An Inexact Coordinate Descent Method for the Weighted l_1-regularized Convex Optimization Problem

Xiaoqin Hua* Nobuo Yamashita†

September 24, 2012

Abstract. The purpose of this paper is to propose an inexact coordinate descent (ICD) method for solving the weighted l_1-regularized convex optimization problem with a box constraint. The proposed algorithm solves a one dimensional subproblem inexactly at each iteration. We give criteria of the inexactness under which the sequence generated by the proposed method converges to an optimal solution and its convergence rate is at least R-linear without assuming the uniqueness of solutions.
Keywords. l_1-regularized convex optimization, inexact coordinate descent method, linear convergence, error bound.

1 Introduction

We consider the following weighted l_1-regularized convex optimization problem:

$$\begin{aligned} &\text{minimize} \;\; F(x) := g(Ax) + \langle b, x\rangle + \sum_{i=1}^{n} \tau_i |x_i| \\ &\text{subject to} \;\; l \le x \le u, \end{aligned} \tag{1.1}$$

where $g : \mathcal{R}^m \to (-\infty, \infty]$ is a strictly convex function on $\operatorname{dom} g$, $A \in \mathcal{R}^{m\times n}$ and $b \in \mathcal{R}^n$. Moreover, τ, l and u are n-dimensional vectors such that $l_i \in [-\infty, \infty)$, $u_i \in (-\infty, \infty]$, $\tau_i \in [0, \infty)$ and $l_i < u_i$ for each $i = 1, \cdots, n$. The nonnegative scalar constant τ_i is called weight and the term $\sum_{i=1}^{n} \tau_i |x_i|$ is called the l_1-regularization function. For convenience, we denote the differentiable term of F by f, that is, $f(x) := g(Ax) + \langle b, x\rangle$.

*School of Mathematics and Physics, Jiangsu University of Science and Technology, Zhenjiang 212003, CHINA. Current address: Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, JAPAN. E-mail: hua.xiaoqin.22r@st.kyoto-u.ac.jp

†Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, JAPAN. E-mail:nobuo@i.kyoto-u.ac.jp

The problem (1.1) contains many well-known problems as special cases [7, 16, 15]. When $\tau_i = 0$ for all index i, the problem (1.1) is reduced to the differentiable convex problem with a box constraint. When $l_i = -\infty$ and $u_i = \infty$ for each i, it is reduced to the unconstrained l_1-regularized convex problem. When τ_i is a fixed positive constant $\hat{\tau}$ for all i, $b = 0$ and $g(y) := \frac{1}{2}\|y - z\|^2$ with some $z \in \mathcal{R}^m$, it is the famous l_1-l_2 problem. Another important special case is the l_1-regularized logistic regression problem where g is given by $g(y) := \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-y_i))$. Each special case has wide applications in the real life such as the compressed sensing [15], the feature selection in the data classification [7], the data mining [9], geophysics [1] and so on. Typically, the scales of these weighted l_1-regularized convex optimization problems are very large and the objective functions are not differentiable everywhere due to the regularization function. Moreover, the optimal solutions are possibly not unique because the matrix A may not have full column rank. Thus the Newton-type methods such as the interior point method can not be applied directly.

In the past, the coordinate descent (CD) method is verified to be one of the feasible methods for the large scale optimization problems [4, 11, 14, 16]. The CD method minimizes the objective function with respect to one variable while all the others are fixed at each iteration. The idea of this method is very simple and the storage of calculations is little. In some special cases, it can be implemented in parallel. Luo and Tseng [16] proved its global and linear convergence for a smooth problem, that is, $\tau_i = 0$ for all i. Note that the problem (1.1) can be reformulated as a smooth problem (see the problem (2.6) in Section 2). However, the reformulated problem has twice variables. In 2001, Tseng [11] showed the global convergence of a block coordinate descent (BCD) method for minimizing an nondifferentiable function with certain separability. But the exact minimizers of the subproblem must be found on each iteration in [11, 16]. It is possible for the l_1-l_2 problem, while usually it is hard for the general l_1-regularized convex problem.

In this paper, we propose an inexact coordinate descent (ICD) method and extend the results of Luo and Tseng [16]. Roughly speaking, we extend in the following three aspects:

- The smooth convex problem is extended to that with the l_1-regularized function.
- On each iteration step, we accept an inexact solution of a subproblem instead of the exact solution.
- The linear convergence rate is extended to the nonsmooth problem.

In this paper, under the same basic assumptions as in [16], we show that the proposed ICD method is not only globally convergent but also with at least R-linear convergence rate under the almost cycle rule.

This paper is organized as follows. In Section 2, we derive the optimal conditions of the problem (1.1) and also define the ε-optimality conditions which are related to an inexact solution. In Section 3, we present a framework of the ICD method and make some assumptions for the “inexact solutions”. The global convergence and linear convergence rate are established in Section 4. Finally, we conclude this paper in Section 5.

Throughout this paper, we use the following notations. For a differentiable function h, ∇h denotes the gradient of h and $\nabla^2 h$ denotes the Hessian matrix of h. $\nabla_i h$ denotes the ith coordinate of the gradient vector ∇h. If h is convex and nondifferentiable, ∂h denotes the subdifferential of h. For any real number x, $|x|$ denotes the absolute value of x. For a given vector $x \in \mathcal{R}^n$, we denote by x_i the ith coordinate of x. We denote the 2-norm of x by $\|x\|$. For any matrix A, the transpose of A is denoted by A^T and A_j denotes the jth column. For the function $F : \mathcal{R}^n \to \mathcal{R}$ and a vector $x \in \mathcal{R}^n$, we sometimes use a notation $F(x_1, \cdots, x_n)$ instead of $F(x)$.

2 Preliminaries

Throughout the paper, we make the following basic assumptions for the problem (1.1).

Assumption 2.1. *For the problem (1.1), we assume that*

(a) A_j *is a nonzero vector for all* $j \in \{1, 2, \cdots, n\}$.

(b) $l_i < 0 < u_i$ *for all* $i \in \{1, 2, \cdots, n\}$.

(c) The set of the optimal solutions, denoted by X^*, *is nonempty.*

(d) The effective domain of g, *denoted by* $\operatorname{dom} g$, *is nonempty and open.*

(e) g *is twice continuously differentiable on* $\operatorname{dom} g$.

(f) $\nabla^2 g(Ax^*)$ *is positive definite for every optimal solution* $x^* \in X^*$.

We make a few remarks on these assumptions. In Part (a), if A_j is zero, then x_j^* of the optimal solution x^* can be easily determined. Thus we can remove x_j from the problem (1.1). Part (b) is just for simplification. If both l_i and u_i are positive for some $i \in \{1, 2, \cdots, n\}$, we may replace x_i, l_i and u_i by $\bar{x}_i + \frac{l_i + u_i}{2}$, $\frac{l_i - u_i}{2}$ and $\frac{u_i - l_i}{2}$. Then the problem (1.1) is reformulated into the case without l_1-regularized term for the index i. If g is strongly convex and twice differentiable on $\operatorname{dom} g$, then Part (e) and (f) are satisfied automatically. For example, a quadratic function, an exponential function, and even some complicate functions in the l_1-regularized logistic regression problem satisfy (e) and (f). Note that we do not assume the boundness of the optimal solution set X^*.

Next, we present some properties under Assumption 2.1 that are used in the subsequent sections. From Assumption 2.1(e) and (f), there exists a sufficiently small closed neighborhood $B(Ax^*)$ of Ax^* such that $B(Ax^*) \subseteq \operatorname{dom} g$ and $\nabla^2 g$ is positive definite in $B(Ax^*)$. Furthermore, it implies that g is strongly convex in $B(Ax^*)$, i.e., there exists a scalar $\sigma > 0$ such that

$$g(y) - g(z) - \langle \nabla g(z), y - z \rangle \geq \sigma \|y - z\|^2, \ \forall y, z \in B(Ax^*). \tag{2.1}$$

2.1 Optimality conditions

The KKT conditions [13] for the problem (1.1) are described as follows.

$$\begin{aligned} &\nabla_i f(x) + \tau_i \partial |x_i| - \mu_i + \nu_i \ni 0, \\ &x_i \geq l_i, \mu_i \geq 0, \mu_i(x_i - l_i) = 0, \quad i = 1, \cdots, n, \\ &x_i \leq u_i, \nu_i \geq 0, \nu_i(u_i - x_i) = 0, \end{aligned} \tag{2.2}$$

where $\partial | \cdot |$ is the subdifferential of the absolute value function. Since the problem (1.1) is convex, x satisfying (2.2) is a solution of the problem (1.1). The KKT conditions (2.2) can be rewritten as follows.

Lemma 2.1. *A vector x is an optimal solution of the problem (1.1) if and only if one of the following statements holds for each i.*

(i) $\nabla_i f(x) \geq \tau_i$ *and* $x_i = l_i$.

(ii) $\nabla_i f(x) = \tau_i$ *and* $l_i \leq x_i \leq 0$.

(iii) $|\nabla_i f(x)| \leq \tau_i$ *and* $x_i = 0$.

(iv) $\nabla_i f(x) = -\tau_i$ *and* $0 \leq x_i \leq u_i$.

(v) $\nabla_i f(x) \leq -\tau_i$ *and* $x_i = u_i$.

Next, we represent these conditions as a fixed point of some operator. To this end, we first define a mapping $T_\tau : \mathcal{R}^n \to \mathcal{R}^n$ as

$$T_\tau(x)_i := (|x_i| - \tau_i)_+ \operatorname{sgn}(x_i), \tag{2.3}$$

where the scalar function $(a)_+$ is defined by $(a)_+ := \max(0, a)$ and $\mathrm{sgn}(a)$ is a sign function defined as follows:

$$\mathrm{sgn}(a) := \begin{cases} -1 & \text{if } a < 0, \\ 0 & \text{if } a = 0, \\ 1 & \text{if } a > 0. \end{cases}$$

It can be verified that T_τ is nonexpensive, i.e., $\|T_\tau(y) - T_\tau(z)\| \leq \|y - z\|$, for any $y, z \in \mathrm{dom}\, F$.

Let $[x]^+_{[l,u]}$ denote the orthogonal projection of a vector x onto the box $[l, u]$. This projection is also nonexpensive and its ith coordinate can be written as $[x_i]^+_{[l_i,u_i]} := \mathrm{mid}\{x_i, l_i, u_i\}$, where $\mathrm{mid}\{x_i, l_i, u_i\}$ is defined by $\mathrm{mid}\{x_i, l_i, u_i\} := \max\{l_i, \min\{u_i, x_i\}\}$.

By using the mappings T_τ and $[\cdot]^+_{[l,u]}$, we define a mapping $P_{\tau,l,u}(x) : \mathcal{R}^n \to \mathcal{R}^n$ by

$$P_{\tau,l,u}(x) := [T_\tau(x - \nabla f(x))]^+_{[l,u]}. \tag{2.4}$$

Since $[x]^+_{[l,u]}$ and T_τ are nonexpensive, we have that

$$\|P_{\tau,l,u}(y) - P_{\tau,l,u}(z)\| \leq \|y - z - \nabla f(y) + \nabla f(z)\|, \ \forall y, z \in \mathrm{dom}\, F. \tag{2.5}$$

Now, the optimal solutions can be described as a fixed point of the mapping $P_{\tau,l,u}$.

