

Article

$O(1/k^2)$ Complexity of MABCD/iMABCD Algorithm for Two-Block Composite Convex Optimization Problem

Zixuan Chen ¹, Hongxi Chen ² and Xiaoliang Song ^{2,*}

¹ College of Sciences, Northeastern University, Shenyang 110819, China; chenzixuan@mail.neu.edu.cn

² School of Mathematical Sciences, Dalian University of Technology, Dalian 116025, China; chen hongxi@mail.dlut.edu.cn

* Correspondence: songxiaoliang@dlut.edu.cn

Abstract

This paper focuses on the two-block composite convex optimization problems prevalent in signal recovery, image processing and related fields. Although Alternating Block Coordinate Descent (ABCD) is simple to implement and low in computational cost, its objective function often oscillates during iterations. We propose monotone ABCD (MABCD) to guarantee the monotonic descent of the objective. On this basis, an inexact variant iMABCD is presented, which allows inexact subproblem solutions to lower computational overhead and adaptation to realistic application scenarios. Rigorous theoretical analysis of numerical experiments demonstrates that both algorithms attain the $O(1/k^2)$ convergence rate. The monotonic constraint effectively removes undesirable iterative fluctuations and stabilizes the optimization process, while the inexact updating rule greatly improves practical applicability.

Keywords: ABCD algorithm; monotone strategy; inexact solution; complexity analysis

MSC: 65K05; 90C06; 90C25

1. Introduction

Large-scale composite convex optimization problems with two-block separable nonsmooth structures are ubiquitous in modern signal recovery, image processing, statistical learning, and intelligent sensing systems [1–3]. The standard formulation of such problems can be expressed as:

$$\min_{w=(u,v) \in \mathbb{U} \times \mathbb{V}} \theta(w) = h(w) + p(u) + q(v) \quad (1)$$

where $\mathbb{U}$ and $\mathbb{V}$ are finite-dimensional Euclidean spaces. The smooth coupling term $h : \mathbb{U} \times \mathbb{V} \rightarrow \mathbb{R}$ is assumed to have a globally Lipschitz continuous gradient, i.e., there exists $L > 0$ such that

$$\|\nabla h(w_1) - \nabla h(w_2)\| \leq L\|w_1 - w_2\|, \quad \forall w_1, w_2 \in \mathbb{U} \times \mathbb{V},$$

where $\|\cdot\|$ denotes the Euclidean norm. The nonsmooth functions $p : \mathbb{U} \rightarrow \mathbb{R} \cup \{+\infty\}$ and $q : \mathbb{V} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper closed convex (i.e., lower semicontinuous and not identically $+\infty$). Due to the inherent block-separable characteristic of nonsmooth terms, block coordinate splitting optimization has become the mainstream solution framework for this class of problems, which avoids the high computational cost of joint variable optimization and adapts well to large-scale data scenarios [4].



Academic Editor: Wei-Shih Du

Received: 14 June 2026

Revised: 13 July 2026

Accepted: 19 July 2026

Published: 20 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

As a classic representative of block-wise optimization methods, the Alternating Block Coordinate Descent (ABCD) algorithm [5] has attracted extensive attention in recent years for its simple iterative framework, low per-iteration computational overhead, and excellent adaptability to two-block-structured optimization problems. Different from traditional alternating direction method of multipliers (ADMM), which relies on auxiliary variables and penalty parameter tuning [2], and forward-backward splitting (FBS), which suffers from restricted Lipschitz constant constraints [6], ABCD implements alternating iterative updates of block variables u and v by combining gradient descent for smooth coupling terms and proximal mapping for nonsmooth regularization terms [7]. This alternating update mechanism effectively decouples the complex coupled optimization problem and significantly reduces the difficulty of subproblem solving, leading to ABCD being widely applied in matrix factorization, image reconstruction, and regularized regression tasks [8].

Despite the practical advantages of standard ABCD, it still exhibits defects in iterative stability and objective function monotonicity, which limit its optimization performance and convergence robustness. In the original ABCD iterative framework, the update rules for the block variables only focus on the gradient descent optimization of the current iteration, lacking effective constraint mechanisms for the overall objective function [9]. Affected by ill-conditioned matrix distribution, uneven Lipschitz constants for different blocks, and complex nonsmooth term superposition, the objective function value of standard ABCD frequently fluctuates and rises during the iterative process [10]. This non-monotonic iteration phenomenon not only slows the practical convergence speed and increases the number of redundant iterations but also easily causes the algorithm to fall into local suboptimal solutions.

Recent years have witnessed growing interest in block coordinate descent (BCD) and inexact proximal methods for large-scale composite convex optimization. A number of accelerated BCD variants have been developed to achieve the optimal $\mathcal{O}(1/k^2)$ convergence rate. Notably, the inexact majorized accelerated block coordinate descent (imABCD) framework proposed in [11] establishes $\mathcal{O}(1/k^2)$ complexity while allowing flexible inexact subproblem solutions via adjustable error tolerances. Unlike methods that enforce monotonic objective decrease through line search [3,12], this accelerated framework prioritizes convergence rate guarantees over monotonicity, reflecting a deliberate algorithmic trade-off. Meanwhile, monotone BCD approaches typically employ Armijo-type or backtracking line search to ensure nonincreasing objectives [3,12], which may introduce additional computational overhead per iteration. The integration of acceleration mechanisms with monotonicity safeguards, as well as the design of parameter-free monotonic selection rules, remains an active area of investigation [13,14].

To address the monotonicity deficiency of the original ABCD algorithm and further enhance its iterative stability, this paper proposes a monotone ABCD (MABCD) optimization algorithm. In the full-gradient setting, ref. [15] introduced a monotone version of FISTA via a simple two-point comparison that ensures nonincreasing objective values without line search. In this work, we extend this monotone correction to the two-block ABCD framework. By embedding monotone judgment, the algorithm strictly restricts the update direction of block variables and ensures that the overall loss function maintains a nonincreasing trend in the entire iterative process. Unlike existing ABCD variants that either accept non-monotone sequences or rely on costly line search procedures, this improvement completely solves the objective function fluctuation problem of the original ABCD algorithm and retains the advantages of low computational complexity and simple deployment of the original framework, without introducing additional complex iterative operations.

On the basis of the proposed MABCD framework, we further develop an inexact monotone ABCD (iMABCD) algorithm to improve practical applicability. Different from

MABCD, which requires exact solutions for all subproblems in each iteration, the improved inexact version allows the subproblems to be solved within a reasonable tolerance. This inexact updating strategy possesses significant practical advantages for large-scale optimization tasks. First, it greatly reduces the computational overhead per iteration, since strict exact subproblem solving usually involves expensive operations. Second, the inexact scheme enhances the flexibility and robustness of the algorithm in real-world scenarios, where exact analytical solutions are often unavailable or computationally prohibitive. Benefiting from the combination of strict monotonic descent guarantee and tolerance-based inexact updating, the proposed iMABCD achieves a better trade-off between computational efficiency and optimization accuracy, making it more suitable for real-time and large-scale engineering applications.

Under standard assumptions, this paper proves the $O(1/k^2)$ complexity of the proposed MABCD and iMABCD algorithms. Theoretical analysis verifies that the introduced monotone skill does not compromise the convergence accuracy of the original algorithm. On the contrary, it provably refines the iterative optimization path, leading to more stable and monotonic progress toward the optimum without altering the asymptotic precision. In particular, the monotone skill eliminates undesirable oscillations or temporary increases in the objective value, ensuring that every iteration yields a nonincreased cost while preserving $O(1/k^2)$ complexity.

The remainder of this paper is organized as follows: Section 2 presents a detailed review of the original ABCD algorithm framework and its existing limitations. Then, the specific iterative update rules of the proposed MABCD algorithm and the theoretical convergence analysis are provided. In Section 3, we introduce inexactness and give the framework of the iMABCD algorithm and its convergence analysis. Section 4 summarizes the work and outlines prospects the future research directions.

Notation Overview

For the convenience of readers, we summarize all core mathematical notations used in the subsequent algorithms and theoretical analysis in Table 1.

Table 1. Notation used in the algorithm and its analysis (operators, matrices, norms, and iterates).

Symbol	Description
$\mathcal{Q}$	Self-adjoint positive semidefinite linear operator satisfying the lower bound in (5); characterizes the curvature lower bound of h .
$\hat{\mathcal{Q}}$	Self-adjoint positive semidefinite linear operator satisfying the upper bound in (5); used to construct the majorization upper model of h .
$\mathcal{Q}_{11}, \mathcal{Q}_{22}, \mathcal{Q}_{12}$	2×2 block decomposition of $\mathcal{Q}$; $\mathcal{Q}_{11} : \mathbb{U} \rightarrow \mathbb{U}$, $\mathcal{Q}_{22} : \mathbb{V} \rightarrow \mathbb{V}$, $\mathcal{Q}_{12} : \mathbb{V} \rightarrow \mathbb{U}$.
$\hat{\mathcal{Q}}_{11}, \hat{\mathcal{Q}}_{22}, \hat{\mathcal{Q}}_{12}$	2×2 block decomposition of $\hat{\mathcal{Q}}$ with analogous mapping properties.
$\mathcal{D}_1 : \mathbb{U} \rightarrow \mathbb{U}$ $\mathcal{D}_2 : \mathbb{V} \rightarrow \mathbb{V}$	Positive definite operators such that $\hat{\mathcal{Q}} = \mathcal{Q} + \text{Diag}(\mathcal{D}_1, \mathcal{D}_2)$. (Assumption 1)
$\mathcal{H}$	Key positive semidefinite operator used in the complexity analysis; $\mathcal{H} \triangleq \text{Diag}(\mathcal{D}_1, \mathcal{D}_2 + \mathcal{Q}_{22})$.
$\hat{h}(w; \bar{w})$	Majorization quadratic upper bound of h at $\bar{w}$; defined in (5).
$w^{k+1} = (u^{k+1}, v^{k+1})$	Extrapolated point at iteration k for Algorithms 1–3 (after Nesterov acceleration).
$\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k)$	Computed solution of the subproblems at iteration k for Algorithm 1.
$\tilde{z}^k = (\tilde{x}^k, \tilde{y}^k)$	Computed solution of the subproblems at iteration k for Algorithm 2 or inexactly computed solution of the subproblems at iteration k for Algorithm 3.

Table 1. Cont.

Symbol	Description
$\tilde{w}^k = (\tilde{u}^k, \tilde{v}^k)$	Selected solution to the monotone step at iteration k for Algorithm 2 or 3.
$\hat{z}^k = (\hat{u}^k, \hat{v}^k)$	Exactly computed auxiliary solution of the subproblems for analysis; defined in (21)
ϵ_k	Non-negative nonincreasing error tolerance for the inexact subproblem solutions; satisfies $\sum i\epsilon_i < \infty$.
δ_u^k, δ_v^k	Residual errors incurred when solving the u and v subproblems inexactly; they satisfy $\ \hat{Q}_{11}^{-1/2}\delta_u^k\ \leq \epsilon_k$ and $\ \hat{Q}_{22}^{-1/2}\delta_v^k\ \leq \epsilon_k$.

