainder of the paper is organized as follows. Section 2 introduces the necessary concepts and results, and reviews the convergence analysis framework under the Kurdyka-Łojasiewicz (KL) property. Section 3 presents a novel **F**ramework for **A**nalyzing the **C**omplexity of **I**dentification (FACI) and provides a rigorous proof. In Section 4, we establish the theory of partly smoothness and lower bounds for L_0 and L_p regularization models (1), and verify that property (A-iii) holds within the abstract framework (FACI) for these nonconvex regularization models. Section 5 is devoted to validating that PG, APG, and ADMM satisfy all remaining assumptions in the FACI framework, and thus can apply the derived complexity results. Finally, Section 6 concludes the paper.

2 Preliminaries and Basic Concepts

Let bold letters denote vectors, e.g., $\mathbf{x} \in \mathbb{R}^n$, and let x_i and $|\mathbf{x}|$ denote the i-th entry and the absolute value of the vector, respectively. Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes their standard inner product: $\sum_{i=1}^n x_i y_i$. The notation $\|\mathbf{x}\|_p$ refers to the p-norm, defined as $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ for $0 < p < \infty$, and the subscript p is omitted when $p = 2$. $\text{sign}(\mathbf{x})$ is defined as a vector of the same length as $\mathbf{x}$ with its i-component equal to the set $[-1, 1]$ if $x_i = 0$; otherwise being the sign of each component of $\mathbf{x}$. $I(\mathbf{x}) := \{i \mid x_i \neq 0\}$ denotes its support set. Given a set of $\mathcal{D}$, $\overline{\mathcal{D}}$ denotes its complementary. Define $[n] := \{1, \cdots, n\}$. Given a matrix $A \in \mathbb{R}^{n\times n}$ or a vector $\mathbf{x} \in \mathbb{R}^n$ and an index set $\Lambda \subseteq [n]$, $A_{\Lambda,\Lambda}$, $\mathbf{x}|_\Lambda$ and $\lambda_{\min}(A)$ to denote $A(i, j)_{i,j\in\Lambda}$, $\mathbf{x}(i)_{i\in\Lambda}$ and the minimum nonzero eigenvalue of A, respectively. $A \succ \mathbf{0}$ means positive definite.

A set $\mathcal{M}$ is a manifold around a point $\mathbf{x}$ if $\mathbf{x} \in \mathcal{M}$ and there is an open set V containing $\mathbf{x}$ such that $\mathcal{M} \cap V = \{\mathbf{x} \in V \mid \Phi(\mathbf{x}) = 0\}$ where the smooth function Φ has a surjective derivative throughout V. An extended-real-valued function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is said to be proper if $\text{dom} f = \{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) < \infty\}$ is nonempty. We denote the extended reals by $\overline{\mathbb{R}} = [-\infty, +\infty]$.

We review some definitions from [33]. Consider a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at a point $\mathbf{x} \in \mathbb{R}^n$, the subderivative $dh(\mathbf{x})(\cdot) : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined by

$$dh(\mathbf{x})(\bar{\mathbf{w}}) = \liminf_{\tau\downarrow 0 \mathbf{w}\to\bar{\mathbf{w}}} \frac{h(\mathbf{x} + \tau\mathbf{w}) - h(\mathbf{x})}{\tau},$$

and the regular subdifferential $\hat{\partial} h(\hat{\mathbf{x}})$ and the limiting subdifferential $\partial h(\hat{\mathbf{x}})$ are defined as

$$\hat{\partial} h(\hat{\mathbf{x}}) = \left\{ \mathbf{v} \middle| \liminf_{\mathbf{x}\to\hat{\mathbf{x}}} \frac{h(\mathbf{x}) - h(\hat{\mathbf{x}}) - \langle \mathbf{v}, \mathbf{x} - \hat{\mathbf{x}} \rangle}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2} \geq 0 \right\},$$

$$\partial h(\hat{\mathbf{x}}) = \left\{ \mathbf{v} \middle| \exists\, \mathbf{x}^r \to \hat{\mathbf{x}}, h(\mathbf{x}^r) \to h(\hat{\mathbf{x}}), \mathbf{v}^r \in \hat{\partial} h(\mathbf{x}^k), \mathbf{v}^k \to \mathbf{v} \right\},$$

respectively. The horizon subdifferential is defined by

$$\partial^\infty h(\mathbf{x}) = \left\{ \lim_r \lambda_r \mathbf{v}^r \;\middle|\; \mathbf{v}^r \in \partial h(\mathbf{x}^r),\ \mathbf{x}^r \to \mathbf{x},\ h(\mathbf{x}^r) \to h(\mathbf{x}),\ \lambda_r \downarrow 0 \right\}.$$

Suppose that $h(\hat{\mathbf{x}})$ is finite and $\partial h(\hat{\mathbf{x}}) \neq \emptyset$, h is *regular* at $\hat{\mathbf{x}}$ if and only if h is locally lower semicontinuous at $\hat{\mathbf{x}}$ with $\partial h(\hat{\mathbf{x}}) = \hat{\partial} h(\hat{\mathbf{x}})$ and $\partial^\infty h(\hat{\mathbf{x}}) = \hat{\partial} h(\hat{\mathbf{x}})^\infty$ [33, Corollary 8.11], where $\hat{\partial} h(\hat{\mathbf{x}})^\infty$ is recession cone (in the sense of convex analysis). Given a vector

$\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_0 = \sum_{i=1}^n |x_i|_0$, and

$$\hat{\partial}\|\mathbf{x}\|_0 = \partial\|\mathbf{x}\|_0 = \left\{\mathbf{v} \in \mathbb{R}^n | \ v_i = \left\{\begin{array}{l} 0, \ i \in \mathrm{supp}(\mathbf{x}) \\ \mathbb{R}, \ i \notin \mathrm{supp}(\mathbf{x}) \end{array}\right\}\right\}.$$

and let $0 < p < 1$,

$$\hat{\partial}(\|\mathbf{x}\|_p^p) = \partial(\|\mathbf{x}\|_p^p) = \left\{\mathbf{v} \in \mathbb{R}^n | \ v_i = \left\{\begin{array}{ll} px_i^{p-1}\,\mathrm{sign}(x_i), & i \in \mathrm{supp}(\mathbf{x}) \\ \mathbb{R}, & i \notin \mathrm{supp}(\mathbf{x}) \end{array}\right\}\right\}.$$

Let f be a proper lower semicontinuous (l.s.c.) function and $\alpha > 0$, its proximal mapping defines as

$$\mathrm{Prox}_{f/\alpha}(\mathbf{v}) = \arg\min_{\mathbf{x}} \left\{f(\mathbf{x}) + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{v}\|_2^2\right\}.$$

We assume that f is μ-strongly convex so that for some $\mu > 0$, we have $f(\mathbf{y}) \geq f(\mathbf{x}) + (\mathbf{y}-\mathbf{x})^\top \mathbf{p} + \frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$, for any $\mathbf{p} \in \partial f(\mathbf{x})$. Second, we assume that the gradient of f is ℓ-Lipschitz continuous, meaning that

$$\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\| \leq \ell\|\mathbf{x} - \mathbf{y}\|, \ \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Next, we review the concepts of partly smooth [17, Definition 2.7] and the Kurdyka-Łojasiewicz (KL) property [34, 35] from variational analysis.

Definition 1. *(Partly smooth) Suppose that the set $\mathcal{M} \subset \mathbb{R}^n$ contains the point $\mathbf{x}$. The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is partly smooth at $\mathbf{x}$ relative to $\mathcal{M}$ if $\mathcal{M}$ is a manifold around $\mathbf{x}$ and the following four properties hold:*

(i) (Restricted Smoothness) the restriction $f|_\mathcal{M}$ is smooth around $\mathbf{x}$;
(ii) (Regularity) at every point close to $\mathbf{x}$ in $\mathcal{M}$, the function f is regular and has a subgradient;
(iii) (Normal Sharpness) $df(\mathbf{x})(-\mathbf{w}) > -df(\mathbf{x})(\mathbf{w})$ for all nonzero directions $\mathbf{w}$ in $N_\mathcal{M}(\mathbf{x})$;
(iv) (Subgradient Continuity) the subdifferential map ∂f is continuous at $\mathbf{x}$ relative to $\mathcal{M}$.

Definition 2. *(Kurdyka-Lojasiewicz property) We say a proper closed function $h : \mathbb{R}^n \to (-\infty, +\infty]$ satisfies the KL property at a point $\hat{\mathbf{x}} \in \mathrm{dom}\partial h$ if there exist a constant $\alpha \in (0, \infty]$, a neighborhood U of $\hat{\mathbf{x}}$, and a continuous concave function $\phi : [0, \nu) \to [0, \infty)$ with $\phi(0) = 0$ such that*

(i) ϕ is continuously differentiable on $(0, \nu)$ with $\phi' > 0$ on $(0, \nu)$;
(ii) for every $\mathbf{x} \in U$ with $h(\hat{\mathbf{x}}) < h(\mathbf{x}) < h(\hat{\mathbf{x}}) + \nu$, it holds that

$$\phi'(h(\mathbf{x}) - h(\hat{\mathbf{x}}))\mathrm{dist}(\mathbf{0}, \partial h(\mathbf{x})) \geq 1.$$

Next, we review the framework for proving global sequential convergence using the KL property [34].

Proposition 1 *(General convergence framework) Let $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper lower semicontinuous function. Consider a sequence that satisfies the following three conditions:*

(i) (descent property) There exists $c > 0$ such that

$$\psi(\mathbf{x}^{k+1}) + c\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 \leq \psi(\mathbf{x}^k)$$

holds for all $k \in \mathbb{N}$.
(ii) (relative error) There exist $b > 0$ and $\boldsymbol{\xi}^{k+1} \in \partial\psi(\mathbf{x}^{k+1})$ such that

$$\|\boldsymbol{\xi}^{k+1}\|_2 \leq b\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2$$

holds for any $k \in \mathbb{N}$.
(iii) (continuity condition) There exist a subsequence $\{\mathbf{x}^{k_j} : j \in \mathbb{N}\}$ and $\mathbf{x}^*$ such that

$$\mathbf{x}^{k_j} \to \mathbf{x}^* \text{ and } \psi(\mathbf{x}^{k_j}) \to \psi(\mathbf{x}^*), \text{ as } j \to +\infty.$$

If ψ satisfies the KL property at $\mathbf{x}^*$, then $\sum_{k=1}^{\infty} \|\mathbf{x}^k - \mathbf{x}^{k-1}\| < +\infty$, $\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^*$, and $0 \in \partial\psi(\mathbf{x}^*)$.

The following descent lemma will play a central role in the convergence analysis. For its proof we refer to [26, Lemma 1.2.3].