Theorem 2.1. *For the problem (1.1), a vector x belongs to the optimal solution set X^* if and only if $x = P_{\tau,l,u}(x)$, i.e., $X^* = \{x|\ x \in \mathrm{dom}\, g, x = P_{\tau,l,u}(x)\}$.*

Proof. This theorem is a direct consequence of Theorem 2.2 that will be shown in Subsection 2.2. □

Since the solution set X^* is not necessarily bounded, the level set of F may be not bounded. Nevertheless, as an extension of [16, Lemma 3.3], we can show the compactness of the set $\Omega(\zeta) := \{t|\ t = Ax, F(x) \leq \zeta, x \in [l, u]\}$,

Lemma 2.2. *For a given constant value ζ, the set $\Omega(\zeta)$ is a compact subset of $\mathrm{dom}\, g$.*

Proof. The l_1-regularized convex problem (1.1) can be transformed into a smooth optimization problem with box constraints:

$$\begin{aligned} \underset{x^+,\, x^- \in \mathcal{R}^n}{\text{minimize}} \quad & \bar{F}(x^+, x^-) := g(Ax^+ - Ax^-) + \langle b, x^+ - x^- \rangle + \sum_{i=1}^{n} \tau_i (x_i^+ + x_i^-) \\ \text{subject to} \quad & 0 \leq x_i^+ \leq u_i, i = 1, \cdots, n; \\ & 0 \leq x_i^- \leq |l_i|, i = 1, \cdots, n. \end{aligned} \tag{2.6}$$

Note that if (x^+, x^-) is feasible for the problem (2.6), then $x = x^+ - x^-$ is also feasible for the problem (1.1) due to $l \leq x \leq u$.

Let $\bar{\Omega}(\zeta)$ be defined as follows.

$$\begin{aligned}\bar{\Omega}(\zeta) &:= \{Ax^+ - Ax^- \mid \bar{F}(x^+, x^-) \le \zeta, x^+ \in [0, u], x^- \in [0, |l|]\} \\ &= \{Ax \mid x = x^+ - x^-, \bar{F}(x^+, x^-) \le \zeta, x^+ \in [0, u], x^- \in [0, |l|]\},\end{aligned}$$

where $|l| = (|l_1|, \cdots, |l_n|)^T$. Then $\bar{\Omega}(\zeta)$ is a compact set of $\operatorname{dom} g$ from Appendix in [16].

In the rest part, we only need to show $\bar{\Omega}(\zeta) = \Omega(\zeta)$. In fact, for every $t \in \bar{\Omega}(\zeta)$, there exists (x, x^+, x^-) such that $t = Ax, x = x^+ - x^-, \bar{F}(x^+, x^-) \le \zeta, x^+ \in [0, u], x^- \in [0, |l|]$. Then we have $x \in [l, u]$ and $\zeta \ge \bar{F}(x^+, x^-) \ge F(x)$. It further implies $t \in \Omega(\zeta)$, i.e., $\bar{\Omega}(\zeta) \subseteq \Omega(\zeta)$. Conversely, for every $t \in \Omega(\zeta)$, there exists a x such that $t = Ax, F(x) \le \zeta, x \in [l, u]$. Let $x_i^+ := \max\{x_i, 0\}$, and $x_i^- := \max\{-x_i, 0\}$ for each i. Then we have $x^+ \in [0, u], x^- \in [0, |l|]$, $x = x^+ - x^-$, and $\bar{F}(x^+, x^-) = F(x)$. Therefore, we deduce $t \in \bar{\Omega}(\zeta)$, which implies $\Omega(\zeta) \subseteq \bar{\Omega}(\zeta)$. Consequently, the relation $\bar{\Omega}(\zeta) = \Omega(\zeta)$ holds. $\square$

Next, we show that ∇g is Lipschitz continuous on some compact set including $\Omega(\zeta)$. For this purpose, we define a set $\Omega(\zeta) + B(\epsilon_0)$ as $\Omega(\zeta) + B(\epsilon_0) := \{p + v \mid p \in \Omega(\zeta),\ \|v\| \le \epsilon_0\}$, where ϵ_0 is a positive constant. It is easy to see that the set $\Omega(\zeta) + B(\epsilon_0)$ is a compact set.

Lemma 2.3. *There exist constants $L > 0$ and $\epsilon_0 > 0$ such that $\Omega(\zeta) + B(\epsilon_0) \subseteq \operatorname{dom} g$ and $\|\nabla g(y) - \nabla g(z)\| \le L\|y - z\|$ for all $y,\ z \in \Omega(\zeta) + B(\epsilon_0)$.*

Proof. Since $\Omega(\zeta)$ is closed from Lemma 2.2 and $\operatorname{dom} g$ is open, there exists a positive constant ϵ_0 such that $\Omega(\zeta) + B(\epsilon_0) \subseteq \operatorname{dom} g$. Furthermore, since g is twice continuously differentiable on $\operatorname{dom} g$ and $\Omega(\zeta) + B(\epsilon_0)$ is compact, $\nabla^2 g(x)$ is bounded in $\Omega(\zeta) + B(\epsilon_0)$, that is to say, there exists a constant $L > 0$ such that $\|\nabla^2 g(x)\| \le L$ for all $x \in \Omega(\zeta) + B(\epsilon_0)$. Then, this lemma holds from the mean value theorem. $\square$

Similar to [17, Lemma 2.1], we can prove the following invariant property of the optimal solution set X^*. For simplicity, we omit the proof here.

Lemma 2.4. *For any $x^*, y^* \in X^*$, $Ax^* = Ay^*$.*

2.2 ε-optimality conditions

In this subsection, we give a definition of a relaxed optimality conditions and a relation between the conditions and the mapping $P_{\tau,l,u}$.

Definition 2.1. *We say that the ε-optimality conditions for the problem (1.1) hold at x if one of the following statements holds for each i.*

(i) $\nabla_i f(x) - \tau_i \ge -\varepsilon$ *and* $|x_i - l_i| \le \varepsilon$.

(ii) $|\nabla_i f(x) - \tau_i| \leq \varepsilon$ *and* $l_i - \varepsilon \leq x_i \leq \varepsilon$.

(iii) $|\nabla_i f(x)| \leq \tau_i + \varepsilon$ *and* $|x_i| \leq \varepsilon$.

(iv) $|\nabla_i f(x) + \tau_i| \leq \varepsilon$ *and* $-\varepsilon \leq x_i \leq u_i + \varepsilon$.

(v) $\nabla_i f(x) + \tau_i \leq \varepsilon$ *and* $|x_i - u_i| \leq \varepsilon$.

Definition 2.2. *We say x is an ε-approximate solution of the problem (1.1) if the ε-optimality conditions hold at x.*

Note that the optimality condition in Lemma 2.1 can be obtained by Definition 2.1 with $\varepsilon = 0$.

For convenience, we define the following five index sets:

$$\begin{aligned}
J_1(x, \varepsilon) &:= \{i |\ \nabla_i f(x) - \tau_i \geq -\varepsilon, |x_i - l_i| \leq \varepsilon\};\\
J_2(x, \varepsilon) &:= \{i |\ |\nabla_i f(x) - \tau_i| \leq \varepsilon, l_i - \varepsilon \leq x_i \leq \varepsilon\};\\
J_3(x, \varepsilon) &:= \{i |\ |\nabla_i f(x)| \leq \tau_i + \varepsilon, |x_i| \leq \varepsilon\};\\
J_4(x, \varepsilon) &:= \{i |\ |\nabla_i f(x) + \tau_i| \leq \varepsilon, -\varepsilon \leq x_i \leq u_i + \varepsilon\};\\
J_5(x, \varepsilon) &:= \{i |\ \nabla_i f(x) + \tau_i \leq \varepsilon, |x_i - u_i| \leq \varepsilon\}.
\end{aligned}$$

Then the ε-optimality conditions hold at x if and only if $\bigcup_{i=1}^{5} J_i(x, \varepsilon) = \{1, 2, \cdots, n\}$.

Throughout the paper, for simplicity, we assume that

$$\varepsilon < \frac{1}{2} \min_i \{-l_i, u_i\}. \tag{2.7}$$

The next theorem gives an equivalent description of the ε-optimality conditions, which will be used for constructing an inexact CD method and investigating its convergence properties.

Theorem 2.2. *The ε-optimality conditions hold at x if and only if $|x_i - P_{\tau,l,u}(x)_i| \leq \varepsilon$ holds for each i.*

Proof. By the definition of $T_\tau(x)$ and $P_{\tau,l,u}(x)$ in (2.3) and (2.4), we have that

$$\begin{aligned}
|x_i - P_{\tau,l,u}(x)_i| &= |x_i - \text{mid}\{l_i, u_i, \max\{0, |x_i - \nabla_i f(x)| - \tau_i\}\text{sgn}(x_i - \nabla_i f(x))\}|\\
&= \begin{cases}
|x_i - l_i| & \text{if } x_i - \nabla_i f(x) \in (-\infty, l_i - \tau_i],\\
|\nabla_i f(x) - \tau_i| & \text{if } x_i - \nabla_i f(x) \in (l_i - \tau_i, -\tau_i],\\
|x_i| & \text{if } x_i - \nabla_i f(x) \in (-\tau_i, \tau_i],\\
|\nabla_i f(x) + \tau_i| & \text{if } x_i - \nabla_i f(x) \in (\tau_i, u_i + \tau_i],\\
|x_i - u_i| & \text{if } x_i - \nabla_i f(x) \in (u_i + \tau_i, \infty).
\end{cases}
\end{aligned} \tag{2.8}$$

We firstly consider the " if " part of this theorem. It is sufficient to show that if $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$ holds for each $i \in \{1, 2, \cdots, n\}$, then for each $i \in \{1, 2, \cdots, n\}$ there exists a $j \in \{1, 2, \cdots, 5\}$ such that $i \in J_j(x, \varepsilon)$. We can prove this according to the distinct cases in (2.8). If $x_i - \nabla_i f(x) \in (-\infty, l_i - \tau_i]$, then it follows from $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$ and (2.8) that $|x_i - P_{\tau,l,u}(x)_i| = |x_i - l_i| \le \varepsilon$, that is, $x_i - l_i \ge -\varepsilon$. Moreover, since $x_i - \nabla_i f(x) \in (-\infty, l_i - \tau_i]$ implies that $\nabla_i f(x) - \tau_i \ge x_i - l_i$, we have $\nabla_i f(x) - \tau_i \ge -\varepsilon$. Therefore, $i \in J_1(x, \varepsilon)$ holds. Similarly, we can show that if $x_i - \nabla_i f(x)$ is located in other intervals, the corresponding results also hold.

Conversely, suppose that x is an ε-approximate solution, i.e., for each $i \in \{1, 2, \cdots, n\}$, there exists a $j \in \{1, 2, \cdots, 5\}$ such that $i \in J_j(x, \varepsilon)$. Thus, it is sufficient to show that for each i and j such that $i \in J_j(x, \varepsilon)$, the condition $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$ holds.

Case 1: $i \in J_1(x, \varepsilon)$ **or** $i \in J_5(x, \varepsilon)$. First suppose $i \in J_1(x, \varepsilon)$. Then we have

$$\nabla_i f(x) - \tau_i \ge -\varepsilon \text{ and } |x_i - l_i| \le \varepsilon. \tag{2.9}$$

They imply that $x_i - \nabla_i f(x) \le l_i - \tau_i + 2\varepsilon$. It then follows from (2.7) that $x_i - \nabla_i f(x) \in (-\infty, -\tau_i)$. Thus, we focus on (2.8) in two intervals $(-\infty, l_i - \tau_i]$ and $(l_i - \tau_i, -\tau_i]$. If $x_i - \nabla_i f(x) \in (-\infty, l_i - \tau_i]$, it follows from (2.8) that $|x_i - P_{\tau,l,u}(x)_i| = |x_i - l_i|$. Then the inequality $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$ holds due to (2.9). If $x_i - \nabla_i f(x) \in (l_i - \tau_i, -\tau_i]$, then we have $\nabla_i f(x) - \tau_i < x_i - l_i$ and $|x_i - P_{\tau,l,u}(x)_i| = |\nabla_i f(x) - \tau_i|$, which together with (2.9) implies $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$. A symmetric argument can prove the case with $i \in J_5(x, \varepsilon)$.