Algorithm 1 ABCD: Accelerated Block Coordinate Descent algorithm for problem (1)

- 1: Select an initial point $(u^1, v^1) = (\tilde{u}^0, \tilde{v}^0) \in \text{dom}(p) \times \text{dom}(q)$, and take $t_1 = 1$. Then, for each iteration $k \geq 1$, the following steps are carried out.
- 2: Compute the exact block minimizers:

$$\begin{cases} \tilde{x}^k = \arg \min_{u \in \mathbb{U}} \{p(u) + \hat{h}(u, v^k; w^k)\}, \\ \tilde{y}^k = \arg \min_{v \in \mathbb{V}} \{q(v) + \hat{h}(\tilde{x}^k, v; w^k)\}. \end{cases} \quad (2)$$

- 3: Denote $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k)$.
- 4: Compute acceleration parameters and extrapolated iterate:

$$\begin{cases} t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ w^{k+1} = \tilde{w}^k + \frac{t_k - 1}{t_{k+1}} (\tilde{w}^k - \tilde{w}^{k-1}). \end{cases}$$

Algorithm 2 MABCD: Monotone Accelerated Block Coordinate Descent algorithm for problem (1)

- 1: Select an initial point $(u^1, v^1) = (\tilde{u}^0, \tilde{v}^0) \in \text{dom}(p) \times \text{dom}(q)$, and take $t_1 = 1$. Then, for each iteration $k \geq 1$, the following steps are carried out.
- 2: Compute the exact block minimizers:

$$\begin{cases} \tilde{x}^k = \arg \min_{u \in \mathbb{U}} \{p(u) + \hat{h}(u, v^k; w^k)\}, \\ \tilde{y}^k = \arg \min_{v \in \mathbb{V}} \{q(v) + \hat{h}(\tilde{x}^k, v; w^k)\}. \end{cases} \quad (3)$$

- 3: Denote $\tilde{z}^k = (\tilde{x}^k, \tilde{y}^k)$.
- 4: Monotone constraint selection step:

$$\tilde{w}^k = (\tilde{u}^k, \tilde{v}^k) = \arg \min \left\{ \theta(w) \mid w \in \{(\tilde{x}^k, \tilde{y}^k), (\tilde{u}^{k-1}, \tilde{v}^{k-1})\} \right\}.$$

- 5: Compute acceleration parameters and extrapolated iterate:

$$\begin{cases} t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ w^{k+1} = \tilde{w}^k + \frac{t_k - 1}{t_{k+1}} (\tilde{w}^k - \tilde{w}^{k-1}) + \frac{t_k}{t_{k+1}} (\tilde{z}^k - \tilde{w}^k). \end{cases}$$

Algorithm 3 iMABCD: an inexact Monotone Accelerated Block Coordinate Descent algorithm for problem (1)

- 1: Select an initial point $(u^1, v^1) = (\tilde{u}^0, \tilde{v}^0) \in \text{dom}(p) \times \text{dom}(q)$, a non-negative, nonincreasing sequence of tolerances $\{\varepsilon_k\}$, and take $t_1 = 1$. Then, for each iteration $k \geq 1$, the following steps are carried out.
- 2: Compute inexact block subproblem solutions:

$$\begin{cases} \tilde{x}^k = \arg \min_{u \in \mathbb{U}} \{p(u) + \hat{h}(u, v^k; w^k) + \langle \delta_u^k, u \rangle\}, \\ \tilde{y}^k = \arg \min_{v \in \mathbb{V}} \{q(v) + \hat{h}(\tilde{x}^k, v; w^k) + \langle \delta_v^k, v \rangle\}, \end{cases} \quad (4)$$

where the error pair $(\delta_u^k, \delta_v^k) \in \mathbb{U} \times \mathbb{V}$ satisfies

$$\max \left\{ \|\hat{Q}_{11}^{-1/2} \delta_u^k\|, \|\hat{Q}_{22}^{-1/2} \delta_v^k\| \right\} \leq \varepsilon_k.$$

- 3: Denote $\tilde{z}^k = (\tilde{x}^k, \tilde{y}^k)$.
- 4: Monotone constraint discrete selection step:

$$\tilde{w}^k = (\tilde{u}^k, \tilde{v}^k) = \arg \min \left\{ \theta(w) \mid w \in \{(\tilde{x}^k, \tilde{y}^k), (\tilde{u}^{k-1}, \tilde{v}^{k-1})\} \right\}.$$

- 5: Compute acceleration parameters and extrapolated iterate:

$$\begin{cases} t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ w^{k+1} = \tilde{w}^k + \frac{t_k - 1}{t_{k+1}} (\tilde{w}^k - \tilde{w}^{k-1}) + \frac{t_k}{t_{k+1}} (\tilde{z}^k - \tilde{w}^k). \end{cases}$$

2. Monotone ABCD

In this section, we first present a detailed review of the original ABCD algorithm framework. Subsequently, we introduce the monotone ABCD (MABCD) method, in which all subproblems are solved exactly, and we prove that this algorithm attains a convergence rate of $O(1/k^2)$.

As previously indicated, the objective of this work is to solve the unconstrained composite convex problem (1). In this setting, $p(u)$ and $q(v)$ are proper closed convex functions, while the gradient ∇h is globally Lipschitz. To enable efficient block-coordinate updates, we introduce two self-adjoint positive semidefinite operators $\mathcal{Q}$ and $\hat{\mathcal{Q}} : \mathbb{U} \times \mathbb{V} \rightarrow \mathbb{U} \times \mathbb{V}$, which provide global quadratic lower and upper bounds for h . Intuitively, they act as curvature matrices controlling the step sizes; for a quadratic h , they reduce to its Hessian, and, in general, they can be chosen via the Lipschitz constant of ∇h . Such bounds are standard in majorization–minimization schemes and help decouple the block variables while guaranteeing convergence [11]. Consequently, for all $w, \bar{w} \in \mathbb{U} \times \mathbb{V}$, the following inequalities hold:

$$\begin{cases} h(w) \geq h(\bar{w}) + \langle \nabla h(\bar{w}), w - \bar{w} \rangle + \frac{1}{2} \|w - \bar{w}\|_{\mathcal{Q}}^2, \\ h(w) \leq \hat{h}(w; \bar{w}) := h(\bar{w}) + \langle \nabla h(\bar{w}), w - \bar{w} \rangle + \frac{1}{2} \|w - \bar{w}\|_{\hat{\mathcal{Q}}}^2, \end{cases} \quad (5)$$

where $\|\cdot\|_{\mathcal{Q}}^2 = \langle \cdot, \mathcal{Q} \cdot \rangle$ and similarly for $\hat{\mathcal{Q}}$.

The 2×2 block decomposition of the operators $\mathcal{Q}$ and $\hat{\mathcal{Q}}$ is as follows:

$$\mathcal{Q}w := \begin{pmatrix} \mathcal{Q}_{11} & \mathcal{Q}_{12} \\ \mathcal{Q}_{21}^* & \mathcal{Q}_{22} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}, \quad \hat{\mathcal{Q}}w := \begin{pmatrix} \hat{\mathcal{Q}}_{11} & \hat{\mathcal{Q}}_{12} \\ \hat{\mathcal{Q}}_{21}^* & \hat{\mathcal{Q}}_{22} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}, \quad \forall w = (u, v) \in \mathbb{U} \times \mathbb{V}.$$

Here, $\mathcal{Q}_{11}, \hat{\mathcal{Q}}_{11} : \mathbb{U} \rightarrow \mathbb{U}$ and $\mathcal{Q}_{22}, \hat{\mathcal{Q}}_{22} : \mathbb{V} \rightarrow \mathbb{V}$ are self-adjoint positive semidefinite linear operators; moreover, $\mathcal{Q}_{12}, \hat{\mathcal{Q}}_{12} : \mathbb{V} \rightarrow \mathbb{U}$ are linear, and their adjoints are $\mathcal{Q}_{12}^*$ and $\hat{\mathcal{Q}}_{12}^*$, respectively. In what follows, we impose the following assumption.

Assumption 1 ([11]). *There exist two self-adjoint positive semidefinite linear operators $\mathcal{D}_1 : \mathbb{U} \rightarrow \mathbb{U}$ and $\mathcal{D}_2 : \mathbb{V} \rightarrow \mathbb{V}$ such that*

$$\hat{\mathcal{Q}} := \mathcal{Q} + \text{Diag}(\mathcal{D}_1, \mathcal{D}_2). \quad (6)$$

Furthermore, $\hat{\mathcal{Q}}$ satisfies $\hat{\mathcal{Q}}_{11} \succ 0$ and $\hat{\mathcal{Q}}_{22} \succ 0$.

Below is the framework of the ABCD algorithm for solving problem (1).

Then, we introduce a strict monotone constraint mechanism [15] on the original update of the ABCD algorithm and propose the following monotone Accelerated Block Coordinate Descent (MABCD) algorithm to solve problem (1). The minimization operation over w is a discrete two-point comparison rule: we evaluate objective $\theta(\cdot)$ at two candidate points—the newly updated block minimizer $\tilde{z}^k = (\tilde{x}^k, \tilde{y}^k)$ and the previous iterate $\tilde{w}^{k-1}$. We select the candidate with a smaller objective value as $\tilde{w}^k$, which strictly guarantees a nonincreasing sequence $\{\theta(\tilde{w}^k)\}$ and provides the monotonicity property of MABCD.

In order to establish the $O(\frac{1}{k^2})$ complexity rate for the objective values, we define a positive semidefinite operator $\mathcal{H} : \mathbb{U} \times \mathbb{V} \rightarrow \mathbb{U} \times \mathbb{V}$ as follows:

$$\mathcal{H} := \text{Diag}(\mathcal{D}_1, \mathcal{D}_2 + \mathcal{Q}_{22}). \quad (7)$$

The following lemma holds for the sequence $\{t_k\}_{k \geq 1}$ generated by the MABCD algorithm.

Lemma 1 ([11]). *The following properties hold for the sequence $\{t_k\}_{k \geq 1}$ produced by the MABCD algorithm:*

- (a) $1 - \frac{1}{t_{k+1}} = \frac{t_k^2}{t_{k+1}^2}.$
- (b) $\frac{k+1}{2} \leq t_k \leq \frac{5}{8}k + \frac{3}{8} \leq k.$
- (c) *Given any $\tilde{w}, \bar{w} \in \mathbb{U} \times \mathbb{V}$. Consider $w = (1 - \frac{1}{t_k})\tilde{w} + \frac{1}{t_k}\bar{w}$. Then,*

$$t_k^2[\theta(w) - \theta(\bar{w})] \leq t_{k-1}^2[\theta(\tilde{w}) - \theta(\bar{w})].$$

In the following, we establish the $O(1/k^2)$ complexity of the MABCD algorithm when its subproblems are solved exactly. To this end, we first present a lemma that highlights a crucial property of the objective values along the iterates generated by the algorithm.

Lemma 2. *Under Assumption 1, let sequences $\{\tilde{z}^k := (\tilde{x}^k, \tilde{y}^k)\}$ and $\{\tilde{w}^k := (\tilde{u}^k, \tilde{v}^k)\}$ be produced by the MABCD algorithm. Then, for any $k \geq 1$, we have*

$$\theta(\tilde{w}^k) - \theta(w) \leq \frac{1}{2}\|w - w^k\|_{\mathcal{H}}^2 - \frac{1}{2}\|w - \tilde{z}^k\|_{\mathcal{H}}^2, \quad \forall w \in \mathbb{U} \times \mathbb{V}.$$

Specifically, when w^* is an optimal solution to (1), then

$$\|w^* - \tilde{w}^k\|_{\mathcal{H}} \leq \|w^* - \tilde{z}^k\|_{\mathcal{H}} \leq \|w^* - w^k\|_{\mathcal{H}}.$$

Proof. Invoking the optimality condition for the subproblems in (1), we obtain

$$\begin{cases} 0 \in \partial p(\tilde{x}^k) + \nabla_u h(w^k) + \hat{Q}_{11}(\tilde{x}^k - u^k), \\ 0 \in \partial q(\tilde{y}^k) + \nabla_v h(w^k) + Q_{12}^*(\tilde{x}^k - u^k) + \hat{Q}_{22}(\tilde{y}^k - v^k). \end{cases}$$

Consequently, from the convexity of $p(\cdot)$ and $q(\cdot)$, it follows that

$$\begin{cases} p(u) \geq p(\tilde{x}^k) + \langle u - \tilde{x}^k, -\nabla_u h(w^k) - \hat{Q}_{11}(\tilde{x}^k - u^k) \rangle, \quad \forall u \in \mathbb{U}, \\ q(v) \geq q(\tilde{y}^k) + \langle v - \tilde{y}^k, -\nabla_v h(w^k) - Q_{12}^*(\tilde{x}^k - u^k) - \hat{Q}_{22}(\tilde{y}^k - v^k) \rangle, \quad \forall v \in \mathbb{V}. \end{cases} \quad (8)$$