Lemma 1. *For the differentiable function $\varphi : \mathbb{R}^n \to \mathbb{R}$ with $L_{\nabla\varphi}$-Lipschitz continuous gradient it holds*

$$|\varphi(\mathbf{u}) - \varphi(\mathbf{v}) - \langle \mathbf{u} - \mathbf{v}, \nabla\varphi(\mathbf{v})\rangle| \leq \frac{L_{\nabla\varphi}}{2}\|\mathbf{u} - \mathbf{v}\|^2 \quad \forall\, \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

3 A Framework for Analyzing the Complexity for Identification

Let $\{\mathbf{x}^k\}$ be the sequence generated by Algorithm $\mathcal{A}$, where $\mathbf{x}^k \to \mathbf{x}^\infty$ and $\mathbf{x}^\infty$ denotes the limiting critical point. We first define the *minimum index number* at which the iterates produced by Algorithm $\mathcal{A}$ enter the final active manifold. Subsequently, we develop a unified framework based on five assumptions that characterize the general properties of Algorithm $\mathcal{A}$ and the model (1), under which we analyze and quantify the iteration complexity required to identify the final active manifold.

Minimum Index for Manifold Identification:
Let $\{\mathbf{x}^k\}$ be the sequence generated by Algorithm $\mathcal{A}$ with $\mathbf{x}^k \to \mathbf{x}^\infty$. The *minimum index* J is defined as the smallest integer such that $\mathbf{x}^k \in \mathcal{M}_\infty$ for all $k \geq J$, where $\mathcal{M}_\infty$ denotes the final active manifold associated with the limiting point $\mathbf{x}^\infty$.

Next, we present an abstract framework consisting of a set of assumptions imposed on the sequence $\{\mathbf{x}^k\}$ generated by Algorithm $\mathcal{A}$, the merit function sequence $\psi^k := \psi(\mathbf{x}^k)$, and the proximity structure of r.

Framework for Analyzing the Complexity for Identification (FACI)
Let $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, lower semicontinuous function, and define $\psi^k := \psi(\mathbf{x}^k)$. Let $\{\mathbf{x}^k\}$ be a sequence generated by Algorithm $\mathcal{A}$, with $\mathbf{x}^{k+1} \in \text{prox}_{r/c}(\mathbf{q}^k)$, where $c > 0$ and k denotes iteration number. The following conditions hold:

(A-i) (Convergence to a Critical Point) $\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^\infty$ and $0 \in \partial\psi(\mathbf{x}^\infty)$.
(A-ii) (Descent Property) There exists a constant $c_1 > 0$ such that

$$\psi^{k+1} \leq \psi^k - c_1\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2, \quad \forall k \in \mathbb{N}.$$

(A-iii) (Lower Bound for Nonzero Entries) For $\mathbf{x}^{k+1} \in \text{prox}_{r/c}(\mathbf{q}^k)$, there exists a scalar $\nu > 0$ (depending on c, but independent on $\mathbf{q}^k$) such that

$$|x_i^{k+1}| \geq \nu \quad \text{for all } i \in I(\mathbf{x}^{k+1}),$$

where $I(\cdot)$ denotes the support set.
(A-iv) (Linear Convergence on a Fixed Subspace) There exists a constant $q \in (0, 1)$ such that

$$\psi^{k+1} - \psi(\mathbf{x}^\infty) \leq q\left(\psi^k - \psi(\mathbf{x}^\infty)\right),$$

for all $\mathbf{x}^k, \mathbf{x}^{k+1} \in \Pi_\Lambda$, where

$$\Pi_\Lambda := \{\mathbf{x} \in \mathbb{R}^n \mid I(\mathbf{x}) = \Lambda\}.$$

(A-v) (Finite Manifold Identification) There exists an index K such that

$$\mathbf{x}^k \in \mathcal{M}_\infty \quad \text{for all } k \geq K,$$

where $\mathcal{M}_\infty := \mathcal{M}_{\mathbf{x}^\infty}$ and $\mathbf{x}^\infty$ is the critical point defined in (A-i).

We define the minimum index number by

$$J = \inf_k \left\{k : \ \psi^k < \psi^\infty + c_1\nu^2 \text{ and } \psi^{k-1} - \psi^\infty \geq c_1\nu^2\right\}. \tag{3}$$

It follows that $\mathbf{x}^j \in \mathcal{M}_{\mathbf{x}^\infty}$ for all $j \geq J$. Suppose otherwise. Then there exists $k \geq J$ with $\text{supp}(\mathbf{x}^k) \neq \text{supp}(\mathbf{x}^{k+1})$. In this case, we obtain the following

$$c_1 \sum_{k=J}^{\infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \geq c_1\nu^2,$$

which contradicts

$$c_1 \sum_{k=J}^{\infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \leq \sum_{k=J}^{\infty} (\psi^k - \psi^{k+1}) \leq \psi^J - \psi^{\infty} < c_1 \nu^2.$$

Hence, the definition (3) indeed characterizes the minimum index number J for entering the active manifold.

Theorem 2 *Suppose Algorithm $\mathcal{A}$ is applied to problem* (1) *and the sequence of merit functions satisfies assumptions (A-i)-(A-v). Then the minimum index number J for identifying the active manifold $\mathcal{M}_\infty$ satisfies* $J \leq J^{\diamond} := \frac{\psi^0 - \psi^\infty}{c_1 \nu^2} + \log_{1/q} \frac{\psi^0 - \psi^\infty}{c_1 \nu^2}$.

Proof We outline the main steps to characterize the minimum index number J.

By assumption (A-ii), we have

$$\psi^{k+1} \leq \psi^k - c_1 \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2.$$

By the minimality of J, we know that $\mathbf{x}^k \in \mathcal{M}_{\mathbf{x}^\infty}$ for $k \geq J$, but $\mathbf{x}^{J-1} \notin \mathcal{M}_{\mathbf{x}^\infty}$. Let J denote the minimum index number. Then, for $k \geq J$, we have $I(\mathbf{x}^k) = I(\mathbf{x}^\infty)$. Let the index set $\mathcal{I}_J = \{0, 1, 2, \ldots, J\}$. We divide $\mathcal{I}_J$ into two mutually disjoint sets, $\mathcal{I}_{1,J}$ and $\mathcal{I}_{2,J}$, as follows:

$$\begin{aligned} \mathcal{I}_{1,J} &:= \{k \in \mathcal{I}_J : I(\mathbf{x}^k) = I(\mathbf{x}^{k+1})\}, \\ \mathcal{I}_{2,J} &:= \{k \in \mathcal{I}_J : I(\mathbf{x}^k) \neq I(\mathbf{x}^{k+1})\}. \end{aligned} \tag{4}$$

Clearly, we have $\mathcal{I}_J = \mathcal{I}_{1,J} \cup \mathcal{I}_{2,J}$.

If $k \in \mathcal{I}_{2,J}$, then by invoking assumption (A-iii), we obtain the following:

$$\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \geq \nu,$$

which directly follows from

$$\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \geq \min \left(\inf_{i \in I(\mathbf{x}^k)} |x_i^k|, \inf_{i \in I(\mathbf{x}^{k+1})} |x_i^{k+1}| \right) \geq \nu.$$

If $k \in \mathcal{I}_{1,J}$, the merit function $\psi^k - \psi^\infty$ converges linearly due to being restricted to the space $\Pi_\Lambda = \{\mathbf{x} \mid I(\mathbf{x}) = \Lambda\}$ by invoking assumption (A-iii).

Next, we divide into two cases to prove: **Case (A):** $0 \in \mathcal{I}_{2,J}$, **Case (B):** $0 \in \mathcal{I}_{1,J}$.
Case (A): $0 \in \mathcal{I}_{2,J}$. For the index set $\mathcal{I} = \{0, \ldots, J\}$, we partition it into the following non-overlapping subsets:

$$\mathcal{I}_{1,J} = \left\{ \underbrace{k_1, \ldots, k_1 + i_1 - 1}_{\triangledown(1)}, \underbrace{k_2, \ldots, k_2 + i_2 - 1}_{\triangledown(2)}, \ldots, \underbrace{k_s, \ldots, k_s + i_s - 1}_{\triangledown(s)} \right\},$$

$$\mathcal{I}_{2,J} = \left\{ \underbrace{0, \ldots, k_1 - 1}_{\diamondsuit(1)}, \underbrace{k_1 + i_1, \ldots, k_2 - 1}_{\diamondsuit(2)}, \ldots, \underbrace{k_{s-1} + i_{s-1}, \ldots, k_s - 1}_{\diamondsuit(s-1)} \right\}.$$

The iterate on each unit $\triangledown(t)$ $(1 \leq t \leq s)$, it achieves the linear convergence with the factor $q_{\triangledown(1)}, \cdots, q_{\triangledown(s)} \in (0, 1)$, respectively. Let $q := \max(q_{\triangledown(1)}, \cdots, q_{\triangledown(s)})$. Also, $0 < q < 1$.
Invoking (3) and (4), $J \in \mathcal{I}_{1,J}$. It implies that $J = k_s + i_s - 1$. $J - 1 = k_s + i_s - 2$. To proceed, we further divide it into two cases to verify.

Case (a): $J - 1 \in \mathcal{I}_{1,J}$.