Case 2: $i \in J_2(x, \varepsilon)$ **or** $i \in J_4(x, \varepsilon)$. First suppose $i \in J_2(x, \varepsilon)$. Then we have

$$|\nabla_i f(x) - \tau_i| \le \varepsilon \text{ and } l_i - \varepsilon \le x_i \le \varepsilon. \tag{2.10}$$

We obtain $-\tau_i - \varepsilon \le -\nabla_i f(x) \le \varepsilon - \tau_i$ from the first inequality, which yields $l_i - \tau_i - 2\varepsilon \le x_i - \nabla_i f(x) \le 2\varepsilon - \tau_i$ with the second inequality of (2.10) . With the assumption (2.7) on ε, we have $x_i - \nabla_i f(x) \in [l_i - \tau_i - 2\varepsilon, u_i)$. Dividing the interval $[l_i - \tau_i - 2\varepsilon, u_i)$ into $[l_i - \tau_i - 2\varepsilon, l_i - \tau_i]$, $(l_i - \tau_i, -\tau_i]$, $(-\tau_i, \tau_i]$ and (τ_i, u_i), we can show $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$ by (2.8) and (2.10) in a way similar to Case 1. In the case where $i \in J_4(x, \varepsilon)$, a similar analysis shows $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$.

Case 3: $i \in J_3(x, \varepsilon)$. In this case, we have $|\nabla_i f(x)| \le \tau_i + \varepsilon$ and $|x_i| \le \varepsilon$. They imply $-\tau_i - 2\varepsilon \le x_i - \nabla_i f(x) \le \tau_i + 2\varepsilon$. Moreover, we have by (2.7) that $l_i - \tau_i < x_i - \nabla_i f(x) < u_i + \tau_i$. By dividing the interval $(l_i - \tau_i, u_i + \tau_i)$ into the following three intervals: $(l_i -$

$\tau_i, -\tau_i], (-\tau_i, \tau_i]$ and $(\tau_i, u_i + \tau_i)$, we can prove $|x_i - P_{\tau,l,u}(x)_i| \le \varepsilon$ in a way similar to Cases 1 and 2.

Upon the preceding proof, the necessary condition of this theorem is confirmed. $\square$

3 Inexact coordinate descent (ICD) method

In this section, we will first present a framework for the ICD method, and then give some assumptions for the "inexact solutions".

A general framework of the ICD method can be described as follows.

Inexact coordinate descent (ICD) method:

Step 0: Choose an initial point $x^0 \in [l, u]$ and let $r := 0$.

Step 1: If some termination condition holds, then stop.

Step 2: Choose an index $i(r) \in \{1, \cdots, n\}$, get an approximate solution $x_{i(r)}^{r+1}$ of the following one dimensional subproblem:

$$\underset{x_{i(r)} \in \{l_{i(r)} \le x_{i(r)} \le u_{i(r)}\}}{\text{minimize}} F(x_1^r, x_2^r, \cdots, x_{i(r)-1}^r, x_{i(r)}, x_{i(r)+1}^r, \cdots, x_n^r). \tag{3.1}$$

Step 3: Set $x_j^{r+1} := x_j^r$ for all $j \in \{1, \cdots, n\}$ such that $j \ne i(r)$ and let $r := r + 1$. Go to Step 1.

Note that the exact solution of the subproblem (3.1) is unique from Assumption 2.1(a) and the strict convexity of g. We use the notation $i(r)$ for the index chosen at the rth iteration. For simplicity, we will use i instead of $i(r)$ when $i(r)$ is clear from the context.

For the global convergence of the ICD method, it is important to define the inexactness of the approximate solutions of the subproblem (3.1) and choose an appropriate index $i(r)$ in Step 2.

For the inexactness, we require the following assumptions:

Assumption 3.1. *We assume that the following statements hold:*

(i) $F(x_1^r, x_2^r, \cdots, x_{i-1}^r, x_i^{r+1}, x_{i+1}^r, \cdots, x_n^r) \le \min_{x_i \in \{l_i, 0, u_i, x_i^r\}} F(x_1^r, x_2^r, \cdots, x_{i-1}^r, x_i, x_{i+1}^r, \cdots, x_n^r).$

(ii) x_i^{r+1} *is feasible, i.e.,* $x_i^{r+1} \in [l_i, u_i]$.

(iii) x_i^{r+1} *is an* ε^{r+1}*-approximate solution of the subproblem (3.1).*

(iv) ***Conditions on*** ε^{r+1}*:* $\varepsilon^{r+1} \le \min\{\delta_r, \alpha_r |x_i^{r+1} - x_i^r|, \varepsilon^r\}$*, where* $\{\delta_r\}$ *is a monotonically decreasing sequence such that* $\lim_{r\to\infty} \delta_r = 0$*, and* $\alpha_r \in [0, \bar{\alpha}]$ *with a positive constant* $\bar{\alpha}$*.*

(v) ***Conditions on*** α_r*:* $\alpha_r < \dfrac{\sigma \min\limits_j \|A_j\|^2}{L \max\limits_j \|A_j\|^2 + 1}$ *holds for sufficiently large* r*, where* σ *is a positive constant defined in (2.1), and* L *is the Lipschitz constant of* ∇g *given in Lemma 2.3.*

Here we make a simple explanation. Part (i) enforces not only that $\{F(x^r)\}$ is decreasing but also that $\{F(x^{r+1})\}$ is less than $F(x_1^r, x_2^r, \cdots, x_{i-1}^r, x_i, x_{i+1}^r, \cdots, x_n^r)$ at a point where F is nonsmooth. This condition is easy to check when computing. It also plays a key role for the convergence of $\{x^r\}$ when the objective function is not differentiable. In Part (iii), recall that the ε-optimality conditions for the one dimensional subproblem (3.1) is that one of (i)-(v) in Definition 2.1 holds at $x_{i(r)}$. The assumptions(i)-(iv) are necessary for the global convergence while the assumption (v) for α_r is used to guarantee the linear convergence rate of $\{x^r\}$.

Note that if we obtain the exact solution of the subproblem (3.1) on each iteration, then the sequence $\{x^r\}$ satisfies Assumption 3.1 automatically. Hence, the classical CD method is a special case of the ICD method.

For the choice of the coordinate $i(r)$ in Step 2 , we adopt the following almost cycle rule, which is an extension of the classical cyclic rule in [3].

Almost cyclic rule:
There exists an integer $B \geq n$, such that every coordinate is iterated upon at least once every B successive iterations.

In the next section, we will show the ICD method with almost cycle rule converges R-linearly to a solution under Assumption 2.1 and 3.1.

4 Global and linear convergence

In this section, we will show the global and linear convergence of the ICD method. Compared with the classical exact CD method, the ICD method has many “inexact” factors. Thus we need some preparations.

First of all, we illustrate a brief outline of the proof.

(1) $\lim\limits_{r\to\infty} \{x^{r+1} - x^r\} = 0$. (Lemma 4.3)

(2) $Ax^r \to Ax^*$ where x^* is one of the optimal solutions. (Theorem 4.1)

(3) Sufficient decreasing: $F(x^r) - F(x^{r+1}) \geq \eta \|x^r - x^{r+1}\|^2$ for some positive constant η. (Lemma 4.8)

(4) Error bound: $\|Ax^r - Ax^*\| \leq \kappa \|x^r - P_{\tau,l,u}(x^r)\|$ for some κ. (Lemma 4.9)

(5) Linear convergence. (Theorems 4.2 and 4.3)

Note that since it is not necessary for the matrix A to have full column rank, $Ax^r \to Ax^*$ (Theorem 4.1) does not imply $x^r \to x^*$.

For convenience, we define two vectors $\tilde{x}^{r+1}$ and x^{r+1} as follows.

$$\tilde{x}^{r+1} := (x_1^r, x_2^r, \cdots, x_{i(r)-1}^r, \tilde{x}_{i(r)}^{r+1}, x_{i(r)+1}^r, \cdots, x_n^r) \tag{4.1}$$

and

$$x^{r+1} := (x_1^r, x_2^r, \cdots, x_{i(r)-1}^r, x_{i(r)}^{r+1}, x_{i(r)+1}^r, \cdots, x_n^r), \tag{4.2}$$

where $x_{i(r)}^{r+1}$ and $\tilde{x}_{i(r)}^{r+1}$ are an ε^{r+1}-approximate solution and the exact solution of the subproblem (3.1), respectively.

In the first part of this section, we show $\lim\limits_{r\to\infty}\{F(\tilde{x}^r) - F(x^r)\} = 0$ and $\lim\limits_{r\to\infty}\{x^{r+1} - x^r\} = 0$. To this end, we need the following function $h_i : \mathcal{R}^n \times \mathcal{R}^n \to \mathcal{R}$ and Lemma 4.1.

$$\begin{aligned} h_i(y,z) :=& \nabla_i f(z)(y_i - z_i) + \tau_i(|y_i| - |z_i|) \\ =& \begin{cases} (\nabla_i f(z) + \tau_i)(y_i - z_i) & \text{if } y_i \geq 0, z_i \geq 0, \\ \nabla_i f(z)(y_i - z_i) + \tau_i(y_i + z_i) & \text{if } y_i \geq 0, z_i \leq 0, \\ \nabla_i f(z)(y_i - z_i) + \tau_i(-y_i - z_i) & \text{if } y_i \leq 0, z_i \geq 0, \\ (\nabla_i f(z) - \tau_i)(y_i - z_i) & \text{if } y_i \leq 0, z_i \leq 0. \end{cases} \end{aligned} \tag{4.3}$$

Lemma 4.1. *There exists a positive constant M such that $|x_{i(r)}^{r+1} - \tilde{x}_{i(r)}^{r+1}| \leq \frac{2M}{\|A_{i(r)}\|}$ for all r.*

Proof. By lemma 2.2, we have that the set $\Omega(F(x^0))$ is compact. Since $\{Ax^{r+1}\}$, $\{A\tilde{x}^{r+1}\} \subseteq \Omega(F(x^0))$ holds, we further obtain that $\{Ax^{r+1}\}$ and $\{A\tilde{x}^{r+1}\}$ are bounded , that is to say, there exists a constant $M > 0$ such that $\|Ax^{r+1}\|$, $\|Ax^r\| \leq M$ for all r. Then we deduce

$$\|A_{i(r)}\||x_{i(r)}^{r+1} - \tilde{x}_{i(r)}^{r+1}| = \|Ax^{r+1} - A\tilde{x}^{r+1}\| \leq \|Ax^{r+1}\| + \|A\tilde{x}^{r+1}\| \leq 2M,$$

which implies the conclusion since A_i is nonzero for all i. $\square$

Lemma 4.2. $\lim\limits_{r\to\infty}\{F(\tilde{x}^r) - F(x^r)\} = 0.$

Proof. Since $\tilde{x}_{i(r)}^{r+1}$ is the exact solution of the subproblem (3.1), the inequality

$$F(\tilde{x}^{r+1}) - F(x^{r+1}) \leq 0 \tag{4.4}$$

always holds. On the other hand, by the convexity of f, we have

$$\begin{aligned} F(\tilde{x}^{r+1}) - F(x^{r+1}) \geq& \nabla_{i(r)} f(x^{r+1})(\tilde{x}_{i(r)}^{r+1} - x_{i(r)}^{r+1}) + \tau_{i(r)}(|\tilde{x}_{i(r)}^{r+1}| - |x_{i(r)}^{r+1}|) \\ =& h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}). \end{aligned} \tag{4.5}$$

Let indexes sets Z^A and Z^B be defined by

$$Z^A := \{r| \, |\tilde{x}^r_{i(r)} - x^r_{i(r)}| \le \varepsilon^r\}, \; Z^B := \{r| \, |\tilde{x}^r_{i(r)} - x^r_{i(r)}| > \varepsilon^r\},$$

respectively. First we consider the subsequence $\{x^{r+1}\}_{Z^A}$ of $\{x^r\}$. Since $\{Ax^r\}$ is bounded, $\{\nabla f(x^r)\}$ is also bounded from the continuity of ∇g. It then follows from (4.4), (4.5) and $\varepsilon^{r+1} \to 0$ that $\lim\limits_{r\to\infty,\, r\in Z^A} \{F(\tilde{x}^{r+1}) - F(x^{r+1})\} = 0$.