Utilizing the inequalities stated in (5), we arrive at

$$h(w) - h(\tilde{z}^k) \geq \langle \nabla h(w^k), w - \tilde{z}^k \rangle + \frac{1}{2} \|w - w^k\|_{\hat{Q}}^2 - \frac{1}{2} \|\tilde{z}^k - w^k\|_{\hat{Q}}^2. \quad (9)$$

Summing (8) and (9) results in

$$\begin{aligned} & 2(\theta(w) - \theta(\tilde{z}^k)) \\ &= \left(\|w - w^k\|_{\hat{Q}}^2 - \|\tilde{z}^k - w^k\|_{\hat{Q}}^2 \right) - \left(\|w - w^k\|_{\hat{Q}}^2 - \|w - \tilde{z}^k\|_{\hat{Q}}^2 - \|\tilde{z}^k - w^k\|_{\hat{Q}}^2 \right) \\ & \quad - 2\langle \tilde{x}^k - u, Q_{12}(\tilde{y}^k - v^k) \rangle. \end{aligned} \quad (10)$$

We can prove that the following inequality holds through the Cauchy–Schwarz inequality

$$\begin{aligned} & 2\langle \tilde{x}^k - u, Q_{12}(\tilde{y}^k - v^k) \rangle \\ &= 2\langle Q(\tilde{z}^k - w), \begin{pmatrix} 0 \\ \tilde{y}^k - v^k \end{pmatrix} \rangle - 2\langle Q_{22}(\tilde{y}^k - v), \tilde{y}^k - v^k \rangle, \\ &\leq \left(\|\tilde{z}^k - w\|_{\hat{Q}}^2 + \|\tilde{y}^k - v^k\|_{Q_{22}}^2 \right) - \left(\|\tilde{y}^k - v\|_{Q_{22}}^2 + \|\tilde{y}^k - v^k\|_{Q_{22}}^2 - \|v^k - v\|_{Q_{22}}^2 \right), \\ &= \|\tilde{z}^k - w\|_{\hat{Q}}^2 + \|v^k - v\|_{Q_{22}}^2 - \|\tilde{y}^k - v\|_{Q_{22}}^2. \end{aligned} \quad (11)$$

Inserting the inequality above into (10), we derive

$$\begin{aligned} & 2(\theta(w) - \theta(\tilde{z}^k)) \\ &\geq \left(\|w - w^k\|_{\hat{Q}}^2 - \|\tilde{z}^k - w^k\|_{\hat{Q}}^2 \right) - \left(\|w - w^k\|_{\hat{Q}}^2 - \|w - \tilde{z}^k\|_{\hat{Q}}^2 - \|\tilde{z}^k - w^k\|_{\hat{Q}}^2 \right) \\ & \quad - \left(\|\tilde{z}^k - w\|_{\hat{Q}}^2 + \|v^k - v\|_{Q_{22}}^2 - \|\tilde{y}^k - v\|_{Q_{22}}^2 \right), \\ &= \|w - \tilde{z}^k\|_{\mathcal{H}}^2 - \|w - w^k\|_{\mathcal{H}}^2, \end{aligned} \quad (12)$$

where the last inequality is due to Assumption 1. Then,

$$\theta(\tilde{w}^k) - \theta(w) \leq \theta(\tilde{z}^k) - \theta(w) \leq \frac{1}{2} \|w - w^k\|_{\mathcal{H}}^2 - \frac{1}{2} \|w - \tilde{z}^k\|_{\mathcal{H}}^2. \quad (13)$$

The stated inequality therefore holds. $\square$

Building on the preceding lemma, we now establish the $O(1/k^2)$ complexity of the monotone ABCD algorithm.

Theorem 1. Assume that Assumption 1 holds and that the solution set Ω of problem (1) is nonempty. Take $w^* = (u^*, v^*) \in \Omega$. Then, for the sequence $\{\tilde{w}^k\} = \{(\tilde{u}^k, \tilde{v}^k)\}$ produced by the monotone ABCD algorithm, the following estimate holds for every $k \geq 1$:

$$\theta(\tilde{w}^k) - \theta(w^*) \leq \frac{2\|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2}{(k+1)^2}.$$

Proof. Substituting $w = (1 - \frac{1}{t_k})\tilde{w}^{k-1} + \frac{1}{t_k}w^*$ into Lemma 2 yields, for every $k \geq 2$,

$$t_k^2\theta(\tilde{z}^k) - t_k^2\theta(w) \leq \frac{1}{2}\|(t_k - 1)\tilde{w}^{k-1} + w^* - t_k w^k\|_{\mathcal{H}}^2 - \frac{1}{2}\|(t_k - 1)\tilde{w}^{k-1} + w^* - t_k \tilde{z}^k\|_{\mathcal{H}}^2. \quad (14)$$

Letting $\tilde{w} = \tilde{w}^{k-1}$ and $\tilde{w} = w^*$ in Lemma 1 (c), we derive

$$t_k^2[\theta(w) - \theta(w^*)] \leq t_{k-1}^2[\theta(\tilde{w}^{k-1}) - \theta(w^*)]. \quad (15)$$

Using the fact that

$$w^k = \tilde{w}^{k-1} + \frac{t_{k-1} - 1}{t_k}(\tilde{w}^{k-1} - \tilde{w}^{k-2}) + \frac{t_{k-1}}{t_k}(\tilde{z}^{k-1} - \tilde{w}^{k-1}), \quad (16)$$

we obtain

$$(t_k - 1)\tilde{w}^{k-1} - t_k w^k = (t_{k-1} - 1)\tilde{w}^{k-2} - t_{k-1}\tilde{z}^{k-1}. \quad (17)$$

Summing (14) and (15), for $k \geq 2$, we derive

$$\begin{aligned} & t_k^2[\theta(\tilde{w}^k) - \theta(w^*)] + \frac{1}{2}\|t_k \tilde{z}^k - w^* - (t_k - 1)\tilde{w}^{k-1}\|_{\mathcal{H}}^2 \\ & \leq t_{k-1}^2[\theta(\tilde{w}^{k-1}) - \theta(w^*)] + \frac{1}{2}\|(t_k - 1)\tilde{w}^{k-1} + w^* - t_k w^k\|_{\mathcal{H}}^2, \\ & = t_{k-1}^2[\theta(\tilde{z}^{k-1}) - \theta(w^*)] + \frac{1}{2}\|t_{k-1}\tilde{z}^{k-1} - w^* - (t_{k-1} - 1)\tilde{w}^{k-2}\|_{\mathcal{H}}^2. \end{aligned} \quad (18)$$

Applying Lemma 2 with $k = 1$ and $w = w^*$, we get

$$\theta(\tilde{w}^1) - \theta(w^*) \leq \frac{1}{2}\|w^1 - w^*\|_{\mathcal{H}}^2 - \frac{1}{2}\|\tilde{z}^1 - w^*\|_{\mathcal{H}}^2 = \frac{1}{2}\|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2 - \frac{1}{2}\|\tilde{z}^1 - w^*\|_{\mathcal{H}}^2.$$

Therefore, for any $k \geq 1$,

$$\begin{aligned} & t_k^2[\theta(\tilde{w}^k) - \theta(w^*)] + \frac{1}{2}\|t_k \tilde{z}^k - w^* - (t_k - 1)\tilde{w}^{k-1}\|_{\mathcal{H}}^2 \\ & \leq t_{k-1}^2[\theta(\tilde{w}^{k-1}) - \theta(w^*)] + \frac{1}{2}\|t_{k-1}\tilde{z}^{k-1} - w^* - (t_{k-1} - 1)\tilde{w}^{k-2}\|_{\mathcal{H}}^2, \\ & \leq \dots, \\ & \leq t_1^2[\theta(\tilde{w}^1) - \theta(w^*)] + \frac{1}{2}\|t_1 \tilde{z}^1 - w^* - (t_1 - 1)\tilde{w}^0\|_{\mathcal{H}}^2, \\ & \leq \frac{1}{2}\|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2. \end{aligned} \quad (19)$$

From Lemma 1 (b), we know that $t_k \geq \frac{k+1}{2}$, and, since $\mathcal{H}$ is a positive semidefinite operator, we derive

$$\theta(\tilde{w}^k) - \theta(w^*) \leq \frac{\|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2}{2t_k^2} \leq \frac{2\|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2}{(k+1)^2}. \quad (20)$$

□

3. Inexact Monotone ABCD

In this section, we turn to the scenario where the subproblems are solved inexactly. The notion of incorporating inexactness has been employed in various forms of the block coordinate descent (BCD) and accelerated proximal gradient (APG) methods [5,16–18], and it plays a crucial role in the efficient solution to multi-block problems. The minimization operation over w is a discrete two-point comparison rule: we evaluate the objective $\theta(\cdot)$ at two candidate points—the inexact block minimizer $\tilde{z}^k = (\tilde{x}^k, \tilde{y}^k)$ and the previous iterate $\tilde{w}^{k-1}$. We select the candidate with a smaller objective value as $\tilde{w}^k$, which strictly guarantees a nonincreasing sequence $\{\theta(\tilde{w}^k)\}$ and maintains the monotonicity property of iMABCD. We first give our proposed iMABCD algorithm for solving problem (1).

The following lemma describes the monotone decrease in the objective values for the iMABCD algorithm. Its proof follows the same reasoning as that of the exact counterpart in Lemma 2, with the optimality conditions applied at $(\tilde{x}^k, \tilde{y}^k)$.

Lemma 3. Under Assumption 1, let sequences $\{\tilde{z}^k := (\tilde{x}^k, \tilde{y}^k)\}$ and $\{\tilde{w}^k := (\tilde{u}^k, \tilde{v}^k)\}$ be produced by the iMABCD algorithm. Then, for any $k \geq 1$, we have

$$\theta(\tilde{w}^k) - \theta(w) \leq \frac{1}{2} \|w - w^k\|_{\mathcal{H}}^2 - \frac{1}{2} \|w - \tilde{z}^k\|_{\mathcal{H}}^2 + \varepsilon_k \|w - \tilde{z}^k\|_{\text{Diag}(\hat{\mathcal{Q}}_{11}, \hat{\mathcal{Q}}_{22})}, \quad \forall w \in \mathbb{U} \times \mathbb{V}.$$

At the $(k+1)$ -th iteration, we denote the exact solutions as

$$\begin{cases} \hat{x}^k = \arg \min_{u \in \mathbb{U}} \{p(u) + \hat{h}(u, v^k; w^k)\}, \\ \hat{y}^k = \arg \min_{v \in \mathbb{V}} \{q(v) + \hat{h}(\hat{x}^k, v; w^k)\}, \end{cases} \quad (21)$$

and $\hat{z}^k = (\hat{x}^k, \hat{y}^k)$. We set $(\hat{x}^0, \hat{y}^0) = (\tilde{x}^0, \tilde{y}^0) = (u^1, v^1)$ for consistency. The following lemma characterizes the discrepancy between $(\hat{x}^k, \hat{y}^k)$ and $(\tilde{x}^k, \tilde{y}^k)$.