On the space corresponds to the unit $\triangledown(1)$, we have

$$\psi^{k_1+i_1} - \psi^{\infty} \leq q^{i_1}(\psi^{k_1} - \psi^{\infty}).$$

Analogously, for other units of $\triangledown(t)$ $(2 \leq t \leq s-1)$, we have that

$$\psi^{k_t+i_t} - \psi^{\infty} \leq q^{i_t}(\psi^{k_t} - \psi^{\infty}),$$

For the iterates on the last union of $\triangledown(s)$, we deduce that

$$\psi^{k_s+i_s-2} - \psi^{\infty} \leq q^{i_s-2}(\psi^{k_s} - \psi^{\infty}),$$

respectively. Equivalent transformation of the above leads to

$$i_1 \leq \log_{1/q} \frac{\psi^{k_1} - \psi^{\infty}}{\psi^{k_1+i_1} - \psi^{\infty}}.$$

$$\text{for } 2 \leq t \leq s-1, \quad i_t \leq \log_{1/q} \frac{\psi^{k_t} - \psi^{\infty}}{\psi^{k_t+i_t} - \psi^{\infty}}.$$

$$i_s - 2 \leq \log_{1/q} \frac{\psi^{k_s} - \psi^{\infty}}{\psi^{k_s+i_s-2} - \psi^{\infty}}.$$

Combining these above inequalities, we have that

$$\sum_{t=1}^{s} i_t - 2 \leq \log_{1/q} \left(\Pi_{j=1}^{s-1} \left(\frac{\psi^{k_j} - \psi^{\infty}}{\psi^{k_j+i_j} - \psi^{\infty}} \right) \times \frac{\psi^{k_s} - \psi^{\infty}}{\psi^{k_s+i_s-2} - \psi^{\infty}} \right).$$

Invoking

$$\frac{\psi^{k_j} - \psi^{\infty}}{\psi^{k_{j-1}+i_{j-1}} - \psi^{\infty}} \le 1, \quad j = 2, \cdots, s,$$

we have

$$\sum_{t=1}^{s} i_t - 2 \le \log_{1/q} \frac{\psi^{k_1} - \psi^{\infty}}{\psi^{k_s+i_s-2} - \psi^{\infty}} \le \log_{1/q} \frac{\psi^{k_1} - \psi^{\infty}}{c_1\nu^2} \le \log_{1/q} \frac{\psi^{0} - \psi^{\infty}}{c_1\nu^2}. \tag{5}$$

The the second-to-last inequality is due to $\psi^{J-1} - \psi^{\infty} \ge c_1\nu^2$, and the last is due to $\psi^{k_1} \le \psi^0$. On the other hand, since $\psi^{J-1} - \psi^{\infty} \ge c_1\nu^2$, and $k_s \in \nabla(s)$ where $\nabla(s) \in \mathcal{I}_{1,J}$, it follows from

$$\psi^{k_s} - \psi^{\infty} \ge \psi^{J-1} - \psi^{\infty} \ge c_1\nu^2. \tag{6}$$

Furthermore,

$$\psi^{k_s} = \sum_{t=2}^{s} [(\psi^{k_t} - \psi^{k_{t-1}+i_{t-1}}) + (\psi^{k_{t-1}+i_{t-1}} - \psi^{k_{t-1}})] + (\psi^{k_1} - \psi^0) + \psi^0. \tag{7}$$

Next, we will consider each term inside the parentheses separately. So, we have

$$\begin{aligned} &\text{on each } \Diamond(t-1): \quad \psi^{k_t} - \psi^{k_{t-1}+i_{t-1}} \le -\left(k_t - (k_{t-1} + i_{t-1})\right) c_1\nu^2, \; t = 3, \ldots, s, \\ &\text{on each } \nabla(t): \quad \psi^{k_t+i_t} - \psi^{k_t} \le 0, \; t = 1, \ldots, s, \\ &\text{on } \Diamond(1): \quad \psi^{k_1} - \psi^0 \le -k_1 c_1 \nu^2. \end{aligned} \tag{8}$$

Combining (7) and (8), we have that

$$\psi^{k_s} \le \psi^0 - c_1\nu^2 \left(k_1 + k_2 - (k_1 + i_1) + \cdots + k_s - (k_{s-1} + i_{s-1})\right).$$

On the other hand, (6) leads to

$$\psi^{k_s} \ge \psi^{\infty} + c_1\nu^2.$$

Combining the above two inequalities, we obtain that

$$\psi^0 - c_1\nu^2 \left(k_1 + k_2 - (k_1 + i_1) + \cdots + k_s - (k_{s-1} + i_{s-1})\right) \ge \psi^{\infty} + c_1\nu^2.$$

Thus,

$$(k_s - \sum_{t=1}^{s-1} i_t)c_1\nu^2 + c_1\nu^2 \le \psi^0 - \psi^\infty.$$

Consequently,

$$\begin{aligned} J = k_s + i_s - 1 &\le \frac{\psi^0 - \psi^\infty}{c_1\nu^2} + \sum_{t=1}^{s} i_t - 2 \\ &\le \frac{\psi^0 - \psi^\infty}{c_1\nu^2} + \log_{1/q}\frac{\psi^0 - \psi^\infty}{c_1\nu^2}, \end{aligned}$$

where the last is due to (5).
Case (b): $J - 1 \in \mathcal{I}_{2,J}$.
Then, $i_s = 1$. So, $J = k_s$. Recall the definitions of the units $\nabla(t)$ $(1 \le t \le s-1)$ and $\mathcal{I}_{1,J}$, we have the iterates of $\mathbf{x}^{k_t}, \cdots \mathbf{x}^{k_t+i_t-1}, \mathbf{x}^{k_t+i_t}$ satisfying the following:

$$\text{on } \nabla(t): \quad \psi^{k_t+i_t} - \psi^\infty \le q^{i_t}(\psi^{k_t} - \psi^\infty).$$

Combining all these inequalities above, we have that

$$\sum_{t=1}^{s-1} i_t \le \log_{1/q}\left[\Pi_{j=1}^{s-1}\left(\frac{\psi^{k_j} - \psi^\infty}{\psi^{k_j+i_j} - \psi^\infty}\right)\right] = \log_{1/q}\left(\frac{\psi^{k_1} - \psi^\infty}{\psi^{k_{s-1}+i_{s-1}} - \psi^\infty}\right). \tag{9}$$

Since $k_{s-1} + i_{s-1}$ and $J - 1$ all both belong to $\diamondsuit(s-1)$, then $k_{s-1} + i_{s-1} \le J - 1$. It leads to

$$\psi^{k_{s-1}+i_{s-1}} - \psi^\infty \ge \psi^{J-1} - \psi^\infty \ge c_1\nu^2.$$

Thus, (9) leads to

$$\sum_{t=1}^{s-1} i_t \le \log_{1/q}\frac{\psi^0 - \psi^\infty}{c_1\nu^2}.$$

On the other hand, since $J - 1 = k_s - 1$ and $\psi^{J-1} - \psi^\infty \ge c_1\nu^2$, it yields that

$$\psi^{k_s-1} - \psi^\infty \ge c_1\nu^2. \tag{10}$$

On the other hand, we have

$$\psi^{k_s-1} = \underbrace{(\psi^{k_s-1} - \psi^{k_{s-1}+i_{s-1}})}_{\le(k_s-1-(k_{s-1}+i_{s-1}))c_1\nu^2} + \underbrace{(\psi^{k_{s-1}+i_{s-1}} - \psi^{k_{s-1}})}_{\le 0} + \underbrace{\psi^{k_{s-1}} - \psi^{k_{s-2}+i_{s-2}}}_{\le(k_{s-1}-(k_{s-2}+i_{s-2}))c_1\nu^2}$$

$$+ \underbrace{\psi^{k_{s-2}+i_{s-2}} - \psi^{k_{s-2}}}_{\leq 0} + \cdots + \underbrace{(\psi^{k_2} - \psi^{k_1+i_1})}_{\leq (k_2-(k_1+i_1))c_1\nu^2} + \underbrace{\psi^{k_1+i_1} - \psi^{k_1}}_{\leq 0} + \underbrace{\psi^{k_1} - \psi^0}_{\leq k_1 c_1 \nu^2}$$
$$+ \psi^0.$$

So,

$$\psi^{k_s-1} \leq \psi^0 - c_1\nu^2 \left(k_1 + k_2 - (k_1+i_1) + \cdots + k_s - 1 - (k_{s-1}+i_{s-1})\right).$$

Combining the above inequality with (10), we obtain that

$$\psi^0 - c_1\nu^2 \left(k_1 + k_2 - (k_1+i_1) + \cdots + k_s - 1 - (k_{s-1}+i_{s-1})\right) \geq \psi^\infty + c_1\nu^2.$$

It leads to

$$k_s - \sum_{t=1}^{s-1} i_t \leq \frac{\psi^0 - \psi^\infty}{c_1\nu^2}.$$

Consequently,

$$\begin{aligned} J = k_s &\leq \frac{\psi^0 - \psi^\infty}{c_1\nu^2} + \sum_{t=1}^{s-1} i_t \\ &= \frac{\psi^0 - \psi^\infty}{c_1\nu^2} + \log_{1/q} \frac{\psi^0 - \psi^\infty}{c_1\nu^2}. \end{aligned}$$

Case (B): $0 \in \mathcal{I}_{1,J}$. For the index set $\mathcal{I} = \{0, \ldots, J\}$, we partition it into two non-overlapping subsets:

$$\mathcal{I}_{1,J} = \left\{ \underbrace{0, \ldots, k_1 - 1}_{\triangledown(1)}, \underbrace{k_1 + i_1, \ldots, k_2 - 1}_{\triangledown(2)}, \ldots, \underbrace{k_{s-1} + i_{s-1}, \ldots, k_s}_{\triangledown(s)} \right\},$$
$$\mathcal{I}_{2,J} = \left\{ \underbrace{k_1, \ldots, k_1 + i_1 - 1}_{\diamond(1)}, \underbrace{k_2, \ldots, k_2 + i_2 - 1}_{\diamond(2)}, \ldots, \underbrace{k_{s-1}, \ldots, k_{s-1} + i_{s-1} - 1}_{\diamond(s-1)} \right\}.$$

According to the definition (4), we have $J \in \mathcal{I}_{1,J}$. We further divide into two cases to prove: (a) $J-1 \in \mathcal{I}_{2,J}$; and (b) $J-1 \in \mathcal{I}_{1,J}$.

Case (a):$J-1 \in \mathcal{I}_{2,J}$. Thus, $J = k_{s-1} + i_{s-1}$. Note that

$$J = k_{s-1} + i_{s-1}$$

$$= \underbrace{k_1 + \left(\sum_{t=1}^{s-2}(k_{t+1} - (k_t + i_t))\right)}_{\spadesuit} + \underbrace{\sum_{t=1}^{s-1} i_t}_{\clubsuit}.$$

Next, we estimate the first term $\spadesuit$ in the right-hand-side of the above equality. Consider each space determined by the units in $\mathcal{I}_{1,J}$, we have the following inequalities:

$$\psi^{k_1} - \psi^\infty \le q^{k_1}(\psi^0 - \psi^\infty),$$

$$\psi^{k_2} - \psi^\infty \le q^{k_2-(k_1+i_1)}(\psi^{k_1+i_1} - \psi^\infty),$$

$$\vdots$$

$$\psi^{k_{s-1}} - \psi^\infty \le q^{k_{s-1}-(k_{s-2}+i_{s-2})}(\psi^{k_{s-2}+i_{s-2}} - \psi^\infty).$$

Thus, we can get an upper bound for

$$\spadesuit = k_1 + \sum_{t=1}^{s-2}(k_{t+1} - (k_t + i_t)) \le \log_{1/q}(\frac{\psi^0 - \psi^\infty}{\psi^{k_{s-1}} - \psi^\infty}) \le \log_{1/q}\frac{\psi^0 - \psi^\infty}{c_1\nu^2},$$

where the last inequality is due to $\psi^{k_{s-1}} \ge \psi^{J-1} \ge \psi^\infty + c_1\nu^2$.
On the other hand, we have that

$$\psi^{k_1} - c_1\nu^2\,(i_1 + \cdots + i_{s-1} - 1) - \psi^\infty \ge c_1\nu^2.$$

It leads to

$$\clubsuit = \sum_{t=1}^{s-1} i_t \le \frac{\psi^{k_1} - \psi^\infty}{c_1\nu^2} \le \frac{\psi^0 - \psi^\infty}{c_1\nu^2}.$$

Consequently, we have

$$J \le \frac{\psi^0 - \psi^\infty}{c_1\nu^2} + \log_{1/q}\frac{\psi^0 - \psi^\infty}{c_1\nu^2}.$$

Case (b): $J - 1 \in \mathcal{I}_{1,J}$. The proof is similar, and can get the same bound, thus omitted here.