Next consider the subsequence $\{x^{r+1}\}_{Z^B}$. We will show the following inequality holds.

$$h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) \ge -P\varepsilon^{r+1}, \forall \; r+1 \in Z^B, \tag{4.6}$$

where $P = \frac{2M}{\|A_{i(r)}\|} + 2\tau_{i(r)} + 2\varepsilon^{r+1}$. Then it is easy to show $\lim\limits_{r\to\infty,\, r\in Z^B} \{F(\tilde{x}^{r+1}) - F(x^{r+1})\} = 0$ from (4.4), (4.5), (4.6) and $\varepsilon^r \to 0$.

Recall that $x^{r+1}_{i(r)}$ is an ε^{r+1}-approximate solution of the subproblem (3.1), i.e., there exists a $j \in \{1, 2, \cdots, 5\}$ such that $i(r) \in J_j(x^{r+1}, \varepsilon^{r+1})$. Suppose $r+1 \in Z^B$. In the rest part, we show that (4.6) holds for $i(r) \in J_j(x^{r+1}, \varepsilon^{r+1})$, $j \in \{1, 2, \cdots, 5\}$. For simplicity, we only show the cases $i(r) \in J_j(x^{r+1}, \varepsilon^{r+1})$, $j \in \{1, 2, \cdots, 3\}$. The cases $j \in \{4, 5\}$ can be deduced in a similar way.

Case 1: $i(r) \in J_1(x^{r+1}, \varepsilon^{r+1})$. We have $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \ge -\varepsilon^{r+1}$ and $|x^{r+1}_{i(r)} - l_{i(r)}| \le \varepsilon^{r+1}$. Since $\varepsilon^{r+1} \le \dfrac{1}{2} \min\{-l_{i(r)}, u_{i(r)}\}$, the inequality $x^{r+1}_{i(r)} < 0$ holds.

(a) If $\tilde{x}^{r+1}_{i(r)} \ge 0$, then it follows from (4.3), Lemma 4.1 and $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \ge -\varepsilon^{r+1}$ that

$$\begin{aligned}
h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) =& (\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)})\tilde{x}^{r+1}_{i(r)} - (\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)})x^{r+1}_{i(r)} \\
\ge& (2\tau_{i(r)} - \varepsilon^{r+1})\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)}(-\varepsilon^{r+1}) \\
\ge& -\varepsilon^{r+1}(\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)}) \\
\ge& -\varepsilon^{r+1}\frac{2M}{\|A_{i(r)}\|}.
\end{aligned}$$

(b) If $\tilde{x}^{r+1}_{i(r)} < 0$, then $\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)} > 0$ holds by $|x^{r+1}_{i(r)} - l_{i(r)}| \le \varepsilon^{r+1}$ and $r+1 \in Z^B$. We further have $h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) = (\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)})(\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)}) \ge -\varepsilon^{r+1}\frac{2M}{\|A_{i(r)}\|}$ from (4.3), $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \ge -\varepsilon^{r+1}$ and Lemma 4.1. Therefore, the inequality (4.6) holds when $i(r) \in J_1(x^{r+1}, \varepsilon^{r+1})$.

Case 2: $i(r) \in J_2(x^{r+1}, \varepsilon^{r+1})$. We have $|\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}| \le \varepsilon^{r+1}$ and $l_{i(r)} - \varepsilon^{r+1} \le x^{r+1}_{i(r)} \le \varepsilon^{r+1}$. Now,

$$h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) = (\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)})(\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)}) + T(x^{r+1}_{i(r)}, \tilde{x}^{r+1}_{i(r)}, \tau_{i(r)}), \tag{4.7}$$

where

$$
\begin{aligned}
T(x_{i(r)}^{r+1}, \tilde{x}_{i(r)}^{r+1}, \tau_{i(r)}) :=& \tau_{i(r)} \left(\tilde{x}_{i(r)}^{r+1} + |\tilde{x}_{i(r)}^{r+1}| - x_{i(r)}^{r+1} - |x_{i(r)}^{r+1}| \right) \\
=& \begin{cases} 0 & \text{if } \tilde{x}_{i(r)}^{r+1} \le 0,\ x_{i(r)}^{r+1} \le 0, \\ 2\tau_{i(r)} \tilde{x}_{i(r)}^{r+1} & \text{if } 0 < \tilde{x}_{i(r)}^{r+1},\ x_{i(r)}^{r+1} \le 0, \\ -2\tau_{i(r)} x_{i(r)}^{r+1} & \text{if } \tilde{x}_{i(r)}^{r+1} \le 0,\ 0 < x_{i(r)}^{r+1}, \\ 2\tau_{i(r)} \left(\tilde{x}_{i(r)}^{r+1} - x_{i(r)}^{r+1} \right) & \text{if } 0 < \tilde{x}_{i(r)}^{r+1},\ 0 < x_{i(r)}^{r+1}. \end{cases}
\end{aligned} \tag{4.8}
$$

Suppose first that one of $\tilde{x}_{i(r)}^{r+1}$ and $x_{i(r)}^{r+1}$ is nonpositive. It is easy to see that $T(x_{i(r)}^{r+1}, \tilde{x}_{i(r)}^{r+1}, \tau_{i(r)})$ is no less than $-2\tau_{i(r)}\varepsilon^{r+1}$. It then follows from $|\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}| \le \varepsilon^{r+1}$, Lemma 4.1 and (4.7) that

$$
h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) \ge -\varepsilon^{r+1}(\frac{2M}{\|A_{i(r)}\|} + 2\tau_{i(r)}).
$$

Next suppose that both $\tilde{x}_{i(r)}^{r+1}$ and $x_{i(r)}^{r+1}$ are positive. Then

$$
\begin{aligned}
h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) =& (\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)}) \tilde{x}_{i(r)}^{r+1} - x_{i(r)}^{r+1} (\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)}) \\
\ge& (2\tau_i - \varepsilon^{r+1}) \tilde{x}_{i(r)}^{r+1} - x_{i(r)}^{r+1} (2\tau_{i(r)} + \varepsilon^{r+1}) \\
\ge& -\varepsilon^{r+1} (\frac{2M}{\|A_{i(r)}\|} + x_{i(r)}^{r+1}) - x_{i(r)}^{r+1} (2\tau_{i(r)} + \varepsilon^{r+1}) \\
\ge& -(\frac{2M}{\|A_{i(r)}\|} + 2\tau_{i(r)} + 2\varepsilon^{r+1}) \varepsilon^{r+1},
\end{aligned}
$$

where the first inequality follows from $|\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}| \le \varepsilon^{r+1}$, $\tilde{x}_{i(r)}^{r+1} > 0$ and $x_{i(r)}^{r+1} > 0$, the second inequality follows from $\tilde{x}_{i(r)}^{r+1} > 0$ and Lemma 4.1, and the last inequality holds from $0 \le x_{i(r)}^{r+1} \le \varepsilon^{r+1}$. Thus, the inequality (4.6) is also confirmed.

Case 3: $i(r) \in J_3(x^{r+1}, \varepsilon^{r+1})$. We have $|\nabla_{i(r)} f(x^{r+1})| \le \tau_{i(r)} + \varepsilon^{r+1}$ and $|x_{i(r)}^{r+1}| \le \varepsilon^{r+1}$. Moreover, we deduce $\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)} \in [-\varepsilon^{r+1}, 2\tau_i + \varepsilon^{r+1}]$ from the first inequality. Next we only show that the inequality (4.6) holds when $0 \le x_{i(r)}^{r+1} \le \varepsilon^{r+1}$. A symmetric argument can prove the case where $-\varepsilon^{r+1} \le x_{i(r)}^{r+1} \le 0$.

(a) Suppose that $\tilde{x}_{i(r)}^{r+1} \ge 0$. If $\nabla_i f(x^{r+1}) + \tau_{i(r)} \in [-\varepsilon^{r+1}, 0)$, then we have from Lemma 4.1 that

$$
\begin{aligned}
h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) =& (\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)}) (\tilde{x}_{i(r)}^{r+1} - x_{i(r)}^{r+1}) \\
\ge& -|\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)}| |\tilde{x}_{i(r)}^{r+1} - x_{i(r)}^{r+1}| \\
\ge& -\varepsilon^{r+1} \frac{2M}{\|A_{i(r)}\|}.
\end{aligned}
$$

If $\nabla_i f(x^{r+1}) + \tau_{i(r)} \in [0, 2\tau_{i(r)} + \varepsilon^{r+1}]$, then $\tilde{x}^{r+1}_{i(r)}(\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)}) \geq 0$. Since $0 \leq x^{r+1}_{i(r)} \leq \varepsilon^{r+1}$, we have

$$\begin{aligned} h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) =& \tilde{x}^{r+1}_{i(r)}(\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)}) - x^{r+1}_{i(r)}(\nabla_{i(r)} f(x^{r+1}) + \tau_{i(r)}) \\ \geq & - \varepsilon^{r+1}(\varepsilon^{r+1} + 2\tau_{i(r)}). \end{aligned}$$

(b) Suppose that $\tilde{x}^{r+1}_{i(r)} < 0$. Then it follows from $|\nabla_{i(r)} f(x^{r+1})| \leq \tau_{i(r)} + \varepsilon^{r+1}$, $0 \leq x^{r+1}_{i(r)} \leq \varepsilon^{r+1}$ and Lemma 4.1 that

$$\begin{aligned} h_{i(r)}(\tilde{x}^{r+1}, x^{r+1}) =& (\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)})\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)}(\nabla_i f(x^{r+1}) + \tau_{i(r)}) \\ \geq & \varepsilon^{r+1}\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)}(2\tau_{i(r)} + \varepsilon^{r+1}) \\ =& \varepsilon^{r+1}(\tilde{x}^{r+1}_{i(r)} - x^{r+1}_{i(r)}) - 2\tau_{i(r)} x^{r+1}_{i(r)} \\ \geq & - \varepsilon^{r+1}\left(\frac{2M}{\|A_{i(r)}\|} + 2\tau_{i(r)}\right). \end{aligned}$$

It is clear that $h_{i(r)}(\tilde{x}^{r+1}, x^{r+1})$ in both cases (a) and (b) satisfies (4.6).

□

Using the above lemmas, we can show that $\{x^{r+1} - x^r\}$ converges to 0.

Lemma 4.3. *For the sequence $\{x^r\}$ generated by the ICD method, we have $\lim_{r\to\infty}\{x^{r+1} - x^r\} = 0$.*

Proof. We argue it by contradiction. Suppose $x^{r+1} - x^r \nrightarrow 0$. Then there exist at least one coordinate $i \in \{1, 2, \cdots, n\}$, a scalar $\gamma > 0$ and an infinite subset $\tilde{Z}$ of nonnegative integers such that $|x^{r+1}_i - x^r_i| \geq \gamma$ for all $r \in \tilde{Z}$. Since $\gamma > 0$, the index i is the index $i(r)$ chosen in Step 2 of the ICD method at the rth step. Therefore, for any $j \neq i(r)$, we have $x^{r+1}_j = x^r_j$, which together with the assumption $|x^{r+1}_{i(r)} - x^r_{i(r)}| \geq \gamma$ implies that

$$\|A(x^{r+1} - x^r)\| = \|A_{i(r)}\||x^{r+1}_{i(r)} - x^r_{i(r)}| \geq \|A_{i(r)}\|\gamma, \ \ \forall r \in \tilde{Z}. \tag{4.9}$$

Since $\{Ax^r\}$ is bounded, there exist $t^{1,\infty},\ t^{2,\infty} \in \mathcal{R}^n$ and an infinite set $\mathcal{H} \subseteq \tilde{Z}$ such that

$$\lim_{r\to\infty,\, r\in\mathcal{H}} Ax^r = t^{1,\infty}, \lim_{r\to\infty,\, r\in\mathcal{H}} Ax^{r+1} = t^{2,\infty}. \tag{4.10}$$

Note that $t^{1,\infty} \neq t^{2,\infty}$ due to (4.9). It then follows from the continuity of g on $\Omega(F(x^0))$ and (4.10) that

$$\lim_{r\to\infty,\, r\in\mathcal{H}} g(Ax^r) = g(t^{1,\infty}), \lim_{r\to\infty,\, r\in\mathcal{H}} g(Ax^{r+1}) = g(t^{2,\infty}). \tag{4.11}$$