Lemma 4. Suppose the iMABCD algorithm produces the sequences $\{(\tilde{x}^k, \tilde{y}^k)\}$ and $\{(x^k, y^k)\}$, and let $\{(\hat{x}^k, \hat{y}^k)\}$ be as defined in (21). Then, for any $k \geq 1$, the following inequalities hold:

$$\|\hat{x}^k - u^*\|_{\hat{\mathcal{Q}}_{11}}^2 \leq \|w^k - w^*\|_{\mathcal{H}}^2, \quad (22a)$$

$$\|\hat{\mathcal{Q}}_{11}^{\frac{1}{2}}(\hat{x}^k - \hat{x}^k)\| \leq \varepsilon_k, \quad \|\hat{\mathcal{Q}}_{22}^{\frac{1}{2}}(\hat{y}^k - \hat{y}^k)\| \leq (1 + \sqrt{2})\varepsilon_k, \quad (22b)$$

$$\|\tilde{z}^k - \hat{z}^k\|_{\mathcal{H}} \leq \sqrt{7}\varepsilon_k. \quad (22c)$$

Proof. (a) The optimality conditions of (21) and (1) are

$$0 \in \partial p(\hat{x}^k) + \nabla_u h(w^k) + \hat{\mathcal{Q}}_{11}(\hat{x}^k - u^k) \quad \text{and} \quad 0 \in \partial p(u^*) + \nabla_u h(w^*). \quad (23)$$

Applying the monotonicity of ∂p gives

$$\langle \hat{x}^k - u^*, \nabla_u h(w^k) - \nabla_u h(w^*) + \hat{\mathcal{Q}}_{11}(\hat{x}^k - u^k) \rangle \leq 0. \quad (24)$$

Given the global Lipschitz continuity assumed for ∇h , Clarke's mean value theorem applies. Consequently, one can find a self-adjoint positive semidefinite operator $W^k \in \text{conv}\{\partial^2 h([w^{k-1}, w^k])\}$ satisfying

$$\nabla h(w^k) - \nabla h(w^{k-1}) = W^k(w^k - w^{k-1}), \quad (25)$$

where $\text{conv}\{\partial^2 h([w^{k-1}, w^k])\}$ stands for the convex hull of all points in $\partial^2 h(z)$ with $z \in [w^{k-1}, w^k]$. We express W^k in the block form

$$W^k := \begin{pmatrix} W_{11}^k & W_{12}^k \\ (W_{12}^k)^* & W_{22}^k \end{pmatrix},$$

where $W_{11}^k : U \rightarrow U$ and $W_{22}^k : V \rightarrow V$ are self-adjoint positive semidefinite linear operators, while $W_{12}^k : U \rightarrow V$ is linear.

Using the two conditions

$$\left\langle \begin{pmatrix} \hat{x}^k - u^* \\ v^k - v^* \end{pmatrix}, W^k \begin{pmatrix} \hat{x}^k - u^* \\ v^k - v^* \end{pmatrix} \right\rangle \geq 0 \quad \text{and} \quad \mathcal{Q} \preceq W^k \preceq \hat{\mathcal{Q}}, \quad (26)$$

we derive

$$\begin{aligned} & 2 \left\langle \hat{x}^k - u^*, \nabla_u h(w^k) - \nabla_u h(w^*) \right\rangle \\ & \geq \left(\|\hat{x}^k - u^*\|_{W_{11}^k}^2 + \|u^k - u^*\|_{W_{11}^k}^2 - \|\hat{x}^k - u^k\|_{W_{11}^k}^2 \right) - \left(\|\hat{x}^k - u^*\|_{W_{11}^k}^2 + \|v^k - v^*\|_{W_{22}^k}^2 \right) \\ & \geq \|u^k - u^*\|_{\mathcal{Q}_{11}}^2 - \|\hat{x}^k - u^k\|_{\mathcal{Q}_{11}}^2 - \|v^k - v^*\|_{\mathcal{Q}_{22}}^2. \end{aligned} \quad (27)$$

Now, putting together (24) and (27) yields

$$\begin{aligned} 0 & \geq \|u^k - u^*\|_{\mathcal{Q}_{11}}^2 - \|\hat{x}^k - u^k\|_{\mathcal{Q}_{11}}^2 - \|v^k - v^*\|_{\mathcal{Q}_{22}}^2 + 2 \left\langle \hat{x}^k - u^*, \hat{\mathcal{Q}}_{11}(\hat{x}^k - u^*) \right\rangle \\ & \geq \|\hat{x}^k - u^k\|_{\mathcal{Q}_{11}}^2 - \|u^k - u^*\|_{\mathcal{D}_1}^2 - \|v^k - v^*\|_{\mathcal{Q}_{22}}^2, \\ & \geq \|\hat{x}^k - u^k\|_{\mathcal{Q}_{11}}^2 - \|w^k - w^*\|_{\mathcal{H}}^2 \end{aligned}$$

from which we conclude that

$$\|\hat{x}^k - u^k\|_{\mathcal{Q}_{11}}^2 \geq \|w^k - w^*\|_{\mathcal{H}}^2.$$

(b) Proceeding analogously to the derivation of (24) and invoking the optimality conditions from (4) and (21) in conjunction with the monotonicity of ∂p and ∂q , the following relations are obtained:

$$\left\langle \hat{\mathcal{Q}}_{11}(\hat{x}^k - \hat{x}^k) + \delta_u^k, \hat{x}^k - \hat{x}^k \right\rangle \leq 0 \quad \text{and} \quad \left\langle \mathcal{Q}_{12}^*(\hat{x}^k - \hat{x}^k) + \hat{\mathcal{Q}}_{22}(\hat{y}^k - \hat{y}^k) + \delta_v^k, \hat{y}^k - \hat{y}^k \right\rangle \leq 0.$$

From the first inequality, we obtain

$$\|\hat{x}^k - \hat{x}^k\|_{\hat{\mathcal{Q}}_{11}} = \|\hat{\mathcal{Q}}_{11}^{-\frac{1}{2}}(\hat{x}^k - \hat{x}^k)\| \leq \|\hat{\mathcal{Q}}_{11}^{-\frac{1}{2}}\delta_u^k\| \leq \varepsilon_k. \quad (28)$$

The second inequality leads to

$$\begin{aligned} \|\hat{y}^k - \hat{y}^k\|_{\hat{\mathcal{Q}}_{22}}^2 & \leq \|\hat{\mathcal{Q}}_{22}^{-\frac{1}{2}}\delta_v^k\| \|\hat{\mathcal{Q}}_{22}^{\frac{1}{2}}(\hat{y}^k - \hat{y}^k)\| - \left\langle \mathcal{Q}_{12}^*(\hat{x}^k - \hat{x}^k), \hat{y}^k - \hat{y}^k \right\rangle \\ & \leq \|\hat{\mathcal{Q}}_{22}^{\frac{1}{2}}\delta_v^k\| \|\hat{\mathcal{Q}}_{22}^{-\frac{1}{2}}(\hat{y}^k - \hat{y}^k)\| + \frac{1}{2} \left(\|\hat{x}^k - \hat{x}^k\|_{\hat{\mathcal{Q}}_{11}}^2 + \|\hat{y}^k - \hat{y}^k\|_{\hat{\mathcal{Q}}_{22}}^2 \right), \end{aligned} \quad (29)$$

which implies that

$$\begin{aligned} \|\hat{y}^k - \hat{y}^k\|_{\hat{\mathcal{Q}}_{22}}^2 & \leq 2 \|\hat{\mathcal{Q}}_{22}^{-\frac{1}{2}}\delta_v^k\| \|\hat{\mathcal{Q}}_{22}^{\frac{1}{2}}(\hat{y}^k - \hat{y}^k)\| + \|\hat{x}^k - \hat{x}^k\|_{\hat{\mathcal{Q}}_{11}}^2, \\ & \leq 2\varepsilon_k \|\hat{\mathcal{Q}}_{22}^{\frac{1}{2}}(\hat{y}^k - \hat{y}^k)\| + \varepsilon_k^2. \end{aligned} \quad (30)$$

Solving the resulting quadratic inequality yields

$$\|\hat{\mathcal{Q}}_{22}^{\frac{1}{2}}(\hat{y}^k - \hat{y}^k)\| \leq (1 + \sqrt{2})\varepsilon_k. \quad (31)$$

(c) Combining (a) and (b) yields

$$\|\hat{z}^k - \hat{z}^k\|_{\mathcal{H}}^2 = \|\hat{x}^k - \hat{x}^k\|_{\mathcal{D}_1}^2 + \|\hat{y}^k - \hat{y}^k\|_{\hat{\mathcal{Q}}_{22}}^2 \leq 7\varepsilon_k^2. \quad (32)$$

□

We now proceed to state the principal theorem regarding the $O(1/k^2)$ complexity of the inexact monotone ABCD algorithm.

Theorem 2. Assume that Assumption 1 is satisfied and the solution set Ω of (1) is nonempty. Take w^* as an element of Ω . Further suppose that $\sum_{i=1}^{\infty} i\varepsilon_i < \infty$. Then, the following bound holds for the sequence $\{\tilde{w}^k\} := \{(\tilde{u}^k, \tilde{v}^k)\}$ produced by the iMABCD algorithm:

$$\theta(\tilde{w}^k) - \theta(w^*) \leq \frac{2\|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2 + c_0}{(k+1)^2}, \quad \forall k \geq 1,$$

where $c_0 > 0$ is a constant independent of k .

Proof. Taking $k = 1$ and $w = w^*$ in Lemma 2 yields

$$\begin{aligned} 2[\theta(\hat{z}^1) - \theta(w^*)] &\leq \|w^1 - w^*\|_{\mathcal{H}}^2 - \|\hat{z}^1 - w^*\|_{\mathcal{H}}^2, \\ &= \|\hat{z}^0 - w^*\|_{\mathcal{H}}^2 - \|\hat{z}^1 - w^*\|_{\mathcal{H}}^2. \end{aligned} \quad (33)$$

For any $k \geq 2$, since $\hat{z}^k = (\hat{x}^k, \hat{y}^k)$ are the exact block minimizers, we set $\hat{z}^k = \hat{z}^k$ and $w = \frac{(t_k-1)\hat{z}^{k-1} + w^*}{t_k}$ in Lemma 2. This gives

$$\theta(\hat{z}^k) - \theta(w) \leq \frac{1}{2} \left\| \frac{(t_k-1)\hat{z}^{k-1} + w^*}{t_k} - w^k \right\|_{\mathcal{H}}^2 - \frac{1}{2} \left\| \frac{(t_k-1)\hat{z}^{k-1} + w^*}{t_k} - \hat{z}^k \right\|_{\mathcal{H}}^2, \quad (34)$$

which leads to

$$2t_k^2[\theta(\hat{z}^k) - \theta(w)] \leq \left\| t_k w^k - w^* - (t_k-1)\tilde{w}^{k-1} \right\|_{\mathcal{H}}^2 - \left\| (t_k-1)\hat{z}^{k-1} + w^* - t_k \hat{z}^k \right\|_{\mathcal{H}}^2. \quad (35)$$

Applying Lemma 1(c) with $\tilde{w} = \hat{z}^{k-1}$ and $\hat{w} = w^*$ gives

$$2t_k^2[\theta(w) - \theta(w^*)] \leq 2t_{k-1}^2[\theta(\hat{z}^{k-1}) - \theta(w^*)]. \quad (36)$$