Thus, the conclusion follows directly. □

4 Partly Smoothness of L_0 and L_p Regularization Models

In this section, we illustrate the partly smoothness of several nonconvex nonsmooth functions. To carry out a unified analysis, we consider the function $r(\mathbf{x})$ is defined in Table 1. Since $r(\mathbf{x}) = \sum_{i=1}^{n} r_i(x_i)$, we denote $r_\Lambda(\mathbf{x}) = \sum_{i\in\Lambda} r_i(x_i)$ where $\Lambda \subseteq \{1, \cdots, n\}$. For any vector $\overline{\mathbf{x}}(\neq \mathbf{0})$, we define the manifold

$$\mathcal{M}_{\overline{\mathbf{x}}} = \{\mathbf{x} \in \mathbb{R}^n \mid I(\mathbf{x}) = I(\overline{\mathbf{x}})\}. \tag{11}$$

We first characterize the normal cone (normal space) as $\mathcal{M}_{\overline{\mathbf{x}}}$:

$$N_{\mathcal{M}}(\overline{\mathbf{x}}) = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w}_j = 0,\ j \in I(\overline{\mathbf{x}})\}\,.$$

Table 1 Non-Separable Regularizers

Name	$\sigma(\mathbf{x})$
(a) L_0	$\\|\mathbf{x}\\|_0 = \sum_{i=1}^{n} \\|x_i\\|_0$
(b) L_p $(0 < p < 1)$	$\\|\mathbf{x}\\|_p^p = \sum_{i=1}^{n} \|x_i\|^p$

Proposition 3 *Suppose $\varphi(\mathbf{x})$ is smooth and the function $r(\mathbf{x}) = \gamma\sigma(\mathbf{x})$ where σ is defined in Table 1 and $\gamma > 0$. The objective function $F(\mathbf{x})$ in (1) is partly smooth at $\overline{\mathbf{x}} \neq \mathbf{0}$ relative to $\mathcal{M}_{\overline{\mathbf{x}}}$.*

Proof Since $\varphi(\mathbf{x})$ is smooth, we only need to prove that $r(\mathbf{x})$ is partly smooth at $\overline{\mathbf{x}} \neq \mathbf{0}$ relative to $\mathcal{M}_{\overline{\mathbf{x}}}$. By invoking Definition 1, the properties of (i), (ii), and (iv) are obviously valid. We only need to verify the property (iii) normal sharpness. Denote $\Lambda = I(\overline{\mathbf{x}})$. Let $\mathbf{w}(\neq \mathbf{0})$ in $N_{\mathcal{M}}(\overline{\mathbf{x}})$. Next, we prove the following assertion for each $r(\cdot)$ with $\sigma(\cdot)$ defined in Table 1.

$$\begin{aligned} & dr(\overline{\mathbf{x}})(\mathbf{w}) + dr(\overline{\mathbf{x}})(-\mathbf{w}) \\ = & \liminf_{\tau\downarrow 0 \overline{\mathbf{w}}\to\mathbf{w}} \frac{r(\overline{\mathbf{x}} + \tau\overline{\mathbf{w}}) - r(\overline{\mathbf{x}})}{\tau} + \liminf_{\tau\downarrow 0 \overline{\mathbf{w}}\to\mathbf{w}} \frac{r(\overline{\mathbf{x}} - \tau\overline{\mathbf{w}}) - r(\overline{\mathbf{x}})}{\tau} > 0. \end{aligned} \tag{12}$$

Note that

$$\begin{aligned} dr(\overline{\mathbf{x}})(\mathbf{w}) &= \liminf_{\tau\downarrow 0 \overline{\mathbf{w}}\to\mathbf{w}} \frac{r(\overline{\mathbf{x}} + \tau\overline{\mathbf{w}}) - r(\overline{\mathbf{x}})}{\tau} \\ = & \liminf_{\tau\downarrow 0 \overline{\mathbf{w}}\to\mathbf{w}} \frac{\sum_{i\in\Lambda} r_i(\overline{x}_i + \tau\overline{w}_i) - \sum_{i\in\Lambda} r_i(\overline{x}_i) + \sum_{i\in\Lambda^c} r_i(\tau\overline{w}_i)}{\tau} \\ \geq & \underbrace{\liminf_{\tau\downarrow 0 \overline{\mathbf{w}}\to\mathbf{w}} \frac{\sum_{i\in\Lambda} r_i(\overline{x}_i + \tau\overline{w}_i) - \sum_{i\in\Lambda} r_i(\overline{x}_i)}{\tau}}_{\mathrm{I}} + \underbrace{\liminf_{\tau\downarrow 0 \overline{\mathbf{w}}\to\mathbf{w}} \frac{\sum_{i\in\Lambda^c} r_i(\tau\overline{w}_i)}{\tau}}_{\mathrm{II}}. \end{aligned}$$

(a) $r(\mathbf{x}) = \gamma\|\mathbf{x}\|_0$.

Since $\mathbf{w}(\neq \mathbf{0})$ in $N_{\mathcal{M}}(\overline{\mathbf{x}})$,

$$\mathrm{I} = \liminf_{\tau \downarrow 0 \overline{\mathbf{w}} \to \mathbf{w}} \frac{\sum_{i\in\Lambda} r_i(\overline{x}_i + \tau \overline{w}_i) - \sum_{i\in\Lambda} r_i(\overline{x}_i)}{\tau} = 0,$$

and

$$\mathrm{II} = \liminf_{\tau \downarrow 0 \overline{\mathbf{w}} \to \mathbf{w}} \frac{\sum_{i\in\Lambda^c} r_i(\tau \overline{w}_i)}{\tau} > 0.$$

Therefore,

$$dr(\overline{\mathbf{x}})(\mathbf{w}) = \liminf_{\tau \downarrow 0 \overline{\mathbf{w}} \to \mathbf{w}} \frac{r(\overline{\mathbf{x}} + \tau \overline{\mathbf{w}}) - r(\overline{\mathbf{x}})}{\tau} > 0.$$

Analogously,

$$dr(\overline{\mathbf{x}})(-\mathbf{w}) = \liminf_{\tau \downarrow 0 \overline{\mathbf{w}} \to \mathbf{w}} \frac{r(\overline{\mathbf{x}} - \tau \overline{\mathbf{w}}) - r(\overline{\mathbf{x}})}{\tau} > 0.$$

Thus, (12) also holds.

(b) $r(\mathbf{x}) = \gamma \|\mathbf{x}\|_p^p$. First, for any $\mathbf{w}(\neq \mathbf{0})$ in $N_{\mathcal{M}}(\overline{\mathbf{x}})$, we have

$$\mathrm{I} = \liminf_{\tau \downarrow 0, \overline{\mathbf{w}} \to \mathbf{w}} \frac{\sum_{i\in\Lambda} r_i(\overline{x}_i + \tau \overline{w}_i) - \sum_{i\in\Lambda} r_i(\overline{x}_i)}{\tau} = 0.$$

Second, $\mathrm{II} = \liminf_{\tau \downarrow 0, \overline{\mathbf{w}} \to \mathbf{w}} \frac{\sum_{i\in\Lambda^c} r_i(\tau \overline{w}_i)}{\tau} > 0$. Thus, $dr(\overline{\mathbf{x}})(\mathbf{w}) > 0$.

Analogously, $dr(\overline{\mathbf{x}})(-\mathbf{w}) > 0$. Consequently, (12) is valid. □

Next, we establish that the lower bound properties of the L_0 and L_p proximities validate property (A-iii) within the framework of **FACI** when the model (1) has r defined by the L_0 or L_p regularizer.

Let $\sigma(\cdot)$ be a proper l.s.c. function and prox-bounded with threshold $\lambda_\sigma = +\infty$[1], and a vector $\mathbf{q} \in \mathbb{R}^n$ and $\gamma > 0$, define $\widehat{\mathbf{x}}(\mathbf{q})$ to be any global solution of the following minimization problem:

$$\widehat{\mathbf{x}}(\mathbf{q}) \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \left(\gamma\sigma(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{q}\|_2^2 \right). \tag{13}$$

In the following, we first provide the following two lemmas.

Lemma 2. *[31] Consider the following minimization problem:*

$$x^*(z) \in \arg\min_x \left(\phi(x) = \frac{1}{2}(z - x)^2 + \gamma |x|^p \right)$$

[1] According to Definition 1.23 in [33] and Theorem 1.25 in [33], for any $\gamma > 0$, $\widehat{\mathbf{x}}(\mathbf{q})$ is well-defined.

where $z \in \mathbb{R}$ and $0 < p < 1$. Then, let

$$\underline{\nu} = (2\gamma(1-p))^{\frac{1}{2-p}}, \quad \mu = \underline{\nu} + \gamma p \underline{\nu}^{p-1}, \tag{14}$$

$$x^*(z) = \begin{cases} 0 & \text{if } |z| < \mu \\ \{0, \operatorname{sign}(z)\underline{\nu}\} & \text{if } |z| = \mu \\ \operatorname{sign}(z)\nu_* & \text{if } |z| > \mu \end{cases}$$

where for $|z| > \mu$, ν_* is the root of the equation of $\nu + \gamma p \nu^{p-1} = |z|$ in the interval $\in (\underline{\nu}, |z|)$.

Lemma 3. *[31] Consider the minimization problem (13) with $r(\mathbf{x}) = \gamma\|\mathbf{x}\|_0$, the solution $\widehat{\mathbf{x}} := \widehat{\mathbf{x}}(\mathbf{q})$ to (13) is*

$$\widehat{\mathbf{x}}_i = \begin{cases} 0 & \text{if } |q_i| < \sqrt{2\gamma} \\ \{0, q_i\} & \text{if } |q_i| = \sqrt{2\gamma} \\ q_i & \text{if } |q_i| > \sqrt{2\gamma}. \end{cases}$$

Theorem 4 *Given a vector $\mathbf{q} \in \mathbb{R}^n$ and $\mathbf{q} \neq \mathbf{0}$, let $\widehat{\mathbf{x}} := \widehat{\mathbf{x}}(\mathbf{q})$ and $\widehat{\mathbf{x}}(\mathbf{q})$ be defined in (13). If $r(\mathbf{x}) = \gamma\|\mathbf{x}\|_0$ or $r(\mathbf{x}) = \gamma\|\mathbf{x}\|_p^p$ (with $0 < p < 1$), for any $\gamma > 0$, there exists a constant $\nu > 0$ such that*

$$\widehat{\mathbf{x}}_i \geq \nu, \quad i \in I(\widehat{\mathbf{x}}), \tag{15}$$

where ν depends only on γ, and is independent of $\mathbf{q}$.