Since $F(x^r)$ is monotonically decreasing from Assumption 3.1(i) and $F(x^r) \geq F(x^*)$ for any optimal solution x^*, the sequence $\{F(x^r)\}$ is convergent. Let F^∞ be its limit. Then we have

$$\lim_{r\to\infty,\, r\in\mathcal{H}} F(x^r) = F^\infty, \lim_{r\to\infty,\, r\in\mathcal{H}} F(x^{r+1}) = F^\infty. \tag{4.12}$$

Moreover, by Lemma 4.2 and (4.12), we obtain

$$\lim_{r\to\infty,\, r\in\mathcal{H}} F(\tilde{x}^{r+1}) = \lim_{r\to\infty,\, r\in\mathcal{H}} F(x^{r+1}) - \lim_{r\to\infty,\, r\in\mathcal{H}} (F(x^{r+1}) - F(\tilde{x}^{r+1})) = F^\infty, \tag{4.13}$$

where $\tilde{x}^{r+1}$ is defined in (4.1). Since F is convex and $F(\tilde{x}^{r+1}) \le F(x^{r+1}) \le F(x^r)$, we have

$$F(\tilde{x}^{r+1}) \le F\left(\frac{x^r + x^{r+1}}{2}\right) \le \frac{1}{2}F(x^r) + \frac{1}{2}F(x^{r+1}) \le F(x^r).$$

Taking a limit on these inequalities, we obtain

$$\lim_{r\to\infty,\, r\in\mathcal{H}} F\left(\frac{x^{r+1} + x^r}{2}\right) = F^\infty. \tag{4.14}$$

On the other hand,

$$\begin{aligned}
&\lim_{r\to\infty,\, r\in\mathcal{H}} F\left(\frac{x^{r+1}+x^r}{2}\right)\\
\le&\lim_{r\to\infty,\, r\in\mathcal{H}} g(\frac{Ax^{r+1}+Ax^r}{2}) + \limsup_{r\to\infty,\, r\in\mathcal{H}} \left\{\langle b, \frac{x^{r+1}+x^r}{2}\rangle + \sum_{i=1}^n \tau_{i(r)} \left|\frac{x^{r+1}_{i(r)} + x^r_{i(r)}}{2}\right|\right\}\\
\le& g(\frac{t^{1,\infty}+t^{2,\infty}}{2}) + \frac{1}{2}\limsup_{r\to\infty,\, r\in\mathcal{H}} \{\langle b, x^r\rangle + \sum_{i=1}^n \tau_{i(r)}|x^r_{i(r)}|\} + \frac{1}{2}\limsup_{r\to\infty,\, r\in\mathcal{H}} \{\langle b, x^{r+1}\rangle + \sum_{i=1}^n \tau_{i(r)}|x^{r+1}_{i(r)}|\}\\
=& g(\frac{t^{1,\infty}+t^{2,\infty}}{2}) + \frac{1}{2}\limsup_{r\to\infty,\, r\in\mathcal{H}} \{F(x^r) - g(Ax^r)\} + \frac{1}{2}\limsup_{r\to\infty,\, r\in\mathcal{H}} \{F(x^{r+1}) - g(Ax^{r+1})\}\\
=& g(\frac{t^{1,\infty}+t^{2,\infty}}{2}) + \frac{1}{2}(F^\infty - g(t^{1,\infty})) + \frac{1}{2}(F^\infty - g(t^{2,\infty}))\\
<&\frac{1}{2}(g(t^{1,\infty}) + g(t^{2,\infty})) + \frac{1}{2}(F^\infty - g(t^{1,\infty})) + \frac{1}{2}(F^\infty - g(t^{2,\infty}))\\
=& F^\infty,
\end{aligned}$$

where the second inequality follows from the continuity of g and (4.10), the first equality follows from the definition of F, the second equality follows from (4.11) and (4.12), and the third inequality follows from the strict convexity of g and $t^{1,\infty} \ne t^{2,\infty}$. But this inequality contradicts (4.14). Thus $\lim\limits_{r\to\infty}\{x^{r+1} - x^r\} = 0$. □

In the second part of this section, we will show the convergence of $\{Ax^r\}$. Since $\{Ax^r\}$ is bounded, there exist $t^\infty \in \mathcal{R}^n$ and an infinite set $\mathcal{X}$ such that

$$\lim_{r\to\infty,\, r\in\mathcal{X}} Ax^r = t^\infty. \tag{4.15}$$

Then with the continuity of ∇g, we have

$$\lim_{r\to\infty,\, r\in\mathcal{X}} \nabla f(x^r) = d^\infty, \tag{4.16}$$

where

$$d^\infty := A^T \nabla g(t^\infty) + b. \tag{4.17}$$

For the set $\mathcal{X}$, we have the following result with Lemma 4.3, which provides an interesting property associated with $\{\nabla f(x^r)\}$.

Lemma 4.4. *For any $s \in \{0, 1, \cdots, B-1\}$, where B is the integer defined in the almost cycle rule, we have $\lim\limits_{r\to\infty,\, r\in\mathcal{X}} \nabla f(x^{r-s}) = d^\infty$.*

Proof. For any $s \in \{0, 1, \cdots, B-1\}$, we have $Ax^{r-s} = \sum\limits_{k=0}^{s-1} A(x^{r-s+k} - x^{r-s+k+1}) + Ax^r$. It then follows from Lemma 4.3 and (4.15) that

$$\lim_{r\to\infty,\, r\in\mathcal{X}} Ax^{r-s} = \lim_{r\to\infty,\, r\in\mathcal{X}} \sum_{k=0}^{s-1} A(x^{r-s+k} - x^{r-s+k+1}) + \lim_{r\to\infty, r\in\mathcal{X}} Ax^r = t^\infty.$$

From the continuity of ∇g, we have $\lim\limits_{r\to\infty,\, r\in\mathcal{X}} \nabla f(x^{r-s}) = \lim\limits_{r\to\infty,\, r\in\mathcal{X}} A^T \nabla g(Ax^{r-s}) + b = A^T \nabla g(t^\infty) + b$, which together with (4.17) shows this lemma. $\square$

Lemma 4.4 implies that for each $i \in \{1, 2, \cdots, n\}$, for any $s \in \{0, 1, \cdots, B-1\}$, we have

$$\lim_{r\to\infty,\, r\in\mathcal{X}} \nabla_i f(x^{r-s}) = d_i^\infty. \tag{4.18}$$

Fixed coordinate i, let $\varphi(r, i)$ denote the largest integer $\bar{r}$, which does not exceed r, such that the ith coordinate of x is iterated upon at the $\bar{r}$th iteration, that is to say, for all $r \in \mathcal{X}$, we have

$$x_i^r = x_i^{\varphi(r,i)}. \tag{4.19}$$

Since the coordinate is chosen by the almost cycle rule, the relation $r - B + 1 \leq \varphi(r, i) \leq r$ holds for all $r \in \mathcal{X}$. From (4.18), we further obtain

$$\lim_{r\to\infty,\, r\in\mathcal{X}} \nabla_i f(x^{\varphi(r,i)}) = d_i^\infty. \tag{4.20}$$

Now we define the following six index sets associated with d_i^∞ as

$$\begin{aligned}
J_1^\infty &:= \{i|\ d_i^\infty > \tau_i\};\\
J_2^\infty &:= \{i|\ d_i^\infty < -\tau_i\};\\
J_3^\infty &:= \{i|\ |d_i^\infty| < \tau_i\};\\
J_4^\infty &:= \{i|\ d_i^\infty = \tau_i,\ \tau_i > 0\};\\
J_5^\infty &:= \{i|\ d_i^\infty = -\tau_i,\ \tau_i > 0\};
\end{aligned}$$

$$J_6^\infty := \{i |\ d_i^\infty = 0,\ \tau_i = 0\}.$$

Note that $\bigcup_{i=1}^{6} J_i^\infty = \{1, 2, \cdots, n\}$. Next two lemmas give sufficient conditions under which $\{x_i^r\}_\mathcal{X}$ is fixed or lies in some interval.

Lemma 4.5. *Suppose that Assumption 3.1(i) and (iii) hold, and that $x^{\varphi(r,i)}$ is a vector, where the i-th coordinate is chosen on the $\varphi(r,i)$-th iteration. Let L and ε_0 be the constants given in Lemma 2.3. If $\varepsilon^{\varphi(r,i)} < \varepsilon_0$, then the following statements hold for any fixed i:*

(i) *If $\nabla_i f(x^{\varphi(r,i)}) - \tau_i > L\|A_i\|^2\varepsilon^{\varphi(r,i)}$ and $x_i^{\varphi(r,i)} \le \varepsilon^{\varphi(r,i)} + l_i$, then $x_i^{\varphi(r,i)} = l_i$.*

(ii) *If $\nabla_i f(x^{\varphi(r,i)}) + \tau_i < -L\|A_i\|^2\varepsilon^{\varphi(r,i)}$ and $u_i - \varepsilon^{\varphi(r,i)} \le x_i^{\varphi(r,i)}$, then $x_i^{\varphi(r,i)} = u_i$.*

(iii) *If $\nabla_i f(x^{\varphi(r,i)}) + \tau_i > L\|A_i\|^2\varepsilon^{\varphi(r,i)}$ and $|x_i^{\varphi(r,i)}| \le \varepsilon^{\varphi(r,i)}$, then $x_i^{\varphi(r,i)} \le 0$.*

(iv) *If $\nabla_i f(x^{\varphi(r,i)}) - \tau_i < -L\|A_i\|^2\varepsilon^{\varphi(r,i)}$ and $|x_i^{\varphi(r,i)}| \le \varepsilon^{\varphi(r,i)}$, then $x_i^{\varphi(r,i)} \ge 0$.*

Proof. Here, we only show (i) and (iii). The rest can be obtained similarly.

To show (i), we argue by contradiction. If it is not true, then we have $l_i < x_i^{\varphi(r,i)} \le \varepsilon^{\varphi(r,i)} + l_i$ by Assumption 3.1(ii). From the Lipschitz continuity of ∇g in Lemma 2.3, we obtain that $|\nabla_i f(\hat{x}^{\varphi(r,i)}) - \nabla_i f(x^{\varphi(r,i)})| \le L\|A_i\|^2|l_i - x_i^{\varphi(r,i)}|$, where $\hat{x}^{\varphi(r,i)} := (x_1^r, \cdots, x_{i-1}^r, l_i, x_{i+1}^r, \cdots, x_n^r)$. We further can ensure $\nabla_i f(\hat{x}^{\varphi(r,i)}) - \tau_i \ge -L\|A_i\|^2\varepsilon^{\varphi(r,i)} + \nabla_i f(x^{\varphi(r,i)}) - \tau_i > 0$ with the assumptions $l_i < x_i^{\varphi(r,i)} \le \varepsilon^{\varphi(r,i)} + l_i$ and $\nabla_i f(x^{\varphi(r,i)}) - \tau_i > L\|A_i\|^2\varepsilon^{\varphi(r,i)}$. It then follows from the KKT conditions in Lemma 2.1 that l_i is the exact solution of the subproblem (3.1). Since the solution of the subproblem (3.1) is unique, we have $F(x^{\varphi(r,i)}) - F(\hat{x}^{\varphi(r,i)}) > 0$, which contradicts Assumption 3.1(i). Therefore, we have $x_i^{\varphi(r,i)} = l_i$.