Merging the preceding two inequalities and using the identity $t_k w^k = t_k \tilde{w}^{k-1} + (t_{k-1} - 1)(\tilde{w}^{k-1} - \tilde{w}^{k-2}) + t_{k-1}(\hat{z}^{k-1} - \tilde{w}^{k-1})$, we obtain, for every $k \geq 2$,

$$\begin{aligned} &2t_k^2[\theta(\hat{z}^k) - \theta(w^*)] - 2t_{k-1}^2[\theta(\hat{z}^{k-1}) - \theta(w^*)] \\ &\leq \left\| t_{k-1}\hat{z}^{k-1} - (t_{k-1}-1)\tilde{w}^{k-2} - (t_k-1)(\hat{z}^{k-1} - \tilde{w}^{k-1}) \right\|_{\mathcal{H}}^2 - \left\| t_k \hat{z}^k - w^* - (t_k-1)\hat{z}^{k-1} \right\|_{\mathcal{H}}^2, \\ &= \left\| \lambda^{k-1} \right\|_{\mathcal{H}}^2 - 2 \left\langle \mathcal{H}\lambda^{k-1}, t_{k-1}(\hat{z}^{k-1} - \hat{z}^{k-1}) - (t_k-1)(\hat{z}^{k-1} - \tilde{w}^{k-1}) - (t_{k-1}-1)(\hat{z}^{k-2} - \tilde{w}^{k-2}) \right\rangle \\ &\quad + \left\| t_{k-1}(\hat{z}^{k-1} - \hat{z}^{k-1}) - (t_k-1)(\hat{z}^{k-1} - \tilde{w}^{k-1}) - (t_{k-1}-1)(\hat{z}^{k-2} - \tilde{w}^{k-2}) \right\|_{\mathcal{H}}^2 - \left\| \lambda^k \right\|_{\mathcal{H}}^2, \end{aligned} \quad (37)$$

where $\lambda^k := t_k \hat{z}^k - w^* - (t_k - 1) \hat{z}^{k-1}$. We observe that $\|\hat{z}^k - \tilde{w}^k\|_{\mathcal{H}}^2 \leq \|\hat{z}^k - \tilde{z}^k\|_{\mathcal{H}}^2$ holds. Consequently, by invoking Lemma 1(b), Lemma 4, and the nonincreasing nature of $\{\varepsilon_k\}$, we obtain, for all $k \geq 3$,

$$\begin{aligned} & \left\| t_{k-1}(\hat{z}^{k-1} - \tilde{z}^{k-1}) - (t_k - 1)(\hat{z}^{k-1} - \tilde{w}^{k-1}) - (t_{k-1} - 1)(\hat{z}^{k-2} - \tilde{w}^{k-2}) \right\|_{\mathcal{H}} \\ & \leq (t_{k-1} + t_k - 1) \left\| \hat{z}^{k-1} - \tilde{z}^{k-1} \right\|_{\mathcal{H}} + (t_{k-1} - 1) \left\| \hat{z}^{k-2} - \tilde{z}^{k-2} \right\|_{\mathcal{H}}, \\ & \leq c_1(k-1)\varepsilon_{k-2}, \end{aligned} \quad (38)$$

where $c_1 = 5$. For the particular case $k = 2$, the same bound also holds:

$$\begin{aligned} & \left\| t_{k-1}(\hat{z}^{k-1} - \tilde{z}^{k-1}) - (t_k - 1)(\hat{z}^{k-1} - \tilde{w}^{k-1}) - (t_{k-1} - 1)(\hat{z}^{k-2} - \tilde{w}^{k-2}) \right\|_{\mathcal{H}} \\ & \leq (t_{k-1} + t_k - 1) \left\| \hat{z}^{k-1} - \tilde{z}^{k-1} \right\|_{\mathcal{H}} + (t_{k-1} - 1) \left\| \hat{z}^{k-2} - \tilde{z}^{k-2} \right\|_{\mathcal{H}}, \\ & = t_2 \left\| \hat{z}^1 - \tilde{z}^1 \right\|_{\mathcal{H}} \leq c_1(k-1)\varepsilon_{k-2}, \end{aligned} \quad (39)$$

with the convention $\varepsilon_0 = \varepsilon_1$. Then, applying (37) yields, for every $k \geq 2$,

$$\begin{aligned} & 2t_k^2[\theta(\hat{z}^k) - \theta(w^*)] + \left\| \lambda^k \right\|_{\mathcal{H}}^2 \\ & \leq 2t_{k-1}^2[\theta(\hat{z}^{k-1}) - \theta(w^*)] + \left\| \lambda^{k-1} \right\|_{\mathcal{H}}^2 + 2c_1(k-1)\varepsilon_{k-2} \left\| \lambda^{k-1} \right\|_{\mathcal{H}} + c_1^2(k-1)^2\varepsilon_{k-2}^2, \\ & \leq \dots \\ & \leq 2t_1^2[\theta(\hat{z}^1) - \theta(w^*)] + \left\| \lambda^1 \right\|_{\mathcal{H}}^2 + 2c_1 \sum_{i=1}^{k-1} i\varepsilon_{i-1} \left\| \lambda^i \right\|_{\mathcal{H}} + c_1^2 \sum_{i=1}^{k-1} i^2\varepsilon_{i-1}^2, \\ & \leq \left\| w^1 - w^* \right\|_{\mathcal{H}}^2 + 2c_1 \sum_{i=1}^{k-1} i\varepsilon_{i-1} \left\| \lambda^i \right\|_{\mathcal{H}} + c_1^2 \sum_{i=1}^{k-1} i^2\varepsilon_{i-1}^2, \\ & = \left\| \hat{z}^0 - w^* \right\|_{\mathcal{H}}^2 + 2c_1 \sum_{i=1}^{k-1} i\varepsilon_{i-1} \left\| \lambda^i \right\|_{\mathcal{H}} + c_1^2 \sum_{i=1}^{k-1} i^2\varepsilon_{i-1}^2. \end{aligned} \quad (40)$$

From the above inequalities, the sequence $\{\|\lambda^k\|_{\mathcal{H}}\}$ is bounded. If $\|\lambda^k\|_{\mathcal{H}} \leq 1$ holds for every $k \geq 1$, then the conclusion follows. Otherwise, suppose that, for some arbitrarily large positive integer m , we have

$$\|\lambda^{k_m}\|_{\mathcal{H}} := \max \left\{ \|\lambda^i\|_{\mathcal{H}} : 1 \leq i \leq m \right\} \geq 1. \quad (41)$$

Then, for each $1 \leq k \leq m$, it holds that

$$\begin{aligned} \|\lambda^k\|_{\mathcal{H}} & \leq \|\lambda^{k_m}\|_{\mathcal{H}} \leq \frac{1}{\|\lambda^{k_m}\|_{\mathcal{H}}} \left(\left\| \hat{z}^0 - w^* \right\|_{\mathcal{H}}^2 + 2c_1 \sum_{i=1}^{k-1} i\varepsilon_{i-1} \left\| \lambda^i \right\|_{\mathcal{H}} + c_1^2 \sum_{i=1}^{k-1} i^2\varepsilon_{i-1}^2 \right), \\ & \leq \left\| \hat{z}^0 - w^* \right\|_{\mathcal{H}}^2 + 2c_1 \sum_{i=1}^{k_m-1} i\varepsilon_{i-1} \frac{\|\lambda^i\|_{\mathcal{H}}}{\|\lambda^{k_m}\|_{\mathcal{H}}} + c_1^2 \sum_{i=1}^{k_m-1} i^2\varepsilon_{i-1}^2, \\ & \leq \left\| \hat{z}^0 - w^* \right\|_{\mathcal{H}}^2 + 2c_1 \sum_{i=1}^{\infty} i\varepsilon_{i-1} + c_1^2 \sum_{i=1}^{\infty} i^2\varepsilon_{i-1}^2. \end{aligned} \quad (42)$$

Passing to the limit as $m \rightarrow \infty$ yields

$$\|\lambda^k\|_{\mathcal{H}} \leq c_2 := \max \left\{ 1, \left\| \hat{z}^0 - w^* \right\|_{\mathcal{H}}^2 + 2c_1 \sum_{i=1}^{\infty} i\varepsilon_{i-1} + c_1^2 \sum_{i=1}^{\infty} i^2\varepsilon_{i-1}^2 \right\}, \quad \forall k \geq 1. \quad (43)$$

To obtain an upper bound for $\|\hat{z}^{k+1} - w^*\|_{\mathcal{H}'}$, we take $w = w^*$ and $\tilde{w}^k = \hat{z}^{k+1}$ in Lemma 2, which yields

$$\begin{aligned} t_{k+1} \|\hat{z}^{k+1} - w^*\|_{\mathcal{H}} &\leq t_{k+1} \|w^{k+1} - w^*\|_{\mathcal{H}'} \\ &\leq (t_{k+1} - 1) \|\hat{z}^k - w^*\|_{\mathcal{H}} + \|t_k \hat{z}^k - (t_k - 1) \hat{z}^{k-1} - w^*\|_{\mathcal{H}} \\ &\quad + (t_{k+1} - 1) \|\tilde{w}^k - \hat{z}^k\|_{\mathcal{H}} + t_k \|\hat{z}^k - \hat{z}^{k-1}\|_{\mathcal{H}} + (t_k - 1) \|\hat{z}^{k-1} - \tilde{w}^{k-1}\|_{\mathcal{H}'}, \\ &\leq \frac{t_k^2}{t_{k+1}} \|\hat{z}^k - w^*\|_{\mathcal{H}} + c_2 + c_1 \varepsilon_{k-1} t_{k+1}, \end{aligned}$$

where the equality results from $w^{k+1} = \tilde{w}^k + \frac{t_k - 1}{t_{k+1}} (\tilde{w}^k - \tilde{w}^{k-1}) + \frac{t_k}{t_{k+1}} (\hat{z}^k - \tilde{w}^k)$, and the last inequality is derived via Lemma 1(a). Consequently, the above inequality implies

$$\begin{aligned} \frac{t_k^2}{t_{k+1}^2} \|\hat{z}^k - w^*\|_{\mathcal{H}} &\leq \frac{t_k^2}{t_{k+1}^2} \left(\frac{t_{k-1}^2}{t_k^2} \|\hat{z}^{k-1} - w^*\|_{\mathcal{H}} + \frac{c_2}{t_k} + c_1 \varepsilon_{k-2} \right), \\ \frac{t_k^2}{t_{k+1}^2} \cdot \frac{t_{k-1}^2}{t_k^2} \|\hat{z}^{k-1} - w^*\|_{\mathcal{H}} &\leq \frac{t_k^2}{t_{k+1}^2} \cdot \frac{t_{k-1}^2}{t_k^2} \left(\frac{t_{k-2}^2}{t_{k-1}^2} \|\hat{z}^{k-2} - w^*\|_{\mathcal{H}} + \frac{c_2}{t_{k-1}} + c_1 \varepsilon_{k-3} \right), \\ &\vdots \\ \frac{t_k^2}{t_{k+1}^2} \cdot \frac{t_{k-1}^2}{t_k^2} \cdots \frac{t_2^2}{t_3^2} \|\hat{z}^2 - w^*\|_{\mathcal{H}} &\leq \frac{t_k^2}{t_{k+1}^2} \cdot \frac{t_{k-1}^2}{t_k^2} \cdots \frac{t_2^2}{t_3^2} \left(\frac{t_1^2}{t_2^2} \|\hat{z}^1 - w^*\|_{\mathcal{H}} + \frac{c_2}{t_2} + c_1 \varepsilon_0 \right). \end{aligned}$$

Summing the preceding inequalities yields

$$\begin{aligned} \|\hat{z}^{k+1} - w^*\|_{\mathcal{H}} &\leq \frac{t_1^2}{t_{k+1}^2} \|\hat{z}^1 - w^*\|_{\mathcal{H}} + c_2 \sum_{i=1}^k \frac{t_{i+1}}{t_{k+1}^2} + c_1 \sum_{i=1}^k \varepsilon_{i-1}, \\ &\leq \frac{4}{(k+2)^2} \|\hat{z}^1 - w^*\|_{\mathcal{H}} + 2c_2 + c_1 \sum_{i=1}^k \varepsilon_{i-1}, \\ &\leq c_3 := \frac{4}{9} \|\hat{z}^1 - w^*\|_{\mathcal{H}} + 2c_2 + c_1 \sum_{i=1}^k \varepsilon_{i-1}, \end{aligned}$$

where we employ the inequality $\sum_{i=1}^k \frac{t_{i+1}}{t_{k+1}^2} \leq \frac{(3+k)k}{2(\frac{1}{2}k+1)^2} \leq 2$, $\forall k \geq 1$, which is derived from Lemma 1(b). In addition, we have $\|\hat{w}^1 - w^*\|_{\mathcal{H}} \leq c_3$.