Proof If $r(\mathbf{x}) = \gamma\|\mathbf{x}\|_p^p$, by setting the lower bound $\nu := (2\gamma(1-p))^{\frac{1}{2-p}}$, (15) is satisfied. If $r(\mathbf{x}) = \gamma\|\mathbf{x}\|_0$, by setting the lower bound $\nu := \sqrt{2\gamma}$, (15) is satisfied. □

5 Applications of Complexity Analysis Framework for Specific Algorithms

In this section, we investigate several important properties of the model (1) using the PG, ADMM, and APG. These properties include global convergence, linear convergence restricted to the fixed manifold, and the finite identification property without imposing any nondegeneracy conditions.

5.1 Projection Gradient Method

Next, we present the global convergence of Algorithm 1 in Theorem 5 which proof can follow from Proposition 1 to show that the three assumptions in Proposition 1 are valid.

Theorem 5 *(Global convergence) Let $r(\mathbf{x}) = \gamma\sigma(\mathbf{x})$ and $\sigma(\mathbf{x})$ be defined in Table 1, and $\varphi : \mathbb{R}^n \to \mathbb{R}$ be $L_{\nabla\varphi}$-smooth and sub-analytic function. Suppose that $\alpha > \frac{L_{\nabla\varphi}}{2}$. Let the*

Algorithm 1 Projection Gradient Method for solving (1)

Let $\alpha > 0$, and $\mathbf{x}^0$ be a given starting point. Set $k = 0$.

1: $\mathbf{x}^{k+1} \in \text{Prox}_{r/\alpha}\left(\mathbf{x}^k - \frac{1}{\alpha}\nabla\varphi(\mathbf{x}^k)\right)$

2: Set $k = k + 1$, and go to Line 1.

sequence $\{\mathbf{x}^k\}$ be generated by Algorithm 1 and assume that it is bounded; then it converges to a critical point of (1). Moreover, the sequence $\{\mathbf{x}^k\}$ satisfies $\sum_k \|\mathbf{x}^{k+1} - \mathbf{x}^k\| < +\infty$.

Theorem 6 *(Finite identification and linear convergence) Let $r(\mathbf{x}) = \gamma\sigma(\mathbf{x})$ and $\sigma(\mathbf{x})$ be defined in Table 1, and let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a $L_{\nabla\varphi}$-smooth, sub-analytic function, and strongly convex with $\sigma > \overline{\sigma} := \lambda p(1-p)\frac{1}{\nu^{2-p}} > 0$. Let the sequence $\{\mathbf{x}^k\}$ be generated by Algorithm 1. Assume that the step size $\alpha > \frac{L_{\nabla\varphi}}{2}$ and $\{\mathbf{x}^k\}$ is bounded. Then, the following holds:*

(i) The support set of $I(\mathbf{x}^k)$ changes only finitely often;

(ii) The sequence of $\Delta^k := F(\mathbf{x}^k) - F(\mathbf{x}^{\infty,\Lambda})$ converges linearly to zero when $\mathbf{x}^k$ belongs to the subspace $\Pi_\Lambda = \{\mathbf{x} \mid I(\mathbf{x}) = \Lambda\}$ and

$$\mathbf{x}^{\infty,\Lambda} \in \arg\min_{\mathbf{x}\in\Pi_\Lambda}[r(\mathbf{x}) + \varphi(\mathbf{x})]. \tag{16}$$

Proof (i) First, the subproblem for $\mathbf{x}^{k+1}$ involves solving the following minimization problem:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\Pi_\Lambda}\left[r(\mathbf{x}) + \varphi(\mathbf{x}^k) + \nabla\varphi(\mathbf{x}^k)^\top(\mathbf{x}-\mathbf{x}^k) + \frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^k\|^2\right]. \tag{17}$$

From this we have

$$\begin{aligned} & r(\mathbf{x}^{k+1}) + \varphi(\mathbf{x}^k) + \nabla\varphi(\mathbf{x}^k)^\top(\mathbf{x}^{k+1}-\mathbf{x}^k) + \frac{\alpha}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2 \\ & \le r(\mathbf{x}) + \varphi(\mathbf{x}^k) + \nabla\varphi(\mathbf{x}^k)^\top(\mathbf{x}-\mathbf{x}^k) + \frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^k\|^2 - \frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^{k+1}\|^2 \\ & \le r(\mathbf{x}) + \varphi(\mathbf{x}) + \frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^k\|^2 - \frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^{k+1}\|^2. \end{aligned}$$

So,

$$\begin{aligned} & r(\mathbf{x}^{k+1}) + \varphi(\mathbf{x}^{k+1}) + \frac{\alpha - L_{\nabla\varphi}}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2 \\ & \le r(\mathbf{x}) + \varphi(\mathbf{x}) + \frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^k\|^2 - \frac{\alpha}{2}\|\mathbf{x}-\mathbf{x}^{k+1}\|^2. \end{aligned} \tag{18}$$

It leads to

$$\textcircled{1}_k : F(\mathbf{x}^{k+1}) \le F(\mathbf{x}^k) - \frac{\tilde{\alpha}}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2, \tag{19}$$

where $\tilde{\alpha} := 2\alpha - L_{\nabla\varphi}$. It follows from (19) and (2) that

$$\frac{\tilde{\alpha}}{2}\sum_{k=0}^{+\infty}\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \le F(\mathbf{x}^0) - \underline{F}.$$

Thus, $\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \to 0$ as $k \to +\infty$. By further combining with Theorem 4, the assertion (i) is valid.

(ii) Next, we divide into two cases to prove. Case (a): $r(\mathbf{x}) = \lambda\|\mathbf{x}\|_p^p$. For any $\mathbf{x}$ and $\Lambda := I(\mathbf{x})$, we have:

$$\left[\nabla^2 F(\mathbf{x})\right]_{\Lambda,\Lambda} := \left[\nabla^2 \varphi(\mathbf{x})\right]_{\Lambda,\Lambda} + \lambda p(p-1)\sum_{i\in\Lambda}|x_i|^{p-2}.$$

The positivity of the above matrix follows from the following:

$$\lambda_{\min}\left(\left[\nabla^2\varphi(\mathbf{x})\right]_{\Lambda,\Lambda}\right) \geq \lambda_{\min}(\nabla^2\varphi(\mathbf{x})) > \lambda p(1-p)\frac{1}{\nu^{2-p}},$$

and $\sigma > \overline{\sigma}$, we conclude that $F(\mathbf{x})$ is strongly convex with the strong convexity parameter $\tilde{\sigma} = \sigma - \overline{\sigma}$ on the space Π_Λ. Case (b): $r(\mathbf{x}) = \lambda\|\mathbf{x}\|_0$. We conclude that $F(\mathbf{x})$ is strongly convex with the strong convexity parameter $\tilde{\sigma} = \sigma$ on the space Π_Λ. Second, there exists a positive scalar τ such that

$$\textcircled{2}_k: \ \|\mathbf{x}^{k+1} - \mathbf{x}^{\infty,\Lambda}\| \leq \tau\|\mathbf{x}^k - \mathbf{x}^{k+1}\|. \tag{20}$$

We only prove that $r(\mathbf{x}) = \|\mathbf{x}\|_p^p$. The other case of $r(\mathbf{x}) = \|\mathbf{x}\|_0$ can be proved analogously. Since $\mathbf{x}^{k+1}$ and $\mathbf{x}^{\infty,\Lambda}$ on the same space, we have

$$\begin{cases} \sum_{i\in\Lambda}\partial r_i(x_i^{k+1}) + \left[\nabla\varphi(\mathbf{x}^{k+1})\right]_\Lambda + \left[\nabla\varphi(\mathbf{x}^k)\right]_\Lambda - \left[\nabla\varphi(\mathbf{x}^{k+1})\right]_\Lambda + \alpha(\mathbf{x}_\Lambda^{k+1} - \mathbf{x}_\Lambda^k) = 0 \\ \sum_{i\in\Lambda}\partial r_i(x_i^{\infty,\Lambda}) + \left[\nabla\varphi(\mathbf{x}^{\infty,\Lambda})\right]_\Lambda = 0. \end{cases}$$

Consequently,

$$\begin{aligned} \tilde{\sigma}\|\mathbf{x}_\Lambda^{k+1} - \mathbf{x}_\Lambda^{\infty,\Lambda}\|^2 &\leq \sum_{i\in\Lambda}\langle x_i^{k+1} - x_i^{\infty,\Lambda}, \partial r_i(x_i^{k+1}) - \partial r_i(x_i^{\infty,\Lambda})\rangle \\ &+\langle \mathbf{x}_\Lambda^{k+1} - \mathbf{x}_\Lambda^{\infty,\Lambda}, [\nabla\varphi(\mathbf{x}^{k+1})]_\Lambda - [\nabla\varphi(\mathbf{x}^{\infty,\Lambda})]_\Lambda\rangle \\ &= \langle \mathbf{x}_\Lambda^{k+1} - \mathbf{x}_\Lambda^{\infty,\Lambda}, [\nabla\varphi(\mathbf{x}^{k+1})]_\Lambda - [\nabla\varphi(\mathbf{x}^k)]_\Lambda\rangle + \alpha\langle \mathbf{x}_\Lambda^{k+1} - \mathbf{x}_\Lambda^{\infty,\Lambda}, \mathbf{x}_\Lambda^k - \mathbf{x}_\Lambda^{k+1}\rangle. \end{aligned}$$

It follows that (20) holds with $\tau = \frac{L_{\nabla\varphi}+\alpha}{\tilde{\sigma}}$. Third, we prove that

$$\textcircled{3}_k: \ F(\mathbf{x}^{k+1}) - F(\mathbf{x}^{\infty,\Lambda}) \leq \frac{L_{\nabla F}}{2}\|\mathbf{x}^{\infty,\Lambda} - \mathbf{x}^{k+1}\|^2. \tag{21}$$

Inequality (21) follows due to the gradient of F being Lipschitz continuous and $[\nabla F(\mathbf{x}^{\infty,\Lambda})]_\Lambda = 0$. By combining $\textcircled{1}_k$, $\textcircled{2}_k$ and $\textcircled{3}_k$, we have

$$\Delta^{k+1} \leq \frac{1}{1+\tilde{c}}\Delta^k,$$

where $\tilde{c} = \frac{\tilde{\alpha}}{\tau^2 L_{\nabla F}}$. The conclusion follows directly. □

Remark 1 From Theorem 5, the sequence $\{\mathbf{x}^k\}$ generated by Algorithm 1 satisfies $F(\mathbf{x}^k) \to F(\mathbf{x}^\infty)$, where $\mathbf{x}^\infty$ denotes the limit point of $\{\mathbf{x}^k\}$. If the projection gradient method (i.e., Algorithm 1) performs a finite number of iterations within the subspace Π_Λ, then there exists a constant $\widehat{q}_\Lambda \in (0,1)$ such that

$$F(\mathbf{x}^{k+1}) - F(\mathbf{x}^\infty) \leq \widehat{q}_\Lambda\big(F(\mathbf{x}^k) - F(\mathbf{x}^\infty)\big),$$

for all successive iterates $\mathbf{x}^k$ and $\mathbf{x}^{k+1}$ lying in the same subspace Π_Λ.