For (iii), we also prove by contradiction. Suppose that the contray holds, i.e., $x_i^{\varphi(r,i)} \in (0, \varepsilon^{\varphi(r,i)}]$. Let $\tilde{x}^{\varphi(r,i)} := (x_1^r, \cdots, x_{i-1}^r, 0, x_{i+1}^r, \cdots, x_n^r)$. Then, by Lemma 2.3 and the assumption $x_i^{\varphi(r,i)} \in (0, \varepsilon^{\varphi(r,i)}]$, we have

$$|\nabla_i f(\tilde{x}^{\varphi(r,i)}) - \nabla_i f(x^{\varphi(r,i)})| \le L\|A_i\|^2|0 - x_i^{\varphi(r,i)}| \le L\|A_i\|^2\varepsilon^{\varphi(r,i)},$$

which implies,

$$-L\|A_i\|^2\varepsilon^{\varphi(r,i)} + \nabla_i f(x^{\varphi(r,i)}) \le \nabla_i f(\tilde{x}^{\varphi(r,i)}).$$

By the convexity of f, $0 < x_i^{\varphi(r,i)} \le \varepsilon^{\varphi(r,i)}$ and $\nabla_i f(x^{\varphi(r,i)}) + \tau_i > L\|A_i\|^2\varepsilon^{\varphi(r,i)}$, we further have that

$$F(x^{\varphi(r,i)}) - F(\tilde{x}^{\varphi(r,i)}) \ge \nabla_i f(\tilde{x}^{\varphi(r,i)})(x_i^{\varphi(r,i)} - 0) + \tau_i x_i^{\varphi(r,i)} > 0, \tag{4.21}$$

which contradicts Assumption 3.1(i). $\square$

Lemma 4.6. *Suppose that Assumption 3.1 holds. Then, for sufficiently large r, we have*

$$\{x_i^r\}_{\mathcal{X}} = l_i, \forall i \in J_1^\infty; \tag{4.22}$$

$$\{x_i^r\}_{\mathcal{X}} = u_i, \forall i \in J_2^\infty; \tag{4.23}$$

$$\{x_i^r\}_{\mathcal{X}} = 0, \forall i \in J_3^\infty; \tag{4.24}$$

$$l_i \le \{x_i^r\}_{\mathcal{X}} \le 0, \forall i \in J_4^\infty; \tag{4.25}$$

$$0 \le \{x_i^r\}_{\mathcal{X}} \le u_i, \forall i \in J_5^\infty; \tag{4.26}$$

$$l_i \le \{x_i^r\}_{\mathcal{X}} \le u_i, \forall i \in J_6^\infty. \tag{4.27}$$

Proof. Here we only show (4.22) and (4.25). Since the rest part can be shown in a similar way, we omit their proofs.

Case 1: $i \in J_1^\infty$. To show (4.22), it is sufficient to show

$$\{x_i^{\varphi(r,i)}\}_{\mathcal{X}} = l_i, \tag{4.28}$$

since $x_i^r = x_i^{\varphi(r,i)}$ by (4.19). From (4.20), we have that for $\bar{\varepsilon} = \frac{d_i^\infty - \tau_i}{2} > 0$, $i \in J_1^\infty$, there exists a nonnegative integer $\bar{r}$ such that

$$d_i^\infty - \bar{\varepsilon} \le \nabla_i f(x^{\varphi(r,i)}) \le d_i^\infty + \bar{\varepsilon},\ \forall r \ge \bar{r}, r \in \mathcal{X}.$$

It is easy to see that $d_i^\infty - \tau_i - \bar{\varepsilon}$ is positive. Then we have

$$\nabla_i f(x^{\varphi(r,i)}) - \tau_i \ge d_i^\infty - \tau_i - \bar{\varepsilon} > \max\{1, L\|A_i\|^2\}\varepsilon^{\varphi(r,i)} \ge \varepsilon^{\varphi(r,i)} \tag{4.29}$$

for sufficiently large r since $\varepsilon^r \to 0$ and $\nabla_i f(x^{\varphi(r,i)}) \to d_i^\infty$. Furthermore, we ensure $i \in J_1(x^{\varphi(r,i)}, \varepsilon^{\varphi(r,i)})$, since $x_i^{\varphi(r,i)}$ is an $\varepsilon^{\varphi(r,i)}$-approximate solution of the subproblem (3.1). It implies that $|x_i^{\varphi(r,i)} - l_i| \le \varepsilon^{\varphi(r,i)}$. Then by the Assumption 3.1(ii) and (2.7), we have

$$l_i \le x_i^{\varphi(r,i)} \le \varepsilon^{\varphi(r,i)} + l_i < 0. \tag{4.30}$$

Thus, the equality (4.28) can be deduced by (4.29), (4.30) and Lemma 4.5(i), and (4.22) is confirmed.

Case 2: $i \in J_4$. In this case, we have $d_i^\infty = \tau_i$ and $\tau_i > 0$. Let $\tilde{\varepsilon} = \frac{\tau_i}{2}$. It then follows from (4.20) that there exists a $\tilde{r}$, such that $\frac{1}{2}\tau_i < \nabla_i f(x^{\varphi(r,i)}) < \frac{3}{2}\tau_i$ for all $r \in \mathcal{X}$ such that $r \ge \tilde{r}$. Then for sufficiently large r, the inequalities

$$\nabla_i f(x^{\varphi(r,i)}) + \tau_i > \frac{3}{2}\tau_i > \max\{1, L\|A_i\|^2\}\varepsilon^{\varphi(r,i)} \ge \varepsilon^{\varphi(r,i)} \tag{4.31}$$

hold since $\varepsilon^r \to 0$. We further can obtain $i \in \bigcup_{j=1}^{3} J_j(x^{\varphi(r,i)}, \varepsilon^{\varphi(r,i)})$ from Definition 2.1. Therefore, we have

$$x_i^{\varphi(r,i)} \in [l_i, \varepsilon^{\varphi(r,i)}]. \tag{4.32}$$

It finally follows from (4.31), (4.32) and Lemma 4.5(iii) that $x_i^{\varphi(r,i)} \in [l_i, 0]$. From (4.19), (4.25) is confirmed.

□

Next, we will show $Ax^r \to Ax^*$, where x^* is an arbitrary optimal solution of the problem (1.1). For this purpose, we recall Hoffman's error bound [2].

Lemma 4.7. *Let $B \in \mathcal{R}^{k\times n}$, $C \in \mathcal{R}^{k\times n}$ and $e \in \mathcal{R}^k$, $d \in \mathcal{R}^k$. Suppose that the linear system $By = e, Cy \leq d$ is consistent. There exists a scalar $\theta > 0$ depending only on B and C, and such that, for any $\bar{x} \in [l, u]$, $l,\ u \in \mathcal{R}^n$, there is a point $\bar{y} \in \mathcal{R}^n$ satisfying $B\bar{y} = e, C\bar{y} \leq d$ and $\|\bar{x} - \bar{y}\| \leq \theta(\|B\bar{x} - e\| + \|(C\bar{x} - d)_+\|$, where $(x_i)_+ := \max\{0, x_i\}$.*

Theorem 4.1. *Let x^* be an optimal solution of the problem (1.1). Then we have $\lim_{r\to\infty} Ax^r = Ax^*$.*

Proof. In the first step, we show that $Ax^r \to Ax^*$ holds for $r \in \mathcal{X}$, where $\mathcal{X}$ is an infinite set given in (4.15). To this end, we consider the following linear system of y:

$$Ay = Ax^r,\ \ y_i = x_i^r\ (i \in J_1^\infty \cup J_2^\infty \cup J_3)^\infty,\ y_i \leq 0\ (i \in J_4^\infty),\ \text{and}\ y_i \geq 0\ (i \in J_5^\infty),\ y \in [l, u].$$

For sufficiently large r, it follows from (4.22) - (4.27) that x^r is a solution of this system, that is to say, the system is consistent. For any fixed point $\bar{x}$ in $[l, u]$, by Lemma 4.7, there exists a solution $y^r \in [l, u]$ of the above system and a constant θ, which is independent of x^r, such that

$$\|y^r - \bar{x}\| \leq \theta \left(\|A\bar{x} - Ax^r\| + \sum_{i\in J_1\cup J_2\cup J_3} |\bar{x}_i - x_i^r| + \sum_{i\in J_4} \max\{0, \bar{x}_i\} + \sum_{i\in J_5} \max\{0, -\bar{x}_i\} \right).$$

From the boundness of $\{Ax^r\}$ and (4.22)-(4.24), we further have that the right-hand side of this inequality is bounded. It implies that $\{y^r\}_{\mathcal{X}}$ is also bounded, and hence it has at least one accumulation point. We denote it by y^∞. Furthermore, from (4.15) and Lemma 4.6, we have that y^∞ satisfies the following system:

$$Ay^\infty = t^\infty,\ \ y_i^\infty = l_i\ (i \in J_1),\ \ y_i^\infty = u_i\ (i \in J_2),\ \ y_i^\infty = 0\ (i \in J_3),$$

$$l_i \leq y_i^\infty \leq 0\ (i \in J_4),\ 0 \leq y_i^\infty \leq u_i\ (i \in J_5), l_i \leq y_i^\infty \leq u_i\ (i \in J_6).$$

It then follows from (4.17) that $\nabla f(y^\infty) = A^T \nabla g(Ay^\infty) + b = d^\infty$. Moreover, the relation $y^\infty = P_{\tau,l,u}(y^\infty)$ holds from the above system and Lemma 2.1. Thus, y^∞ is an optimal solution of the problem (1.1) by Lemma 2.1. From Lemma 2.4, we have $Ay^\infty = Ax^*$, i.e., $t^\infty = Ax^*$.

In the second step, we show $\lim\limits_{r\to\infty} Ax^r = Ax^*$. Since $\{Ax^r\}$ is bounded, it is sufficient to show that any accumulation point of $\{Ax^r\}$ is Ax^*. Let $\hat{\mathcal{X}}$ be any subset of nonnegative integers such that $\{Ax^r\}$ is convergent and let $\hat{t}^\infty$ be a limit of$\{Ax^r\}_{\hat{\mathcal{X}}}$. Then we can show $\hat{t}^\infty = Ax^*$ in the set $\hat{\mathcal{X}}$ as Lemma 4.4-4.6. Moreover, the first step of the current proof, i.e., $\{Ax^r\}_{\hat{\mathcal{X}}} \to Ax^*$ holds. Thus, $\{Ax^r\} \to Ax^*$ holds for $r \to \infty$. $\square$

Theorem 4.1 implies that there exists a scalar $\bar{r} > 0$, for any $r \geq \bar{r}$, such that $Ax^r \in B(Ax^*)$, where $B(Ax^*)$ is the closed ball defined just before (2.1). Note that g is strongly convex.

for any small closed ball $B(Ax^*)$, mentioned in the remarks of Assumption 2.1about the positive definite of $\nabla^2 g$, there exists a scalar $\bar{r} > 0$, for any $r \geq \bar{r}$, such that $Ax^r \in B(Ax^*)$.

In the third part of this section, we show the sufficient decreasing of $\{F(x^r)\}$ for sufficiently large r.

Lemma 4.8. *Under Assumption 3.1, there exists a scalar $\eta > 0$ such that $F(x^r) - F(x^{r+1}) \geq \eta\|x^r - x^{r+1}\|^2$ for sufficiently large r.*

Proof. Note that $Ax^r, Ax^{r+1} \in B(Ax^*)$ for sufficiently large r. It then follows from Assumption 2.1 that g is strongly convex in $B(Ax^*)$. Furthermore, we have

$$\begin{aligned} F(x^r) - F(x^{r+1}) =& g(Ax^r) - g(Ax^{r+1}) - \langle A^T \nabla g(Ax^{r+1}), x^r - x^{r+1}\rangle \\ &+ \langle \nabla f(x^{r+1}), x^r - x^{r+1}\rangle + \tau_{i(r)}|x^r_{i(r)}| - \tau_{i(r)}|x^{r+1}_{i(r)}| \\ \geq& \sigma\|A(x^r - x^{r+1})\|^2 + \langle \nabla_{i(r)} f(x^{r+1}), x^r_{i(r)} - x^{r+1}_{i(r)}\rangle + \tau_{i(r)}\left(|x^r_{i(r)}| - |x^{r+1}_{i(r)}|\right) \\ =& \sigma\|A_{i(r)}\|^2 |x^r_{i(r)} - x^{r+1}_{i(r)}|^2 + h_{i(r)}(x^r, x^{r+1}) \\ \geq& \sigma \min_j \|A_j\|^2 \|x^r - x^{r+1}\|^2 + h_{i(r)}(x^r, x^{r+1}), \end{aligned}$$

where $h_{i(r)}$ is defined in (4.3), and $i(r)$ denotes the index chosen on rth step.