We now turn to establishing the boundedness of the quantity $\|t_k \hat{x}^k - u^* - (t_k - 1) \hat{x}^{k-1}\|_{\hat{\mathcal{Q}}_{11}}$. For $k \geq 2$,

$$\begin{aligned} t_k \|\hat{x}^k - u^*\|_{\hat{\mathcal{Q}}_{11}} &\leq t_k \|w^k - w^*\|_{\mathcal{H}'} \\ &= \|(t_k - 1)(\tilde{w}^{k-1} - w^*) + t_{k-1}(\hat{z}^{k-1} - w^*) - (t_{k-1} - 1)(\tilde{w}^{k-2} - w^*)\|_{\mathcal{H}'}, \\ &\leq (t_k - 1) \left(\|\tilde{w}^{k-1} - \hat{z}^{k-1}\|_{\mathcal{H}} + \|\hat{z}^{k-1} - w^*\|_{\mathcal{H}} \right) \\ &\quad + \|t_{k-1}(\hat{z}^{k-1} - \hat{z}^{k-1}) - (t_{k-1} - 1)(\tilde{w}^{k-2} - \hat{z}^{k-2})\|_{\mathcal{H}} \\ &\quad + \|t_{k-1}(\hat{z}^{k-1} - w^*) - (t_{k-1} - 1)(\tilde{w}^{k-2} - w^*)\|_{\mathcal{H}'}, \\ &\leq (t_k - 1)(\sqrt{7}\varepsilon_{k-1} + c_3) + (2t_{k-1} - 1)\sqrt{7}\varepsilon_{k-2} + \|\lambda^{k-1}\|_{\mathcal{H}}, \\ &\leq t_k(8\varepsilon_1 + c_2 + c_3). \end{aligned}$$

When $k = 1$, $\|\hat{x}^k - u^*\|_{\hat{Q}_{11}} \leq \|\hat{z}^k - w^*\|_{\mathcal{H}} \leq c_3$. Consequently, for every $k \geq 2$,

$$\begin{aligned} & \|t_k \tilde{x}^k - u^* - (t_k - 1) \tilde{u}^{k-1}\|_{\hat{Q}_{11}} \\ & \leq t_k \|\tilde{x}^k - u^*\|_{\hat{Q}_{11}} + (t_k - 1) \|\tilde{x}^{k-1} - u^*\|_{\hat{Q}_{11}}, \\ & \leq t_k \left(\|\hat{x}^k - \tilde{x}^k\|_{\hat{Q}_{11}} + \|\hat{x}^k - u^*\|_{\hat{Q}_{11}} \right) + (t_k - 1) \left(\|\hat{x}^{k-1} - u^*\|_{\hat{Q}_{11}} + \|\hat{x}^{k-1} - \tilde{x}^{k-1}\|_{\hat{Q}_{11}} \right), \\ & \leq (2t_k - 1)[9\varepsilon_1 + c_2 + c_3]. \end{aligned}$$

At $k = 1$, it holds that

$$\begin{aligned} \|t_1 \tilde{x}^1 - u^* - (t_1 - 1) \tilde{u}^0\|_{\hat{Q}_{11}} &= \|\tilde{x}^1 - u^*\|_{\hat{Q}_{11}} \\ &\leq \|\hat{x}^1 - u^*\|_{\hat{Q}_{11}} + \|\tilde{x}^1 - \hat{x}^1\|_{\hat{Q}_{11}}, \\ &\leq \|w^1 - w^*\|_{\mathcal{H}} + \varepsilon_1, \\ &\leq c_2 + \varepsilon_1, \\ &\leq (2t_k - 1)[9\varepsilon_1 + c_2 + c_3]. \end{aligned}$$

Meanwhile,

$$\begin{aligned} & \|t_k \tilde{y}^k - v^* - (t_k - 1) \tilde{v}^{k-1}\|_{\hat{Q}_{22}} \\ & \leq \|t_k(\tilde{y}^k - v^*) - (t_k - 1)(\tilde{v}^{k-1} - v^*)\|_{\hat{Q}_{22}}, \\ & \leq \|t_k(\tilde{y}^k - v^*) - (t_k - 1)(\tilde{y}^{k-1} - v^*)\|_{\hat{Q}_{22}} + \|t_k(\tilde{y}^k - \tilde{y}^{k-1}) - (t_k - 1)(\tilde{v}^{k-1} - \tilde{y}^{k-1})\|_{\hat{Q}_{22}}, \\ & \leq \|\lambda^k\|_{\mathcal{H}} + t_k \|\tilde{y}^k - \tilde{y}^{k-1}\|_{\hat{Q}_{22}} + (t_k - 1) \|\tilde{v}^{k-1} - \tilde{y}^{k-1}\|_{\hat{Q}_{22}}, \\ & \leq c_2 + (2t_k - 1)(1 + \sqrt{2})\varepsilon_{k-1}, \\ & \leq (2t_k - 1)[c_2 + 3c_3]. \end{aligned}$$

Substituting $w = \frac{(t_k - 1)\tilde{w}^{k-1} + w^*}{t_k}$ into Lemma 3 and employing Lemma 1(c) yields

$$\begin{aligned} & t_k^2 [\theta(\tilde{w}^k) - \theta(w)] + \frac{1}{2} \|(t_k - 1)\tilde{w}^{k-1} + w^* - t_k \tilde{z}^k\|_{\mathcal{H}}^2 \\ & \leq \frac{1}{2} \|(t_k - 1)\tilde{w}^{k-1} + w^* - t_k \tilde{w}^k\|_{\mathcal{H}}^2 + \varepsilon_k \|(t_k - 1)\tilde{w}^{k-1} + w^* - t_k \tilde{z}^k\|_{\text{Diag}(\hat{Q}_{11}, \hat{Q}_{22})}, \end{aligned} \quad (44)$$

and,

$$t_k^2 [\theta(w) - \theta(w^*)] \leq t_{k-1}^2 [\theta(\tilde{w}^{k-1}) - \theta(w^*)]. \quad (45)$$

Adding the foregoing two inequalities gives

$$\begin{aligned}
& t_k^2 [\theta(\tilde{w}^k) - \theta(w^*)] + \frac{1}{2} \|t_k \tilde{z}^k - w^* - (t_k - 1) \tilde{w}^{k-1}\|_{\mathcal{H}}^2 \\
& + \varepsilon_k \|t_k \tilde{z}^k - w^* - (t_k - 1) \tilde{w}^{k-1}\|_{\text{Diag}(\hat{Q}_{11}, \hat{Q}_{22})}, \\
& \leq t_{k-1}^2 [\theta(\tilde{w}^{k-1}) - \theta(w^*)] + \frac{1}{2} \|t_{k-1} \tilde{z}^{k-1} - w^* - (t_{k-1} - 1) \tilde{w}^{k-2}\|_{\mathcal{H}}^2 \\
& + \varepsilon_k \|t_k \tilde{z}^k - w^* - (t_k - 1) \tilde{w}^{k-1}\|_{\text{Diag}(\hat{Q}_{11}, \hat{Q}_{22})}, \\
& \leq \dots \\
& \leq t_1^2 [\theta(\tilde{w}^1) - \theta(w^*)] + \frac{1}{2} \|t_1 \tilde{z}^1 - w^* - (t_1 - 1) \tilde{w}^0\|_{\mathcal{H}}^2 \\
& + \sum_{i=2}^k \varepsilon_i \|t_i \tilde{z}^i - w^* - (t_i - 1) \tilde{w}^{i-1}\|_{\text{Diag}(\hat{Q}_{11}, \hat{Q}_{22})}, \\
& \leq \frac{1}{2} \|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2 + \sum_{i=1}^k (2t_i - 1) [12\varepsilon_1 + 2c_2 + c_3] \varepsilon_i, \\
& \leq \frac{1}{2} \|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2 + \frac{1}{4} c_0,
\end{aligned} \tag{46}$$

where

$$c_0 := 4 \sum_{i=1}^{\infty} (2t_i - 1) [12\varepsilon_1 + 2c_2 + c_3] \varepsilon_{i-1} \tag{47}$$

is a finite value since $\sum_{i=1}^{\infty} t_i \varepsilon_i \leq \infty$. Noting $t_k \geq \frac{k+1}{2}$, we have

$$\theta(\tilde{w}^k) - \theta(w^*) \leq \frac{2\|\tilde{w}^0 - w^*\|_{\mathcal{H}}^2 + c_0}{(k+1)^2}, \quad \forall k \geq 1. \tag{48}$$

□

4. Numerical Experiments

In this section, we evaluate the numerical performance of the proposed MABCD method and its inexact variant iMABCD on two problems: the extended Lasso problem with sparse outliers and an ill-conditioned non-negative sparse regression problem. The following methods are compared: the original ABCD, its monotone variant MABCD, the inexact monotone method iMABCD, FISTA, and ADMM. All algorithms were implemented in MATLAB and run on a workstation equipped with an Intel(R) Core(TM) i9-10920X CPU at 3.50 GHz and 128 GB of RAM.

4.1. Extended Lasso with Sparse Outliers

Consider the linear measurement model with sparse outliers [19]

$$y = X\beta^* + \sqrt{n}e^* + \alpha, \tag{49}$$

where $X \in \mathbb{R}^{n \times p}$ is the measurement or design matrix, $\beta^* \in \mathbb{R}^p$ is the unknown sparse regression vector, $e^* \in \mathbb{R}^n$ is a sparse vector of outlier errors, and $\alpha \in \mathbb{R}^n$ is random noise. The extended Lasso formulation for robust recovery reads

$$\min_{\beta \in \mathbb{R}^p, e \in \mathbb{R}^n} F(\beta, e) := \frac{1}{2n} \|y - X\beta - \sqrt{n}e\|_2^2 + \lambda_\beta \|\beta\|_1 + \lambda_e \|e\|_1, \tag{50}$$

where $\lambda_\beta > 0$ and $\lambda_e > 0$ are regularization parameters. Let $z = (\beta, e)$, then (50) falls exactly into the two-block composite convex framework studied in this paper, with

$$h(z) = \frac{1}{2n} \|y - X\beta - \sqrt{n}e\|_2^2, \quad p(\beta) = \lambda_\beta \|\beta\|_1, \quad q(e) = \lambda_e \|e\|_1.$$

The Hessian of the smooth term is

$$\mathcal{Q} = \nabla^2 h = \begin{pmatrix} \frac{1}{n} X^\top X & \frac{1}{\sqrt{n}} X^\top \\ \frac{1}{\sqrt{n}} X & I \end{pmatrix}.$$

Following Algorithms 2 and 3, at each iterate $z^k = (\beta^k, e^k)$, we construct the following majorization model $\hat{h}(z; z^k)$:

$$\hat{h}(z; z^k) = h(z^k) + \langle \nabla h(z^k), z - z^k \rangle + \frac{1}{2} |z - z^k|_{\hat{\mathcal{Q}}}^2,$$

where $\hat{\mathcal{Q}}$ is a self-adjoint operator satisfying $\hat{\mathcal{Q}} \succeq \mathcal{Q}$. Specifically, we choose

$$\hat{\mathcal{Q}} = \begin{pmatrix} L_\beta I & \frac{1}{\sqrt{n}} X^\top \\ \frac{1}{\sqrt{n}} X & L_e I \end{pmatrix} = \mathcal{Q} + \begin{pmatrix} L_\beta I - \frac{1}{n} X^\top X & 0 \\ 0 & (L_e - 1)I \end{pmatrix},$$

where $L_\beta \geq \frac{1}{n} \lambda_{\max}(X^\top X)$, and $L_e \geq 1$. Both block subproblems then reduce to proximal operators with ℓ_1 regularization, which admit closed-form solutions via the soft-thresholding operator $\mathcal{S}_\lambda : \mathbb{R}^d \rightarrow \mathbb{R}^d$, defined component-wise by

$$[\mathcal{S}_\lambda(a)]_i = \text{sign}(a_i) \max\{|a_i| - \lambda, 0\}.$$

Specifically, letting $r^k = X\beta^k + \sqrt{n}e^k - y$, the updates are

$$\tilde{\beta}^k = \mathcal{S}_{\lambda_\beta/L_\beta} \left(\beta^k - \frac{1}{nL_\beta} X^\top r^k \right), \quad \tilde{e}^k = \mathcal{S}_{\lambda_e/L_e} \left(e^k - \frac{1}{L_e \sqrt{n}} r^k \right),$$

where $r_\beta^k = X\tilde{\beta}^k + \sqrt{n}e^k - y$.