Remark 2 When Algorithm $\mathcal{A}$ is specified as the projection gradient (PG) method and we set $\psi(\mathbf{x}^{k+1}) := F(\mathbf{x}^{k+1})$, the assumptions (A-i)-(A-v) in the FACI framework are automatically satisfied as a consequence of Theorem 5, inequality (19), Theorem 6, Remark 1, and Theorem 4.

5.2 Alternating Direction Method of Multipliers

To solve the unconstrained problem in (1) via the following equivalent reformulation

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathcal{X}} & r(\mathbf{x}) + \varphi(\mathbf{y}) \\ s.t. & \mathbf{x} = \mathbf{y}, \end{array} \tag{22}$$

and we employ the alternating direction multiplier method (ADMM) to solve (22), as outlined in Algorithm 2.

Algorithm 2 Alternating Direction Method of Multipliers for solving (1)

Let $\beta > 0$, and $\mathbf{y}^0$ and $\mathbf{z}^0$ be given. Set $k = 0$.

1: $\mathbf{x}^{k+1} \in \mathrm{Prox}_{r/\beta}\left(\mathbf{y}^k - \frac{1}{\beta}\mathbf{z}^k\right)$
2: $\mathbf{y}^{k+1} = \mathrm{Prox}_{\varphi/\beta}(\mathbf{x}^{k+1} + \frac{\mathbf{z}^k}{\beta})$
3: $\mathbf{z}^{k+1} = \mathbf{z}^k + \beta\left(\mathbf{x}^{k+1} - \mathbf{y}^{k+1}\right)$
4: Set $k = k + 1$, and go to Line 1.

Theorem 7 *[Global Convergence] Let $r(\mathbf{x}) = \gamma\sigma(\mathbf{x})$, where $\sigma(\mathbf{x})$ is defined in Table 1, and let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be an $L_{\nabla\varphi}$-smooth and subanalytic function. Consider the sequence $\{\mathbf{w}^k := (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)\}$ generated by Algorithm 2. If $\beta > 2L_{\nabla\varphi}$ and $\{\mathbf{x}^k\}$ is bounded, then the sequence $\{\mathbf{w}^k\}$ has finite length, i.e., $\sum_{k=1}^{\infty} \|\mathbf{w}^{k+1} - \mathbf{w}^k\| < \infty$, and thus converges to a limit point $\mathbf{w}^\infty := (\mathbf{x}^\infty, \mathbf{y}^\infty, \mathbf{z}^\infty)$ satisfying*

$$\begin{cases} 0 \in \partial r(\mathbf{x}^\infty) + \mathbf{z}^\infty, \\ \nabla\varphi(\mathbf{y}^\infty) - \mathbf{z}^\infty = 0, \\ \mathbf{x}^\infty = \mathbf{y}^\infty. \end{cases}$$

Moreover, the sequence $\{\mathbf{x}^k\}$ converges to a stationary point $\mathbf{x}^\infty$ of problem (1).

The proof of Theorem 7 is similar to Theorem [36, Theorem 5.8] and Theorem [37, Theorem 8], thus omitted here.

Theorem 8 (Finite Identification) *Let $r(\mathbf{x}) = \gamma\sigma(\mathbf{x})$, where $\sigma(\mathbf{x})$ is defined in Table 1, and let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be an $L_{\nabla\varphi}$-smooth, subanalytic, and strongly convex function with $\sigma > \overline{\sigma} := \lambda p(1-p)/\nu^{2-p} > 0$. Suppose that the sequence $\{\mathbf{w}^k\}$ is generated by Algorithm 2. If $\beta > 2L_{\nabla\varphi}$ and $\{\mathbf{x}^k\}$ is bounded, then the support set $I(\mathbf{x}^k)$ changes only finitely many times.*

Proof First, we define

$$\mathcal{T}(\mathbf{x}, \mathbf{y}) = r(\mathbf{x}) + \varphi(\mathbf{x}) + \frac{\beta}{2}\|\mathbf{x} - \mathbf{y}\|_2^2.$$

Similarly to the proof in [37], we have that

$$\mathcal{T}^{k+1} \leq \mathcal{T}^k - \frac{3L_{\nabla\varphi}}{8}\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2^2. \tag{23}$$

Invoking $\|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2 \leq L_{\nabla\varphi}\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2$, we further have

$$F^{k+1} \leq F^k - \frac{L_{\nabla\varphi}}{8}\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2^2. \tag{24}$$

It follows from (24) that $\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2 \to 0,$ as $k \to +\infty$. Consequently,

$$\|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2 \to 0, \text{ as } k \to +\infty.$$

Additionally, using $\mathbf{z}^{k+1} = \mathbf{z}^k + \beta\left(\mathbf{x}^{k+1} - \mathbf{y}^{k+1}\right)$ and $\mathbf{z}^k = \mathbf{z}^{k-1} + \beta\left(\mathbf{x}^k - \mathbf{y}^k\right)$, we have

$$\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \leq \|\mathbf{y}^k - \mathbf{y}^{k+1}\| + \frac{1}{\beta}\left(\|\mathbf{z}^k - \mathbf{z}^{k+1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\|\right).$$

Therefore,

$$\|\mathbf{x}^k - \mathbf{x}^{k+1}\| \to 0, \text{ as } k \to +\infty. \tag{25}$$

Suppose the conclusion does not hold. Then, there exists a subsequence $\{\mathbf{x}^{k_j}\}$ such that $I(\mathbf{x}^{k_j})$ changes. So, $\|\mathbf{x}^{k_j} - \mathbf{x}^{k_j+1}\| \geq \nu$ which contradicts (25). □

Theorem 9 (Linear Convergence) *Let $r(\mathbf{x}) = \gamma\sigma(\mathbf{x})$, where $\sigma(\mathbf{x})$ is defined in Table 1, and let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be an $L_{\nabla\varphi}$-smooth, subanalytic, and strongly convex function satisfying $\sigma > \overline{\sigma} := \lambda p(1-p)/\nu^{2-p} > 0$. Let $\{\mathbf{w}^k\}$ denote the sequence generated by Algorithm 2. Assume that $\beta > 2L_{\nabla\varphi}$ and that $\{\mathbf{x}^k\}$ is bounded. Define the sequence $P_{k+1} := \mathcal{T}^{k+1} + \frac{3\kappa}{\beta^2}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2$, where $\mathbf{x}^{\infty,\Lambda}$ is defined in (16) and $\kappa := \frac{\kappa_1}{\kappa_2}$ with $\kappa_1 = \frac{3L_{\nabla\varphi}}{16}$ and $\kappa_2 = 3 + \frac{6L_{\nabla\varphi}^2}{\beta^2}$. Then, the sequence $\{P_k\}$ converges linearly to $F(\mathbf{x}^{\infty,\Lambda})$ whenever the iterate $\mathbf{x}^k$ lies in the subspace Π_Λ.*

Proof First, we have the following inequality:

$$\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \leq 3\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 + \frac{3}{\beta^2}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2 + \frac{3}{\beta^2}\|\mathbf{z}^k - \mathbf{z}^{k-1}\|^2. \tag{26}$$

Next, multiplying both sides by κ and combining with equation (23), we obtain the following:

$$P_{k+1} \leq P_k - \left(\frac{3}{8}L_{\nabla\varphi} - 3\kappa - \frac{6\kappa}{\beta^2}L_{\nabla\varphi}^2\right)\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 - \kappa\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2. \tag{27}$$

Note that $\frac{3}{8}L_{\nabla\varphi} - 3\kappa - \frac{6\kappa}{\beta^2}L_{\nabla\varphi}^2 = \frac{3}{16}L_{\nabla\varphi}$. Consequently, let $\mathbf{u} = (\mathbf{x}, \mathbf{y})$,

$$P_{k+1} \leq P_k - \frac{\frac{3L_{\nabla\varphi}}{16}}{3 + \frac{6}{\beta^2}L_{\nabla\varphi}^2}\|\mathbf{u}^k - \mathbf{u}^{k+1}\|^2. \tag{28}$$

The merit function is strongly convex when restricted to Π_Λ, i.e. (let $\mathcal{T}^{\infty,\Lambda} := \mathcal{T}(\mathbf{x}^{\infty,\Lambda}, \mathbf{x}^{\infty,\Lambda})$)

$$\mathcal{T}(\mathbf{v}) \geq \mathcal{T}(\mathbf{u}) + \langle \mathbf{v} - \mathbf{u},\ \xi\rangle + \frac{\hat{\sigma}}{2}\|\mathbf{v} - \mathbf{u}\|^2,\ \forall\, \xi \in \partial\mathcal{T}(\mathbf{u}),\ \hat{\sigma} = \min(\beta, \sigma - \bar{\sigma}).$$

Let $\mathbf{v} = \mathbf{u} - \frac{1}{\hat{\sigma}}\xi$ substitute into the left hand side of the above inequality, it leads to

$$\mathcal{T}(\mathbf{v}) \geq \mathcal{T}(\mathbf{u}) - \frac{1}{2\hat{\sigma}}\mathrm{dist}^2(\mathbf{0}, \partial\mathcal{T}(\mathbf{u})).$$

Leaving $\mathbf{v} = (\mathbf{x}^{\infty,\Lambda}, \mathbf{x}^{\infty,\Lambda})$ in the above, we have found that

$$\mathcal{T}(\mathbf{u}) - \mathcal{T}(\mathbf{x}^{\infty,\Lambda}, \mathbf{x}^{\infty,\Lambda}) \leq \frac{1}{2\hat{\sigma}}\mathrm{dist}^2(\mathbf{0}, \partial\mathcal{T}(\mathbf{u}))$$