Next, we show the inequality

$$h_{i(r)}(x^r, x^{r+1}) \geq -\alpha_r \tilde{L}(x^r_{i(r)} - x^{r+1}_{i(r)})^2, \tag{4.33}$$

where $\tilde{L} := \max\limits_j\{1, L\|A_j\|^2\}$, and α_r is given in Assumption 3.1(v). Note that $\tilde{L} \geq 1$.

We show it by considering 6 cases: $i(r) \in J^\infty_j$, $j = 1, 2, \cdots, 6$. First, we have from Lemma 4.6

$$h_{i(r)}(x^r, x^{r+1}) = 0, \ \forall\, i(r) \in \bigcup_{j=1}^{3} J^\infty_j,$$

and hence (4.33) holds for $i(r) \in J_j^{\infty},\ j = 1, 2, 3$. Then, we only need to consider the other three cases $i \in J_4^{\infty}$, $i \in J_5^{\infty}$ and $i \in J_6^{\infty}$. Here, for simplicity, we only show the case $i \in J_4^{\infty}$. The rest two cases can be obtained in a similar way.

If $i(r) \in J_4^{\infty}$, then it follows from Lemma 4.6 that for the sufficiently large r, $x_{i(r)}^r, x_{i(r)}^{r+1} \in [l_{i(r)}, 0]$. Then we have

$$\begin{aligned} h_{i(r)}(x^r, x^{r+1}) =& \langle \nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}, x_{i(r)}^r - x_{i(r)}^{r+1} \rangle \\ \geq& -|\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}||x_{i(r)}^r - x_{i(r)}^{r+1}|. \end{aligned} \tag{4.34}$$

From the proof of (4.25) in Lemma 4.6, we have $i(r) \in \bigcup_{j=1}^{3} J_j(x^{r+1}, \varepsilon^{r+1})$. Thus we show (4.33) by considering the following three distinct cases.

Case 1: $i(r) \in J_1(x^{r+1}, \varepsilon^{r+1})$. We have by Assumption 3.1(ii)

$$\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \geq -\varepsilon^{r+1} \text{ and } l_{i(r)} \leq x_{i(r)}^{r+1} \leq l_{i(r)} + \varepsilon^{r+1}. \tag{4.35}$$

The first inequality means that $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \in [-\varepsilon^{r+1}, \infty) = [-\varepsilon^{r+1}, \tilde{L}\varepsilon^{r+1}] \cup (\tilde{L}\varepsilon^{r+1}, \infty)$. First suppose that $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \in [-\varepsilon^{r+1}, \tilde{L}\varepsilon^{r+1}]$. It then follows from (4.34) and Assumption 3.1(iv) that $h_{i(r)}(x^r, x^{r+1}) \geq -\tilde{L}\varepsilon^{r+1}|x_{i(r)}^r - x_{i(r)}^{r+1}| \geq -\alpha_r \tilde{L}|x_{i(r)}^r - x_{i(r)}^{r+1}|^2$, which satisfies (4.33).

Next suppose that $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \in (\tilde{L}\varepsilon^{r+1}, \infty)$. Then $x_{i(r)}^{r+1} = l_{i(r)}$ holds from $l_{i(r)} \leq x_{i(r)}^{r+1} \leq l_{i(r)} + \varepsilon^{r+1}$ and Lemma 4.5(i). Therefore, $h_{i(r)}(x^r, x^{r+1}) = \langle \nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}, x_{i(r)}^r - l_{i(r)} \rangle \geq 0$, which implies (4.33) obviously.

Case 2: $i(r) \in J_2(x^{r+1}, \varepsilon^{r+1})$. In this case, we have $|\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}| \leq \varepsilon^{r+1}$ and $l_{i(r)} \leq x_{i(r)}^{r+1} \leq 0$. From Assumption 3.1(iv) and (4.34), we can ensure $h_{i(r)}(x^r, x^{r+1}) \geq -\varepsilon^{r+1}|x_{i(r)}^r - x_{i(r)}^{r+1}| \geq -\alpha_r |x_{i(r)}^r - x_{i(r)}^{r+1}|^2$, which also implies (4.33).

Case 3: $i(r) \in J_3(x^{r+1}, \varepsilon^{r+1})$. We have $|\nabla_{i(r)} f(x^{r+1})| \leq \tau_{i(r)} + \varepsilon^{r+1}$ and $-\varepsilon^{r+1} \leq x_{i(r)}^{r+1} \leq 0$, and hence $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \in [-2\tau_{i(r)} - \varepsilon^{r+1}, \varepsilon^{r+1}]$. If $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \in [-\tilde{L}\varepsilon^{r+1}, \varepsilon^{r+1}]$, then (4.33) holds from Assumption 3.1(iv). If $\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)} \in [-2\tau_{i(r)} - \varepsilon^{r+1}, -\tilde{L}\varepsilon^{r+1})$, then we have $x_{i(r)}^{r+1} = 0$ from Lemma 4.5 and $x_{i(r)}^{r+1} \in [-\varepsilon^{r+1}, 0]$. Hence, we can ensure $h_{i(r)}(x^r, x^{r+1}) = (\nabla_{i(r)} f(x^{r+1}) - \tau_{i(r)}) x_{i(r)}^r \geq 0 \geq -\alpha_r \tilde{L}(x_{i(r)}^r - x_{i(r)}^{r+1})^2$.

Consequently, the inequality (4.33) holds.

The sequence $\{\alpha_r\}$ satisfies $\alpha_r < \dfrac{\sigma \min_j \|A_j\|^2}{\max_j \{1, L\|A_j\|^2\}}$ for sufficiently large r from the Assumption 3.1(v). Then the inequality of this theorem holds with $\eta := \sigma \min_j \|A_j\|^2 - \alpha_r \max_j \{1, L\|A_j\|^2\} >$

0 for sufficiently large r. $\square$

In the last part of this section, before showing the global and linear convergence of $\{x^r\}$, we first recall a kind of the Lipschitz error bound in [10, 11, 17].

Lemma 4.9. *There exists a scalar constant $\kappa > 0$ such that for any $Ax^r \in B(Ax^*)$,*

$$\|Ax^r - Ax^*\| \leq \kappa \|x^r - P_{\tau,l,u}(x^r)\|. \tag{4.36}$$

Proof. Since g is strongly convex on $B(Ax^*)$ and ∇g is Lipschitz continuous, there exists a constant $\hat{\kappa} > 0$ such that $\|x^r - x^*(r)\| \leq \hat{\kappa}\|x^r - P_{\tau,l,u}(x^r)\|$, where $x^*(r)$ is a nearest solution from x^r [17, Lemma 4.4]. It then follows from Lemma 2.4 and $\|Ax^r - Ax^*\| \leq \|A\|\|x^r - x^*\|$ that (4.36) holds with $\kappa := \|A\|\hat{\kappa}$. $\square$

The following result is a direct extension of [16, Lemma 4.5(a)] to the problem (1.1).

Lemma 4.10. *Under Assumption 3.1, there exists a constant $\omega > 0$ such that the inequality $\|Ax^r - Ax^*\| \leq \omega \sum_{h=r}^{r+B-1} \|x^h - x^{h+1}\|$ holds for sufficiently large r.*

Proof. To show this lemma, by Lemmas 4.9, it is sufficient to show that there exists a constant $\hat{\omega} > 0$ such that $\|x^r - P_{\tau,l,u}(x^r)\| \leq \hat{\omega} \sum_{h=r}^{r+B-1} \|x^h - x^{h+1}\|$. Since $\|x^r - P_{\tau,l,u}(x^r)\| \leq \sqrt{n} \max_i |x_i^r - P_{\tau,l,u}(x^r)_i|$, we only need to show that there exists a constant $\tilde{\omega} > 0$ such that $|x_i^r - P_{\tau,l,u}(x^r)_i| \leq \tilde{\omega} \sum_{h=r}^{r+B-1} \|x^h - x^{h+1}\|$ holds for each $i \in \{1, 2, \cdots, n\}$.

Note that $Ax^r \in B(Ax^*)$ for sufficiently large r. For any fixed index $i \in \{1, 2, \cdots, n\}$, let $\psi(r,i)$ be the smallest integer N $(N \geq r)$ such that x_i^r is updated on the Nth step. Then, we have

$$\begin{aligned}
&|x_i^r - P_{\tau,l,u}(x^r)_i| \\
&= \left| \sum_{h=r}^{\psi(r,i)-1} \left[(x_i^h - P_{\tau,l,u}(x^h)_i) - (x_i^{h+1} - P_{\tau,l,u}(x^{h+1})_i) \right] + (x_i^{\psi(r,i)} - P_{\tau,l,u}(x^{\psi(r,i)})_i) \right| \\
&\leq \sum_{h=r}^{\psi(r,i)-1} \left| \left[(x_i^h - P_{\tau,l,u}(x^h)_i) - (x_i^{h+1} - P_{\tau,l,u}(x^{h+1})_i) \right] \right| + \left| x_i^{\psi(r,i)} - P_{\tau,l,u}(x^{\psi(r,i)})_i \right|,
\end{aligned}$$

where the inequality follows from the triangle inequality.

It then follows from the the nonexpensive property (2.5) of the projection $P_{\tau,l,u}(x)$, Assumption 3.1(iv) and Theorem 2.2 that

$$|x_i^r - P_{\tau,l,u}(x^r)_i| \leq \sum_{h=r}^{\psi(r,i)-1} \left(2\left| x_i^h - x_i^{h+1} \right| + \left| \nabla_i f(x^h) - \nabla_i f(x^{h+1}) \right| \right) + \alpha_r \left| x_i^{\psi(r,i)} - x_i^{\psi(r,i)-1} \right|.$$

Since $r+1 \leq \psi(r,i) \leq r+B$ by the almost cycle rule, we obtain

$$|x_i^r - P_{\tau,l,u}(x^r)_i| \leq \sum_{h=r}^{r+B-1} \left(2\left|x_i^h - x_i^{h+1}\right| + \left|\nabla_i f(x^h) - \nabla_i f(x^{h+1})\right|\right) + \alpha_r \left|x_i^{\psi(r,i)} - x_i^{\psi(r,i)-1}\right|.$$

It then follows from the Lipschitz continuity of ∇g and Assumption 3.1 that

$$\begin{aligned}|x_i^r - P_{\tau,l,u}(x^r)_i| \leq& (2+\|A\|^2 L) \sum_{h=r}^{r+B-1} \left\|x^h - x^{h+1}\right\| + \alpha_r \left\|x^{\psi(r,i)} - x^{\psi(r,i)-1}\right\| \\ \leq& \left(2+\|A\|^2 L + \frac{\sigma \min_j \|A_j\|^2}{\max_j\{1, L\|A_j\|^2\}}\right) \sum_{h=r}^{r+B-1} \left\|x^h - x^{h+1}\right\|,\end{aligned}$$

where the first inequality follows from $\|x^h - x^{h+1}\| \geq |x_i^h - x_i^{h+1}|$.

Let $\tilde{\omega} := 2+\|A\|^2 L + \frac{\sigma \min_j \|A_j\|^2}{\max_j\{1, L\|A_j\|^2\}}$. Then $\tilde{\omega} > 0$. Thus the inequality of this lemma holds with $\omega = \kappa\sqrt{n}\tilde{\omega}$, where κ is given in Lemma 4.9. □

Now we are ready to show the linear convergence of $\{F(x^r)\}$ and $\{x^r\}$.

Theorem 4.2. *Suppose that $\{x^r\}$ is generated by the ICD method with the almost cycle rule. Let F^* denote the optimal value of the problem (1.1). Then $\{F(x^r)\}$ converges to F^* at least B-step Q-linearly.*

Proof. In the first step, we show the global convergence of the sequence $\{F(x^r)\}$. Let x^* be an optimal solution of the problem (1.1). Then we have $F^* = F(x^*)$. It follows from the mean value theorem that there exists ξ, which is on the line segment that joins x^r with x^*, such that $g(Ax^r) - g(Ax^*) = \langle A^T \nabla g(A\xi), x^r - x^*\rangle$.