In our experiments, we set $n = 1024$ and $p = 2048$. The entries of X are i.i.d. standard Gaussian. The support sets of β^* and e^* are drawn uniformly without replacement with sparsity ratios 10% and 5%, respectively. Nonzero entries are generated as $\beta_j^* = a + b\zeta_j$, with $a = 1, b = 10$ and $\zeta_j \sim \mathcal{N}(0, 1)$, and $e_i^* = c + d\zeta_i$ with $c = 3, d = 2$ and $\zeta_i \sim \mathcal{N}(0, 1)$. The random noise is $\alpha = 0.01\eta$ with $\eta \sim \mathcal{N}(0, I_n)$. The regularization parameters are fixed at $\lambda_\beta = 3.0 \times 10^{-3}$, $\lambda_e = 4.0 \times 10^{-3}$. All algorithms are terminated when either the maximum number of outer iterations $K_{\max} = 3000$ is reached or the KKT residual is below 10^{-5} . For

$$r^k = X\beta^k + \sqrt{n}e^k - y, \quad g_\beta^k = \frac{1}{n} X^\top r^k, \quad g_e^k = \frac{1}{\sqrt{n}} r^k,$$

the residual is defined by

$$\text{KKT}_k := \max \left\{ \left\| \beta^k - \mathcal{S}_{\lambda_\beta}(\beta^k - g_\beta^k) \right\|_\infty, \left\| e^k - \mathcal{S}_{\lambda_e}(e^k - g_e^k) \right\|_\infty \right\}.$$

FISTA is applied to the joint variable (β, e) using the reciprocal of the Lipschitz constant of ∇h as the step size. ADMM uses an adaptive penalty parameter, and its linear systems are solved by Cholesky factorization.

The solution quality is assessed by the relative errors

$$\text{RelErr}_\beta = \frac{\|\hat{\beta} - \beta^*\|_2}{\max\{\varepsilon, \|\beta^*\|_2\}}, \quad \text{RelErr}_e = \frac{\|\hat{e} - e^*\|_2}{\max\{\varepsilon, \|e^*\|_2\}},$$

and the F1 score for support recovery. Let $S_\beta^* = \{j : \beta_j^* \neq 0\}$ and $\hat{S}_\beta = \{j : |\hat{\beta}_j| > \delta\}$, with threshold $\delta = 10^{-2}$; S_e^* and $\hat{S}_e$ are defined analogously. Treating the union of the two blocks as a single joint support, we compute the true positives (TPs), false positives (FPs), and false negatives (FNs) and define

$$\text{F1} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}.$$

Table 2 reports the numerical results for one representative random instance. This table shows that all algorithms recover the true support almost perfectly and yield comparably small relative errors. Comparing MABCD with ABCD, the monotone version has fewer iterations and shorter computing time, while attaining essentially the same objective value. This acceleration is attributed to the monotone descent constraint, which suppresses the oscillatory behavior of the original ABCD and thus produces a more direct path to the optimum, as further illustrated in Figures 1 and 2. In all subsequent figures, F^* denotes the optimal objective value. FISTA requires more outer iterations than the block-coordinate methods but remains competitive in CPU time because of its low per-iteration cost. ADMM is the slowest method for this instance in terms of both iteration count and CPU time.

Table 2. Single-instance performance comparison on the extended Lasso problem.

Method	Iter	Time (s)	Objective	F1	RelErr _β	RelErr _e
MABCD	1064	0.7539	5.382327	1.0000	8.494×10^{-4}	2.489×10^{-3}
ABCD	1117	0.7832	5.382326	1.0000	8.485×10^{-4}	2.486×10^{-3}
FISTA	1296	0.8473	5.382326	1.0000	8.469×10^{-4}	2.483×10^{-3}
ADMM	1876	5.8896	5.382326	1.0000	8.467×10^{-4}	2.482×10^{-3}

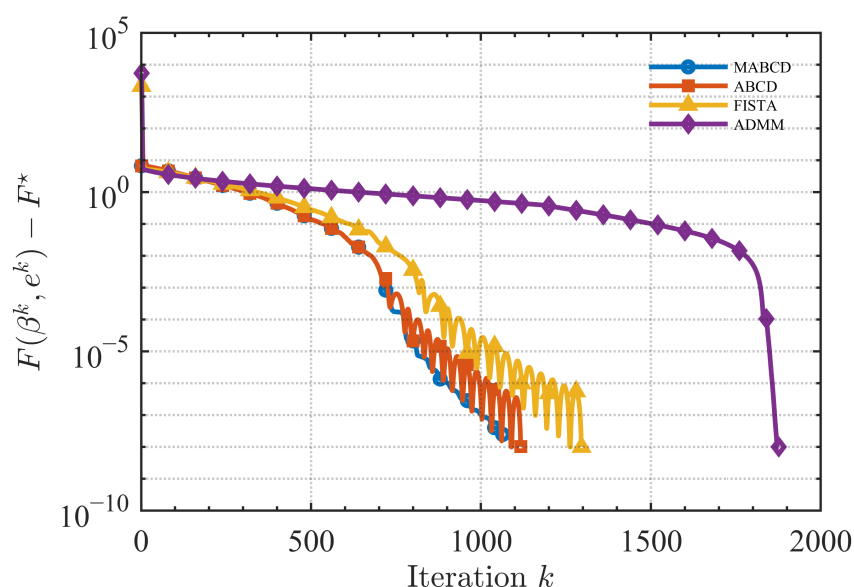


Figure 1. Objective gap versus iteration number for the extended Lasso problem.

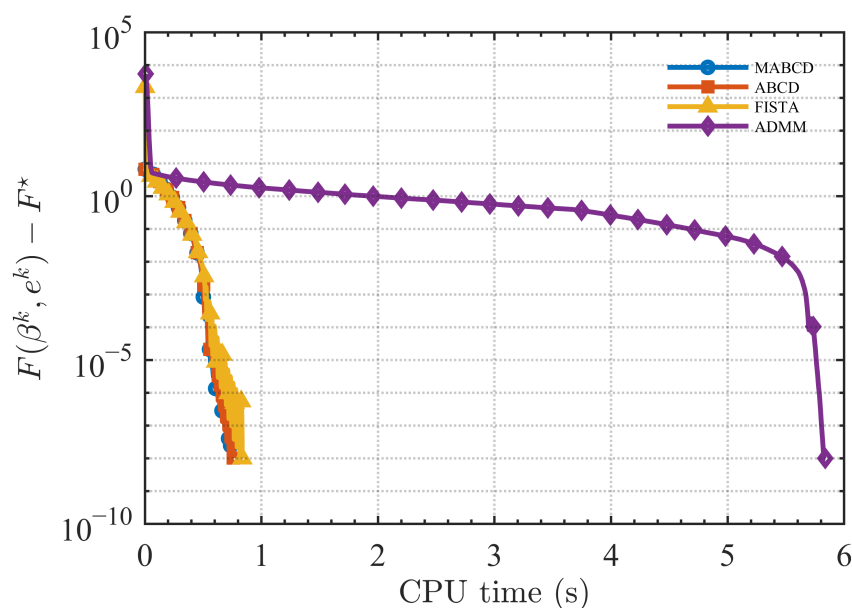


Figure 2. Objective gap versus CPU time for the extended Lasso problem.

To assess robustness with respect to random data generation, the experiment was repeated on 50 independent instances. Table 3 reports the mean results. All algorithms attain the same mean objective value for the reported precision, and their mean F1 scores and relative errors are nearly indistinguishable. MABCD achieves the shortest average running time, which is approximately 3.7% shorter than that of ABCD and 15.5% shorter than that of FISTA. Moreover, MABCD is about 8.24 times faster than ADMM. These results show that MABCD maintains the same level of recovery accuracy as the competing methods while providing the best average computational efficiency over 50 random instances.

Table 3. Mean performance over 50 random trials for the extended Lasso problem.

Method	Objective	F1	RelErr _β	RelErr _e	Time(s)
MABCD	5.565720	0.9986	8.098×10^{-4}	2.368×10^{-3}	0.7649
ABCD	5.565720	0.9986	8.110×10^{-4}	2.371×10^{-3}	0.7945
FISTA	5.565720	0.9986	8.111×10^{-4}	2.371×10^{-3}	0.9049
ADMM	5.565720	0.9986	8.093×10^{-4}	2.367×10^{-3}	6.3008

4.2. Ill-Conditioned Non-Negative Sparse Regression Problem

Consider the following two-block non-negative sparse regression problem:

$$\min_{u \in \mathbb{R}^{p_1}, v \in \mathbb{R}^{p_2}} F(u, v) := \frac{1}{2} \|Au + Bv - b\|_2^2 + \lambda_u \mathbf{1}^\top u + \lambda_v \mathbf{1}^\top v, \quad u \geq 0, v \geq 0, \quad (51)$$

where $A \in \mathbb{R}^{n \times p_1}$ and $B \in \mathbb{R}^{n \times p_2}$ are two data blocks, $b \in \mathbb{R}^n$ is the observation vector, and $\lambda_u, \lambda_v \geq 0$ are regularization parameters. This problem falls into the two-block composite convex optimization framework, with smooth part

$$h(u, v) = \frac{1}{2} \|Au + Bv - b\|_2^2,$$

and nonsmooth convex parts

$$p(u) = \lambda_u \mathbf{1}^\top u + \delta_{\mathbb{R}_+^{p_1}}(u), \quad q(v) = \lambda_v \mathbf{1}^\top v + \delta_{\mathbb{R}_+^{p_2}}(v),$$

where $\delta_{\mathbb{R}_+^d}$ denotes the indicator function of the non-negative orthant. The Hessian of h is

$$\mathcal{Q} = \nabla^2 h = \begin{pmatrix} A^\top A & A^\top B \\ B^\top A & B^\top B \end{pmatrix}.$$

The following majorization scheme is adopted to construct the majorized model:

$$\hat{\mathcal{Q}} = \mathcal{Q} + \text{Diag}(\rho_u I, \rho_v I), \quad \rho_u, \rho_v > 0.$$

The corresponding block subproblems are

$$\tilde{u}^k \approx \arg \min_{u \geq 0} \left\{ \frac{1}{2} \|Au + Bv^k - b\|_2^2 + \lambda_u \mathbf{1}^\top u + \frac{\rho_u}{2} \|u - u^k\|_2^2 \right\}, \quad (52)$$

and

$$\tilde{v}^k \approx \arg \min_{v \geq 0} \left\{ \frac{1}{2} \|A\tilde{u}^k + Bv - b\|_2^2 + \lambda_v \mathbf{1}^\top v + \frac{\rho_v}{2} \|v - v^k\|_2^2 \right\}. \quad (53)$$

Both are strongly convex non-negative quadratic programs. They are solved by an active-set Newton-type method. For MABCD, the inner problems are solved to a fixed high accuracy; for iMABCD, the inner stopping tolerance is chosen as

$$\varepsilon_k = \frac{C}{(k+1)^{2+\gamma}}, \quad C > 0, \quad \gamma > 0. \quad (54)$$

In our experiments, we set $n = 2000$ and $p_1 = p_2 = 500$. The full design matrix $X = [A \ B] \in \mathbb{R}^{n \times (p_1 + p_2)}$ is generated through a prescribed singular value decomposition $X = U\Sigma V^\top$, where the singular values decay geometrically from 1 to $1/\kappa_X$. With $\kappa_X = 10^5$, the spectral condition number of $X^\top X$ is approximately 10^{10} . This construction produces a severely ill-conditioned test problem. The vectors u^* and v^* are non-negative and sparse, with sparsity ratios 10% and 5%, respectively. Their support sets are sampled uniformly at random, and the nonzero entries are generated from $u_i^* \sim \mathcal{U}(1, 11)$, $v_i^* \sim \mathcal{U}(3, 5)$. The observation vector is

$$b = Au^* + Bv^* + \xi, \quad \xi \sim \mathcal{N}(0, \sigma^2 I_n), \quad \sigma = 10^{-2}.$$

We set $\lambda_u = \lambda_v = 10^{-3}$, $C = 0.5$, $\gamma = 0.5$, and $K_{\max} = 3000$. The first-order optimality conditions are monitored through the projected KKT residual

$$\text{KKT}(u, v) = \max\{\|u - \max(0, u - g_u)\|_\infty, \|v - \max(0, v - g_v)\|_\infty\}, \quad (55)$$

where

$$g_u = A^\top (Au + Bv - b) + \lambda_u \mathbf{1}, \quad g_v = B^\top (Au + Bv - b) + \lambda_v \mathbf{1}.$$

The algorithms terminate when $\text{KKT}(u^k, v^k) \leq 10^{-6}$ or when $K_{\max}$ is reached.