On the other hand,

$$P_{k+1}-\mathcal{T}^{\infty,\Lambda} \leq \frac{1}{2\hat{\sigma}}\mathrm{dist}^2(\mathbf{0},\partial\mathcal{T}(\mathbf{u}^{k+1})+\frac{3\kappa}{\beta^2}L_{\nabla\varphi}^2\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2$$
$$\leq \frac{1}{2\hat{\sigma}}c_{\mathcal{T}}^2\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2+\frac{3\kappa}{\beta^2}L_{\nabla\varphi}^2\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2$$
$$=(\frac{c_{\mathcal{T}}^2}{2\hat{\sigma}}+\frac{3\kappa}{\beta^2}L_{\nabla\varphi}^2)\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2$$
$$\leq \frac{(\frac{c_{\mathcal{T}}^2}{2\hat{\sigma}}+\frac{3\kappa}{\beta^2}L_{\nabla\varphi}^2)}{\frac{\frac{3L_{\nabla\varphi}}{16}}{3+\frac{6}{\beta^2}L_{\nabla\varphi}^2}}\frac{\frac{3L_{\nabla\varphi}}{16}}{3+\frac{6}{\beta^2}L_{\nabla\varphi}^2}\|\mathbf{u}^{k+1}-\mathbf{u}^k\|^2$$
$$\leq \frac{(\frac{c_{\mathcal{T}}^2}{2\hat{\sigma}}+\frac{3\kappa}{\beta^2}L_{\nabla\varphi}^2)}{\frac{\frac{3L_{\nabla\varphi}}{16}}{3+\frac{6}{\beta^2}L_{\nabla\varphi}^2}}(P_k-P_{k+1}).$$

The second inequality is due to

$$\mathrm{dist}(\mathbf{0},\partial\mathcal{T}(\mathbf{u}^{k+1})\leq c_{\mathcal{T}}\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2$$

which can be proved similarly to [36, Lemma 5.7] and [37, Lemma 4]. So, the claim holds directly. □

Remark 3 From Theorem 7, the sequence $\{\mathbf{x}^k\}$ generated by Algorithm 2 satisfies $P^k \to F(\mathbf{x}^\infty)$, where $\mathbf{x}^\infty$ denotes the limit point of $\{\mathbf{x}^k\}$. If Algorithm 2 performs a finite number of iterations within the same subspace Π_Λ, then there exists a constant $\widetilde{q}_\Lambda \in (0,1)$ such that

$$P^{k+1}-F(\mathbf{x}^\infty)\leq \widetilde{q}_\Lambda\big(P^k-F(\mathbf{x}^\infty)\big),$$

for all successive iterates $\mathbf{x}^k$ and $\mathbf{x}^{k+1}$ lying in the same subspace Π_Λ.

Remark 4 When Algorithm $\mathcal{A}$ is instantiated as the Alternating Direction Method of Multipliers (ADMM), we define

$$\psi(\mathbf{x}^{k+1},\mathbf{y}^{k+1},\mathbf{z}^{k+1},\mathbf{z}^k):=r(\mathbf{x}^{k+1})+\varphi(\mathbf{x}^{k+1})+\frac{\beta}{2}\|\mathbf{x}^{k+1}-\mathbf{y}^{k+1}\|_2^2+\frac{3\kappa}{\beta^2}\|\mathbf{z}^{k+1}-\mathbf{z}^k\|_2^2.$$

Under this setting, the assumptions (A-i)-(A-v) in the FACI framework are automatically fulfilled as a direct consequence of Theorem 7, inequality (28), Remark 3, Theorem 9 and Theorem 4.

5.3 Accelerated Projection Gradient Method

In the following, we consider using the accelerated projection gradient method to solve the unconstrained model (1), and summarize in Algorithm. 3.

Theorem 10 (Global convergence) *Let $r(\mathbf{x})=\gamma\sigma(\mathbf{x})$, where $\sigma(\mathbf{x})$ is defined in Table 1. Assume that $\varphi:\mathbb{R}^n\to\mathbb{R}$ is an $L_{\nabla\varphi}$-smooth, convex, and subanalytic function. Suppose that the step size α in Algorithm 3 satisfies $\alpha > 2L_{\nabla\varphi}\overline{\beta}^2$ for some $\overline{\beta}\in(0,1)$. If the sequence $\{\mathbf{x}^k\}$ generated by Algorithm 3 is bounded, then it converges to a critical point of F defined in (1). Moreover, the sequence $\{\mathbf{x}^k\}$ satisfies $\sum_{k=0}^{\infty}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|<+\infty$.*

Algorithm 3 Accelerated Projection Gradient Method for solving (1)

Let $\alpha > 0$, and $\mathbf{x}^0$ be a given starting point. Let $\{\beta_k\} \subseteq [0,1)$ with $\overline{\beta} = \sup_k \beta_k < 1$. Set $k = 0$.

1: $\mathbf{y}^k = \mathbf{x}^k + \beta_k(\mathbf{x}^k - \mathbf{x}^{k-1})$.

2: $\mathbf{x}^{k+1} \in \mathrm{Prox}_{r/\alpha}(\mathbf{x}^k - \frac{1}{\alpha}\nabla\varphi(\mathbf{y}^k))$.

3: Set $k = k+1$, and go to Line 1.

Theorem 11 (Finite identification and linear convergence) *Let $r(\mathbf{x}) = \gamma\sigma(\mathbf{x})$, where $\sigma(\mathbf{x})$ is defined in Table 1, and let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be an $L_{\nabla\varphi}$-smooth, subanalytic, and strongly convex function with modulus $\sigma > \overline{\sigma} := \lambda p(1-p)\nu^{-(2-p)} > 0$. Suppose that $\alpha > 2L_{\nabla\varphi}\overline{\beta}^2$ for some $\overline{\beta} \in (0,1)$, and let the sequence $\{\mathbf{x}^k\}$ be generated by Algorithm 3. Assume further that $\{\mathbf{x}^k\}$ is bounded. Then the following statements hold:*

(i) The support set $I(\mathbf{x}^k)$ changes only finitely many times.

(ii) The sequence $\{\delta^k\}$, defined by

$$\delta^k := F(\mathbf{x}^k) + (\hat{c} + L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 - F(\mathbf{x}^{\infty,\Lambda}), \quad \text{where } \hat{c} = \tfrac{1}{2}\big(\tfrac{\alpha}{2} - L_{\nabla\varphi} - L_{\nabla\varphi}\overline{\beta}^2\big),$$

converges linearly to zero when $\mathbf{x}^k$ belongs to the subspace $\Pi_\Lambda = \{\mathbf{x} \mid I(\mathbf{x}) = \Lambda\}$ and $\mathbf{x}^{\infty,\Lambda}$ is defined in (16).

Proof (i) Since the subproblem for $\mathbf{x}^{k+1}$ amounts to solving

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\Pi_\Lambda} \Big[r(\mathbf{x}) + \nabla\varphi(\mathbf{y}^k)^\top(\mathbf{x} - \mathbf{y}^k) + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{x}^k\|^2\Big],$$

we have

$$r(\mathbf{x}^{k+1}) + \nabla\varphi(\mathbf{y}^k)^\top(\mathbf{x}^{k+1} - \mathbf{x}^k) + \frac{\alpha}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \le r(\mathbf{x}^k).$$

By Lemma 1, it holds that

$$\varphi(\mathbf{x}^{k+1}) - \varphi(\mathbf{y}^k) - \frac{L_{\nabla\varphi}}{2}\|\mathbf{x}^{k+1} - \mathbf{y}^k\|^2 \le \langle \mathbf{x}^{k+1} - \mathbf{y}^k, \nabla\varphi(\mathbf{y}^k)\rangle.$$

In addition,

$$\varphi(\mathbf{y}^k) - \varphi(\mathbf{x}^k) \le \langle \mathbf{y}^k - \mathbf{x}^k, \nabla\varphi(\mathbf{y}^k)\rangle.$$

Adding the above three inequalities yields

$$r(\mathbf{x}^{k+1}) + \varphi(\mathbf{x}^{k+1}) - \varphi(\mathbf{x}^k) - \frac{L_{\nabla\varphi}}{2}\|\mathbf{x}^{k+1} - \mathbf{y}^k\|^2 + \frac{\alpha}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \le r(\mathbf{x}^k).$$

Hence, we obtain the descent inequality

$$\textcircled{1}_k: \quad F(\mathbf{x}^{k+1}) + (\hat{c} + L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \le F(\mathbf{x}^k) + (\hat{c} + L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 - \hat{c}\big(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2\big). \quad (29)$$

The remaining proof follows the same reasoning as in the assertion (i) of Theorem 6 and is thus omitted.

(ii) We first show that there exist positive constants τ_1 and τ_2 such that

$$\textcircled{2}_k: \quad \|\mathbf{x}^{k+1} - \mathbf{x}^{\infty,\Lambda}\| \le \tau_1\|\mathbf{x}^{k+1} - \mathbf{x}^k\| + \tau_2\|\mathbf{x}^k - \mathbf{x}^{k-1}\|.$$

Analogously to the proof of (20), we obtain

$$\sum_{i\in\Lambda}\langle x_i^{k+1}-x_i^{\infty,\Lambda},\partial r_i(x_i^{k+1})-\partial r_i(x_i^{\infty,\Lambda})\rangle+\langle \mathbf{x}_\Lambda^{k+1}-\mathbf{x}_\Lambda^{\infty,\Lambda},[\nabla\varphi(\mathbf{x}^{k+1})]_\Lambda-[\nabla\varphi(\mathbf{x}^{\infty,\Lambda})]_\Lambda\rangle$$
$$-\langle \mathbf{x}^{k+1}-\mathbf{x}^{\infty,\Lambda},\nabla\varphi(\mathbf{x}^{k+1})-\nabla\varphi(\mathbf{y}^k)\rangle=\alpha\langle \mathbf{x}^{k+1}-\mathbf{x}^{\infty,\Lambda},\mathbf{x}^k-\mathbf{x}^{k+1}\rangle.$$

On the other hand,

$$\begin{aligned}\tilde\sigma\|\mathbf{x}_\Lambda^{k+1}-\mathbf{x}_\Lambda^{\infty,\Lambda}\|^2 \le &\sum_{i\in\Lambda}\langle x_i^{k+1}-x_i^{\infty,\Lambda},\partial r_i(x_i^{k+1})-\partial r_i(x_i^{\infty,\Lambda})\rangle\\ &+\langle \mathbf{x}_\Lambda^{k+1}-\mathbf{x}_\Lambda^{\infty,\Lambda},[\nabla\varphi(\mathbf{x}^{k+1})]_\Lambda-[\nabla\varphi(\mathbf{x}^{\infty,\Lambda})]_\Lambda\rangle.\end{aligned}$$