Since $Ax^r \to Ax^*$and $\nabla f(x^r) \to d^\infty$, we have

$$d^\infty = \lim_{x\to\infty} \nabla f(x^r) = \lim_{x\to\infty} A^T \nabla g(Ax^r) + b = A^T \nabla g(Ax^*) + b = \nabla f(x^*). \tag{4.37}$$

Thus, we have

$$\begin{aligned}&F(x^r) - F^* \\ =&\langle A^T\nabla g(A\xi) - A^T\nabla g(Ax^*), x^r - x^*\rangle + \langle A^T\nabla g(Ax^*) + b, x^r - x^*\rangle + \sum_{i=1}^n \tau_i(|x_i^r| - |x_i^*|) \\ \leq& L\|A\xi - Ax^*\|\|A(x^r - x^*)\| + \langle A^T\nabla g(Ax^*) + b, x^r - x^*\rangle + \sum_{i=1}^n \tau_i(|x_i^r| - |x_i^*|) \\ \leq& L\|A(x^r - x^*)\|^2 + \langle d^\infty, x^r - x^*\rangle + \sum_{i=1}^n \tau_i(|x_i^r| - |x_i^*|) \\ =& L\|A(x^r - x^*)\|^2 + \sum_{i=1}^n \left[d_i^\infty(x_i^r - x_i^*) + \tau_i(|x_i^r| - |x_i^*|)\right],\end{aligned} \tag{4.38}$$

where the first inequality follows from the Lipschitz continuity of ∇g, and the second inequality follows from (4.37).

With the special structure of the problem (1.1), we can show that for sufficiently large r,

$$d_i^{\infty}(x_i^r - x_i^*) + \tau_i(|x_i^r| - |x_i^*|) = 0, \ \forall i \in \{1, 2, \cdots, n\}. \tag{4.39}$$

We prove this by considering the distinct cases about the index sets $J_j^{\infty}, \ j = \{1, 2, \cdots, 6\}$ since $\{1, 2, \cdots, n\} = \bigcup_{j=1}^{6} J_j^{\infty}$. For simplicity, we only prove the cases $i \in J_1^{\infty}$ and $i \in J_4^{\infty}$. The other cases can be shown in a similar way. If $i \in J_1^{\infty}$, that is, $d_i^{\infty} > \tau_i$, then it follows from Lemma 4.6 that $x_i^r = l_i$ for sufficiently large r. On the other hand, we have $\nabla_i f(x^*) > \tau_i$ by (4.37). It then follows from Lemma 2.1 that $x_i^* = l_i$. These two relations imply that (4.39) holds. If $i \in J_4$, i.e., $d_i^{\infty} = \tau_i$, it then follows from Lemma 4.6 that for sufficiently large r, $l_i \leq x_i^r \leq 0$. On the other hand, we have $\tau_i = \nabla_i f(x^*)$ by (4.37). It further implies that $l_i \leq x^* \leq 0$ from Lemma 2.1. Combining these three relations, we can ensure that (4.39) holds.

Consequently, we have $0 \leq F(x^r) - F^* \leq L\|A(x^r - x^*)\|^2$ by (4.38) and (4.39). It implies $F(x^r) \to F^*$ since $Ax^r \to Ax^*$, that is to say, $\{F(x^r)\}$ is globally convergent.

In the second step, we show the B-step Q-linear convergence rate of $\{F(x^r)\}$. To this end, we need to ensure that there exists a constant $c \in (0, 1)$ such that

$$F(x^{r+B}) - F^* \leq c\left(F(x^r) - F^*\right). \tag{4.40}$$

From (4.38), (4.39) and Lemma 4.10, we have

$$F(x^r) - F^* \leq L\omega^2 \left(\sum_{h=r}^{r+B-1} \|x^h - x^{h+1}\|\right)^2.$$

Letting $k := h - r + 1$, we further have that

$$\begin{aligned} F(x^r) - F^* \leq & L\omega^2 \left(\sum_{k=1}^{B} \|x^{k+r-1} - x^{k+r}\|\right)^2 \\ \leq & L\omega^2 B \sum_{k=1}^{B} \left(\|x^{k+r-1} - x^{k+r}\|\right)^2. \end{aligned}$$

It then follows from Lemma 4.8 that

$$\begin{aligned} F(x^r) - F^* \leq & \frac{L\omega^2 B}{\eta} \sum_{k=1}^{B} \left(F(x^{k+r-1}) - F(x^{k+r})\right) \\ = & \frac{L\omega^2 B}{\eta} \left(F(x^r) - F(x^{r+B})\right). \end{aligned}$$

By rearranging the items of the above inequality, we have

$$F(x^{r+B}) - F^* \leq c\left(F(x^r) - F^*\right), \tag{4.41}$$

where $c = 1 - \frac{\eta}{L\omega^2 B}$. Since $\frac{\eta}{L\omega^2 B} > 0$ and $c < 1$, it means that $\{F(x^r)\}$ converges to F^* at least B-step Q-linearly. $\square$

Theorem 4.3. *Suppose that $\{x^r\}$ is generated by the ICD method with the almost cycle rule. Then there exists an optimal solution x^* of the problem (1.1) such that $\{x^r\}$ converges to x^* at least R-linearly.*

Proof. First we show $\{x^r\}$ is convergent. Let F^* be the optimal value of the problem (1.1). Since $F(x^r)$ converges to F^* at least Q-linearly by Theorem 4.2, we have that $F(x^r)$ converges to F^* at least R-linearly, that is to say, there exist constants $K > 0$, and $\hat{c} \in (0, 1)$ such that

$$F(x^r) - F^* \leq K\hat{c}^r. \tag{4.42}$$

From Lemma 4.8, we have for sufficiently large r,

$$0 \leq \|x^r - x^{r+1}\|^2 \leq \frac{1}{\eta}\left(F(x^r) - F^*\right) + \frac{1}{\eta}\left(F^* - F(x^{r+1})\right) \leq \frac{1}{\eta}\left(F(x^r) - F^*\right), \tag{4.43}$$

where the last inequality holds since $F^* - F(x^{r+1}) \leq 0$.

By combining (4.42) and (4.43), we can deduce that $\|x^r - x^{r+1}\|^2 \leq \frac{K}{\eta}\hat{c}^r$, that is to say , $\|x^r - x^{r+1}\| \leq \sqrt{\frac{K}{\eta}}\hat{c}^{\frac{r}{2}}$. Let $\bar{c} := \hat{c}^{\frac{1}{2}}$. Then, we have $\bar{c} \in (0, 1)$. Moreover, we obtain, for any positive integer m, n and $m > n$,

$$\|x^m - x^n\| \leq \sum_{k=0}^{m-n-1} \|x^{m-k} - x^{m-k-1}\| \leq \sqrt{\frac{K}{\eta}} \sum_{k=0}^{m-n-1} \bar{c}^{m-k-1} = \sqrt{\frac{K}{\eta}} \frac{\bar{c}^n - \bar{c}^m}{1 - \bar{c}} \leq \sqrt{\frac{K}{\eta}} \frac{\bar{c}^n}{1 - \bar{c}}.$$

It implies that $\{x^r\}$ is a cauchy sequence since $0 < \bar{c} < 1$. Therefore, $\{x^r\}$ is convergent.

In the rest, we show that $\{x^r\}$ converges to an optimal solution at least R-linearly. Let x^∞ denote the limit point of $\{x^r\}$. Since $\|x^m - x^n\| \leq \sqrt{\frac{K}{\eta}} \frac{\bar{c}^n - \bar{c}^m}{1 - \bar{c}}$, we have

$$\|x^\infty - x^n\| = \lim_{m \to \infty} \|x^m - x^n\| \leq \lim_{m \to \infty} \sqrt{\frac{K}{\eta}} \frac{\bar{c}^n - \bar{c}^m}{1 - \bar{c}} = \sqrt{\frac{K}{\eta}} \frac{\bar{c}^n}{1 - \bar{c}}.$$

It implies that $\{x^r\}$ converges to x^∞ at least R-linearly since $0 < \bar{c} < 1$. Finally, we can complete the proof by showing that the x^∞ is an optimal solution.

With the continuity of F, we have $\lim_{r \to \infty} F(x^r) = F(x^\infty)$. It then follows from $F(x^r) \to F^*$ in Theorem 4.2 that $F(x^\infty) = F^*$, that is to say, x^∞ is also an optimal solution of the problem (1.1). $\square$

5 Conclusions

In this paper, we have presented a framework of the ICD method for solving l_1-regularized convex optimization (1.1). We also have established the R-linear convergence rate of this method under the almost cycle rule. The key to the ICD method lies in Assumption 3.1 for the "inexact solution". On each iteration step, we only need to find an approximate solution, that raises the possibility to solve general l_1-regularized convex problem.

The proposed ICD method solves an one-dimensional subproblem on each iteration. The Block Coordinate Descent method, which solves a small scale multi-dimensional subproblem, is efficient for some practical problems. Thus it is interesting to extend the proposed ICD method to the "inexact" block CD method.

References

[1] A. Gholami and H. R. Siahkoohi, Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints, Geophysical Journal International, Vol.180, No.2, pp. 871–882, 2010.

[2] A. J. Hoffman, On approximate solutions of systems of linear inequalities, Journal of Research of the National Bureau of standards, Vol. 49, No.4, pp. 263–265, 1952.

[3] D. G. Luenberger, Linear and nonlinear programming, Addision-Wesley, Reading, Massachusetts, 1973.

[4] H. Liu, M. Palatucci, and J. Zhang, Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery, Proceedings of the 26th Annual International Conference on Machine Learning, pp. 649-656, 2009.

[5] J. M. Borwein and A. S. Lewis. Convex analysis and nonlinear optimization: theory and examples, second edition, New York: spinger-verlag, 2000, Canadian mathematical society books in mathematics.

[6] J. M. Ortega and W. C. Rheinboldt, Iterative solution of nonlinear equations in several variables, Reprinted by SIAM, Philadelphia, 2000.

[7] K. Koh, S. J. Kim, and S. Boyd, An interior-point method for large- scale l_1-regularized logistic regression, Journal of Machine Learning Research 8, pp.1519-1555, 2007.

[8] M. A. T. Figueiredo, R. D. Nowak and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE Journal of Selected Topics in Signal Processing, Vol.1, No.4, pp. 586-597, 2007.

[9] M. Y. Park, and T. Hastie, L1-regularization path algorithm for generalized linear models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 69, No.4, pp. 659–677, 2007.

[10] P. Tseng, Approximation accuracy, gradient methods, and error bound for structured convex optimization, Mathematical Programming: Series A and B - 20th International Symposium on Mathematical Programming – ISMP 2009, Vol. 125 N0. 2, pp. 263–295, 2010.

[11] P. Tseng, Convegence of a block coordinate descent method for nondifferentiable minimization, Journal of Optimization and Applications, Vol.109, pp. 475-494, 2001.

[12] P. Tseng and S. Yun. A coordinate gradient descent method for nonsmooth separable minimization, Math Programming, Vol. 117, pp. 387-423, 2009.

[13] R. T. Rockafellar, Convex analysis, Princeton University Press, Princeton, New Jersey, 1970.

[14] T. T. Wu, and K. Lange, Coordinate descent algorithms for lasso penalized regression, The Annals of Applied Statistics, Vol.2, No.1, pp. 224-244, 2008.

[15] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, Bregman iterative algorithms for l_1-minimization with applications to compressed sensing, SIAM J. Imaging Sciences, Vol.1, No.1, pp. 143-168, 2008.

[16] Z. Q. Luo and P. Tseng, On the convergence of the coordinate descent method for convex differentiable minimization, Journal of Optimization and Applications, Vol.72, pp. 7-35, 1992.

[17] Z. Q. Luo and P. Tseng, On the linear convergence of descent methods for convex essentially smooth minimization, SIAM J. Control and Optimization, Vol.30, pp. 408-425, 1992.

[18] Yu. Nesterov, Introductory lectures on convex optimization: a basic course, 2004.