Table 4 reports the results for one representative ill-conditioned instance. ABCD and its variants are substantially more effective than FISTA and ADMM for this problem. MABCD improves upon ABCD in both iteration count and CPU time, while iMABCD achieves the best overall performance, requiring approximately 18.6% less CPU time than ABCD. All three methods attain nearly identical objective values and KKT residuals. In contrast, FISTA reaches the maximum iteration limit without satisfying the stopping criterion, whereas ADMM converges but incurs the largest computational cost. These results demonstrate that exploiting the two-block structure, together with the inexact strategy, improves computational efficiency on ill-conditioned problems without sacrificing solution accuracy.

Table 4. Single-instance performance comparison on ill-conditioned non-negative sparse regression problem.

Method	Iter	Time(s)	Objective	KKT
ABCD	279	0.6434	5.779102×10^{-1}	9.90×10^{-7}
MABCD	260	0.6072	5.779102×10^{-1}	9.61×10^{-7}
iMABCD	254	0.5235	5.779102×10^{-1}	9.91×10^{-7}
FISTA	3000	1.9894	5.779103×10^{-1}	1.58×10^{-6}
ADMM	2723	6.7648	5.779102×10^{-1}	9.98×10^{-7}

Figures 3 and 4 display the objective gap versus iteration number and CPU time, respectively. The accepted objective sequences of MABCD, and iMABCD are nonincreasing, whereas the ABCD curves exhibit upward steps during the iteration. The CPU time comparison shows that iMABCD reaches the prescribed objective accuracy earlier than the other methods, indicating that relaxed inner solves are beneficial when the cost of solving the curvature-aware subproblems is non-negligible.

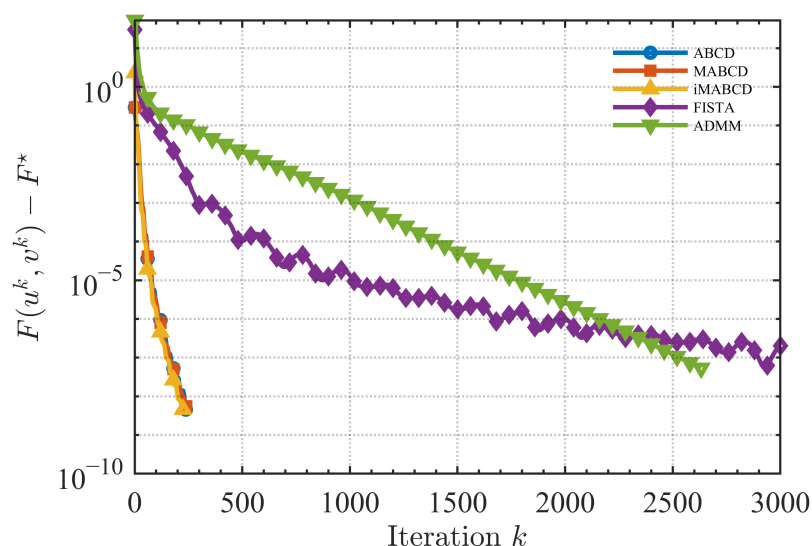


Figure 3. Objective gap versus iteration number for ill-conditioned non-negative sparse regression problem.

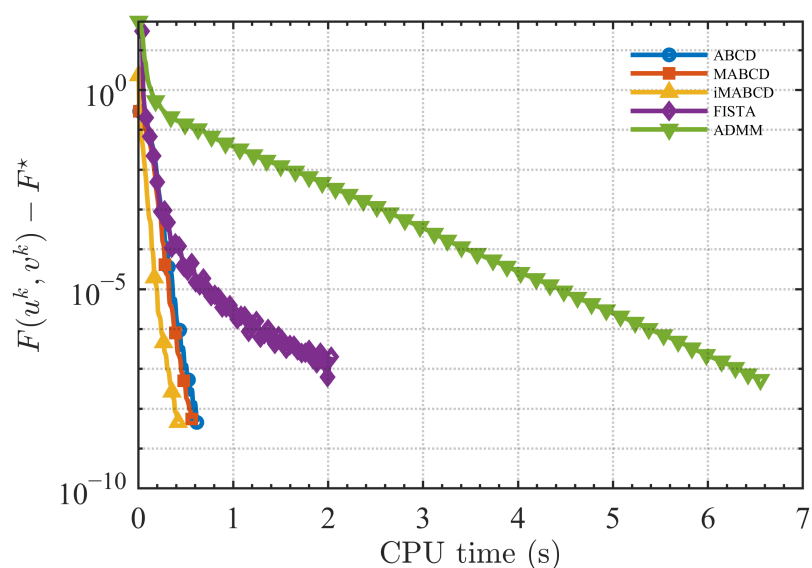


Figure 4. Objective gap versus CPU time for ill-conditioned non-negative sparse regression problem.

5. Conclusions

This paper proposes two improved alternating block coordinate descent algorithms, namely, MABCD and iMABCD, for solving two-block composite convex optimization problems. To overcome the non-monotonic fluctuation issue of the original ABCD method, MABCD incorporates a monotone descent constraint to guarantee stable nonincreasing objective updates throughout the optimization process. Furthermore, the developed iMABCD algorithm relaxes the exact subproblem solving requirement, allowing inexact iterative updates to reduce computational cost and enhance practical flexibility. Under standard assumptions, rigorous theoretical analysis demonstrates that both proposed algorithms achieve the optimal $O(1/k^2)$ convergence complexity, while the monotone mechanism effectively stabilizes the iteration process without sacrificing convergence accuracy. This work provides a stable and flexible alternative for block coordinate optimization, and future research will extend the monotone inexact framework to nonconvex optimization and more general structured learning problems.

Author Contributions: Conceptualization, Z.C. and X.S.; methodology, Z.C. and X.S.; software, H.C.; validation, H.C.; formal analysis, Z.C.; writing—original draft preparation, Z.C.; writing—review and editing, Z.C. and X.S.; visualization, H.C.; supervision, Z.C. and X.S.; funding acquisition, Z.C. and X.S. All authors have read and agreed to the published version of the manuscript.

Funding: Zixuan Chen’s work was supported by the National Natural Science Foundation of China (grant No.12501519). Xiaoliang Song’s work was supported by the National Key R&D Program of China (grant No.2023YFA1011303), the National Natural Science Foundation of China (grant No.12301479), and the General Program of Natural Science Foundation under Liaoning Provincial Science and Technology Plan Joint Plan (grant No.2024-MSLH-063).

Data Availability Statement: Dataset and code will be made available upon request by the authors.

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

References

1. Parikh, N.; Boyd, S. Proximal Algorithms. *Found. Trends Optim.* **2014**, *1*, 127–239. <https://doi.org/10.1561/2400000003>.
2. 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.* **2011**, *3*, 1–122. <https://doi.org/10.1561/2200000016>.
3. Beck, A.; Tetruashvili, L. On the Convergence of Block Coordinate Descent Type Methods. *SIAM J. Optim.* **2013**, *23*, 2037–2060. <https://doi.org/10.1137/120887679>.
4. Nesterov, Y. Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems. *SIAM J. Optim.* **2012**, *22*, 341–362. <https://doi.org/10.1137/100802001>.
5. Sun, D.; Toh, K.-C.; Yang, L. An Efficient Inexact ABCD Method for Least Squares Semidefinite Programming. *SIAM J. Optim.* **2016**, *26*, 2792–2823. <https://doi.org/10.1137/15M1021799>.
6. Combettes, P.L.; Wajs, V.R. Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168–1200. <https://doi.org/10.1137/050626090>.
7. Nikolova, M.; Tan, P. Alternating Structure-Adapted Proximal Gradient Descent for Nonconvex Nonsmooth Block-Regularized Problems. *SIAM J. Optim.* **2019**, *29*, 2053–2078. <https://doi.org/10.1137/17M1142624>.
8. Kim, J.; He, Y.; Park, H. Algorithms for Nonnegative Matrix and Tensor Factorizations: A Unified View Based on Block Coordinate Descent Framework. *J. Glob. Optim.* **2014**, *58*, 285–319. <https://doi.org/10.1007/s10898-013-0035-4>.
9. Grippo, L.; Sciandrone, M. On the Convergence of the Block Nonlinear Gauss-Seidel Method under Convex Constraints. *Oper. Res. Lett.* **2000**, *26*, 127–136. [https://doi.org/10.1016/S0167-6377\(99\)00074-7](https://doi.org/10.1016/S0167-6377(99)00074-7).
10. Lin, Q.; Lu, Z.; Xiao, L. An Accelerated Randomized Proximal Coordinate Gradient Method and Its Application to Regularized Empirical Risk Minimization. *SIAM J. Optim.* **2015**, *25*, 2244–2273. <https://doi.org/10.1137/141000270>.
11. Cui, Y.; Sun, D.F.; Toh, K.C. Computing the Best Approximation Over the Intersection of a Polyhedral Set and the Doubly Nonnegative Cone. *SIAM J. Optim.* **2019**, *29*, 2785–2813. <https://doi.org/10.1137/18M1175562>.
12. Lin, Q.; Xu, Y. Reducing the Complexity of Two Classes of Optimization Problems by Inexact Accelerated Proximal Gradient Method. *SIAM J. Optim.* **2023**, *33*, 1–35. <https://doi.org/10.1137/22M1469584>.

13. Fercoq, O.; Richtárik, P. Accelerated, Parallel, and Proximal Coordinate Descent. *SIAM J. Optim.* **2015**, *25*, 1997–2023. <https://doi.org/10.1137/130949993>.
14. Tappenden, R.; Richtárik, P.; Gondzio, J. Inexact Coordinate Descent: Complexity and Preconditioning. *J. Optim. Theory Appl.* **2016**, *170*, 144–176. <https://doi.org/10.1007/s10957-016-0867-4>.
15. Beck, A.; Teboulle, M. Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems. *IEEE Trans. Image Process.* **2009**, *18*, 2419–2434. <https://doi.org/10.1109/TIP.2009.2028250>.
16. Chen, Z.; Song, X.; Chen, X.; Yu, B. A Warm-Start FE-dABCD Algorithm for Elliptic Optimal Control Problems with Constraints on the Control and the Gradient of the State. *Comput. Math. Appl.* **2024**, *161*, 1–12. <https://doi.org/10.1016/j.camwa.2024.02.038>.
17. Song, X.; Chen, B.; Yu, B. An Efficient Duality-Based Approach for PDE-Constrained Sparse Optimization. *Comput. Optim. Appl.* **2018**, *71*, 663–690. <https://doi.org/10.1007/s10589-017-9951-4>.
18. Jiang, K.; Sun, D.; Toh, K.-C. An Inexact Accelerated Proximal Gradient Method for Large Scale Linearly Constrained Convex SDP. *SIAM J. Optim.* **2012**, *22*, 1042–1064. <https://doi.org/10.1137/110847081>.
19. Nguyen, N.H.; Tran, T.D. Robust Lasso With Missing and Grossly Corrupted Observations. *IEEE Trans. Inf. Theory* **2013**, *59*, 2036–2058. <https://doi.org/10.1109/TIT.2012.2232347>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.