Combining the above inequalities gives

$$\tilde\sigma\|\mathbf{x}^{k+1}-\mathbf{x}^{\infty,\Lambda}\|\le\alpha\|\mathbf{x}^{k+1}-\mathbf{x}^k\|+L_{\nabla\varphi}\big(\|\mathbf{x}^{k+1}-\mathbf{x}^k\|+\overline{\beta}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|\big).$$

Therefore, the relation ② $_k$ holds with

$$\tau_1=\frac{L_{\nabla\varphi}+\alpha}{\tilde\sigma},\qquad \tau_2=\frac{L_{\nabla\varphi}\overline{\beta}}{\tilde\sigma}.$$

Similarly to the proof of (21), we have

$$③_k:\quad F(\mathbf{x}^{k+1})-F(\mathbf{x}^{\infty,\Lambda})\le\frac{L_{\nabla F}}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^{\infty,\Lambda}\|^2.$$

Consequently,

$$F(\mathbf{x}^{k+1})-F(\mathbf{x}^{\infty,\Lambda})\le\frac{L_{\nabla F}}{2}\big(2\tau_1^2\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2+2\tau_2^2\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2\big).$$

It follows that

$$\begin{aligned}&F(\mathbf{x}^{k+1})-F(\mathbf{x}^{\infty,\Lambda})+(\hat c+L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2\\ &\le\tilde\tau\big(\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2+\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2\big)\\ &\le\frac{\tilde\tau}{\hat c}\Big(F(\mathbf{x}^k)+(\hat c+L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2-\big[F(\mathbf{x}^{k+1})+(\hat c+L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2\big]\Big),\end{aligned}$$

where $\tilde\tau=\max\big(L_{\nabla F}\tau_1^2+(\hat c+L_{\nabla\varphi}\overline{\beta}^2),L_{\nabla F}\tau_2^2\big)$. Therefore, $\delta^{k+1}\le\frac{c^*}{1+c^*}\delta^k$, where $c^*=\frac{\tilde\tau}{\hat c}$. □

Remark 5 From Theorem 10, the sequence $\{\mathbf{x}^k\}$ generated by Algorithm 3 converges to a critical point $\mathbf{x}^\infty$, and it satisfies

$$F(\mathbf{x}^k)+(\hat c+L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2\longrightarrow F(\mathbf{x}^\infty).$$

If Algorithm 3 performs a finite number of iterations within the subspace Π_Λ, then there exists a constant $q_\Lambda^*\in(0,1)$ such that

$$\begin{aligned}&F(\mathbf{x}^{k+1})+(\hat c+L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2-F(\mathbf{x}^\infty)\\ &\le q_\Lambda^*\Big(F(\mathbf{x}^k)+(\hat c+L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2-F(\mathbf{x}^\infty)\Big),\end{aligned}$$

for all $\mathbf{x}^j$ $(j=k,k+1)$ lying in the same subspace Π_Λ.

Remark 6 When Algorithm $\mathcal{A}$ is instantiated as the accelerated projection gradient method (APG), we define $\psi(\mathbf{x}^{k+1}):=F(\mathbf{x}^k)+(\hat c+L_{\nabla\varphi}\overline{\beta}^2)\|\mathbf{x}^k$. Under this setting, the assumptions (A-i)-(A-v) in the FACI framework are automatically fulfilled as a direct consequence of Theorem 10, inequality (29), Remark 5, Theorem 11 and Theorem 4.

6 Conclusion

We consider the nonsmooth, nonconvex composite program (1), which frequently arises in machine learning applications and is commonly solved using proximal-type algorithms such as the Projection Gradient (PG), Alternating Direction Method of Multipliers (ADMM), and Accelerated Projection Gradient (APG). These algorithms are known to identify the active manifold after a finite number of iterations, assuming nondegeneracy conditions. However, the quantitative characterization of the iteration complexity required to enter the active manifold has primarily been confined to convex instances of the composite model (1).

In this work, we develop a unified analytical framework based on five fundamental assumptions that jointly capture both the properties of the algorithm and the characteristics of the nonconvex, nonsmooth regularizer r. Within this framework, we provide a rigorous proof of the active-set identification complexity and derive explicit bounds on the number of iterations required to identify the active manifold. To demonstrate the generality of our approach, we show that the PG, ADMM, and APG methods all satisfy the assumptions in the framework of **FACI** imposed on the algorithm. Furthermore, we prove that the L_0 norm and the L_p norm, when used as regularizers, satisfy the assumptions on r outlined in **FACI**. As a result, we derive explicit bounds on the number of iterations required to identify the active manifold for these algorithms. Notably, our analysis extends the understanding of manifold identification and complexity analysis beyond the convex setting.

Funding The research of Min Tao has been supported by the Natural Science Foundation of China (No. 12471289). The research of Xiao-Ping Zhang was supported in part by Shenzhen Key Laboratory of Ubiquitous Data Enabling (No. ZDSYS20220527171406015), and by Tsinghua Shenzhen International Graduate School-Shenzhen Pengrui Endowed Professorship Scheme of Shenzhen Pengrui Foundation.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

[1] Francis, B., Rodolph, J., Julien, M., Guillaume, O.: Optimization with Sparsity-Inducing Penalties. Now Publishers, Hanover, MA, USA (2012)

[2] Boţ, R.I., Nguyen, D.K.: The proximal alternating direction method of multipliers in the nonconvex setting: Convergence analysis and rates. Math. Oper. Res. **45**(2), 682–712 (2020)

[3] Blumensath, T., Davies, M.E.: Iterative thresholding for sparse approximations. J. Fourier Anal. Appl. **14**(5-6), 629–654 (2008)

[4] Lu, Z.: Iterative hard thresholding methods for l_0 regularized convex cone programming. Math. Program. **147**(1-2), 125–154 (2014)

[5] Grishchenko, D., Iutzeler, F., Malick, J.: Proximal gradient methods with adaptive subspace sampling. Math. Oper. Res. **46**, 1303–1323 (2021)

[6] Liang, J., Fadili, J., Peyré, G.: Activity identification and local linear convergence of forward–backward-type methods. SIAM J. Optim. **27**(1), 408–437 (2017)

[7] Chen, X., Ge, D., Wang, Z., Ye, Y.: Complexity of unconstrained L_2-L_p minimization. Math. Program. **143**(1-2), 371–383 (2014)

[8] Parikh, N., Boyd, S.P.: Proximal algorithms. Found. Trends Optim. **1**, 127–239 (2013)

[9] Bertsekas, D.P.: On the Goldstein-Levitin-Polyak gradient projection method. IEEE Trans. Autom. Control **21**(2), 174–184 (1976)

[10] Bertsekas, D.P.: Projected newton methods for optimization problems with simple constraints. SIAM J. Control. Optim. **20**(2), 221–246 (1982)

[11] Sun, Y., Jeong, H., Nutini, J., Schmidt, M.: Are we there yet? manifold identification of gradient-related proximal methods. In: Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 89, pp. 1110–1119. PMLR, Okinawa (2019)

[12] Lee, C.P.: Accelerating inexact successive quadratic approximation for regularized optimization through manifold identification. Math. Program. **13**, 645–655 (2023)

[13] Hare, W., Lewis, A.: Identifying active constraints via partial smoothness and prox-regularity. J. Convex Anal. **11**, 251–266 (2004)

[14] Bareilles, G., Iutzeler, F., Malick, J.: Newton acceleration on manifolds identified by proximal gradient methods. Math. Program. **200**, 37–70 (2023)

[15] Vaiter, S., Golbabaee, M., Fadili, J., Peyré, G.: Model selection with low complexity priors. Inf. Inference **4**(3), 230–287 (2015)

[16] Liang, J., Fadili, J., Peyré, G.: Local linear convergence of forward–backward under partial smoothness. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (eds.) Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2. NIPS'14, vol. 27, pp. 1970–1978. MIT Press, Cambridge, MA, USA (2014)

[17] Lewis, A.S.: Active sets, nonsmoothness, and sensitivity. SIAM J. Optim. **13**(3), 702–725 (2002)

[18] Burke, J.: On the identification of active constraints ii: The nonconvex case. SIAM J. Numer. Anal. **27**(4), 1081–1102 (1990)

[19] Lemaréchal, C., Oustry, F., Sagastizábal, C.: The U-lagrangian of a convex function. Trans. Amer. Math. Soc. **352**, 711–729 (2000)

[20] Daniilidis, A., Drusvyatskiy, D., Lewis, A.S.: Orthogonal invariance and identifiability. SIAM J. Matrix Anal. Appl. **35**(2), 580–598 (2014)

[21] Burke, J.V., More, J.J.: On the identification of active constraints. SIAM J. Numer. Anal. **25**(5), 1197–1211 (1988)

[22] Burke, J.V.: On identification of active constraints II: the nonconvex case. SIAM J. Numer. Anal. **27**, 1081–1102 (1990)

[23] Wright, S.J.: Identifiable surfaces in constrained optimization. SIAM J. Control. Optim. **31**(4), 1063–1079 (1993)

[24] Tao, M., Zhang, X.-P., Xia, Z.-H.: On partial smoothness, activity identification and faster algorithms of l_1 over l_2 minimization. IEEE Trans. Signal Process. **72**, 2874–2889 (2024)

[25] Nutini, J., Schmidt, M., Hare, W.: "active-set complexity" of proximal gradient: How long does it take to find the sparsity pattern? Optim. Lett. **13**, 645–655 (2019)

[26] Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Springer, Media New York (2013)

[27] Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Amer. Math. Soc. **82**(2), 421–439 (1956)

[28] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)

[29] Johnson, R., Zhang, T.: Accelerating stochastic gradient descent using predictive variance reduction. In: Burges, C.J., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems. NIPS'13, vol. 26, pp. 315–323. Curran Associates Inc., Red Hook, NY, USA (2013)

[30] Nutini, J., Laradji, I., Schmidt, M.: Let's make block coordinate descent converge faster: Faster greedy rules, message-passing, active-set complexity, and superlinear convergence. J. Mach. Learn. Res. **23**(131), 1–74 (2022)

[31] Marjanovic, G., Solo, V.: On l_q optimization and matrix completion. IEEE Trans.

Signal Process **60**(11), 5714–5724 (2012)

[32] Chen, X., Xu, F., Ye, Y.: Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. SIAM J. Sci. Comput. **32**(5), 2832–2852 (2010)

[33] Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Berlin, Heidelberg (1998)

[34] Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. **146**, 459–494 (2014)

[35] Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2007)

[36] Tao, M.: Minimization of L_1 over L_2 for sparse signal recovery with convergence guarantee. SIAM J. Sci. Comput. **44**(2), 770–797 (2022)

[37] Tao, M., Zhang, X.-P.: Study on L_1 over L_2 minimization for nonnegative signal recovery. J. Sci. Comp. **95** (2023)