The convergence properties of infeasible inexact proximal alternating linearized minimization

Yukuan Hu$^{1,2}$ & Xin Liu$^{1,2,*}$

$^1$State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; $^2$School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

Email: ykhu@lsec.cc.ac.cn, liuxin@lsec.cc.ac.cn

Received April 21, 2022; accepted December 10, 2022; published online May 17, 2023

Abstract The proximal alternating linearized minimization (PALM) method suits well for solving block-structured optimization problems, which are ubiquitous in real applications. In the cases where subproblems do not have closed-form solutions, e.g., due to complex constraints, infeasible subsolvers are indispensable, giving rise to an infeasible inexact PALM (PALM-I). Numerous efforts have been devoted to analyzing the feasible PALM, while little attention has been paid to the PALM-I. The usage of the PALM-I thus lacks a theoretical guarantee. The essential difficulty of analysis consists in the objective value nonmonotonicity induced by the infeasibility. We study in the present work the convergence properties of the PALM-I. In particular, we construct a surrogate sequence to surmount the nonmonotonicity issue and devise an implementable inexact criterion. Based upon these, we manage to establish the stationarity of any accumulation point, and moreover, show the iterate convergence and the asymptotic convergence rates under the assumption of the Łojasiewicz property.

Keywords proximal alternating linearized minimization, infeasibility, nonmonotonicity, surrogate sequence, inexact criterion, iterate convergence, asymptotic convergence rate, Łojasiewicz property

MSC(2020) 49M27, 65K05, 90C26, 90C30

Citation: Hu Y K, Liu X. The convergence properties of infeasible inexact proximal alternating linearized minimization. Sci China Math, 2023, 66: 2385–2410, https://doi.org/10.1007/s11425-022-2074-7

1 Introduction

In this paper, we focus on the minimization problem with the block structure as

$$
\min_{\bm{x} \in \bigotimes_{i=1}^n \mathbb{R}^{m_i}} f(x_1, \ldots, x_n) \quad \text{s.t.} \quad x_i \in S_i := \{ w_i \in \mathbb{R}^{m_i} : h_i(w_i) \leq 0 \}, \quad i = 1, \ldots, n,
$$

(1.1)

where $f : \bigotimes_{i=1}^n \mathbb{R}^{m_i} \to \mathbb{R}$ is differentiable and not necessarily convex, $\bm{x} := (x_1, \ldots, x_n)$; for $i = 1, \ldots, n$, $x_i \in \mathbb{R}^{m_i}$, $h_i := (h_{i,1}, \ldots, h_{i,p_i})^T : \mathbb{R}^{m_i} \to \mathbb{R}^{p_i}$ is convex differentiable, and $m_i, p_i \in \mathbb{N}$. Problems
sharing this form are ubiquitous (see, e.g., [6,11,12,15,19] and the references therein). We also adopt an extended-valued form of (1.1):

$$\min_{x \in \bigotimes_{i=1}^{n} \mathbb{R}^{m_i}} F(x_1, \ldots, x_n) := f(x_1, \ldots, x_n) + \sum_{i=1}^{n} \delta_{S_i}(x_i),$$

(1.2)

where $\delta_{S_i}$ stands for the indicator function of $S_i$, i.e., $\delta_{S_i}(w)$ equals 0 if $w \in S_i$ and otherwise $\infty$.

In view of the block structure of (1.1), we consider the proximal alternating linearized minimization (PALM) method [5] (see Framework 1, where we impose flexible conditions on the iterate sequence).

**Framework 1**  PALM for solving (1.1)

**Require:** Initial point $x^{(0)} = (x_i^{(0)})_{i=1}^{n} \in \bigotimes_{i=1}^{n} \mathbb{R}^{m_i}$, proximal parameters $\{\sigma_i^{(0)} > 0\}_{i=1}^{n}$.

**Ensure:** An approximate solution $x^{(k)} := (x_i^{(k)})_{i=1}^{n} \in \bigotimes_{i=1}^{n} \mathbb{R}^{m_i}$.

1: Set $k := 0$.
2: while certain conditions are not satisfied do
3:     for $i = 1, \ldots, n$ do
4:         Solve the $i$-th proximal linearized subproblem
5:             $$\min_{x_i \in S_i} \langle \nabla_i f(x_{k+1}^{(k)}, x_{\neq i}^{(k)}), x_i - x_i^{(k)} \rangle + \frac{\sigma_i^{(k)}}{2} \|x_i - x_i^{(k)}\|^2$$
6:             to obtain $x_i^{(k+1)} \in \mathbb{R}^{m_i}$, fulfilling certain conditions.
7:     end for
8:     Update the $i$-th proximal parameter $\sigma_i^{(k)}$ to $\sigma_i^{(k+1)} > 0$ if necessary.
9: end while

When the subproblem (1.3) is exactly solved, we obtain the exact PALM (PALM-E). With properly chosen proximal parameters, one could derive sufficient reduction over the objective value sequence. Based upon this point, the stationarity of any accumulation point follows. This methodology applies to more general frameworks, such as the block successive minimization in [27] and the Bregman distance-based block coordinate proximal gradient methods in [13]. Furthermore, with the aid of the Łojasiewicz property that is shared by a broad swath of functions, one could obtain the iterate convergence in more generic settings (see, e.g., [5,30]).

It is not difficult to check that solving (1.3) in Framework 1 amounts to projecting the point

$$\tilde{x}_i^{(k)} := x_i^{(k)} - \frac{1}{\sigma_i^{(k)}} \langle \nabla_i f(x_{k+1}^{(k)}, x_{\neq i}^{(k)}), x_i - x_i^{(k)} \rangle$$

onto $S_i$. More often, however, the projection is not of closed-form expression. In these contexts, *inexactly* solving (1.3) becomes a much more pragmatic option. Efficient subsolvers for (1.3) could hence be brought to bear.

When the subsolvers inexactly solve (1.3) and yield $x_i^{(k)} \in S_i$ throughout iterations, we obtain the feasible inexact PALM (PALM-F). Most works in this setting enforce the monotonicity of the objective value sequence. Some of them (repeatedly), in one outer iteration, solve the subproblem inexactly to obtain a descent direction and then perform a line search (see, e.g., [6,32]). In [13], Hu and Yamashita treated the solution error as an additional term in the kernel function defining the Bregman distance, and then imposed the assumptions on the solution errors to invoke the results established in the exact settings. Frankel et al. [10] and Ochs [23] put flexibility in solving (1.3) in the sense that the relative error conditions are relaxed while maintaining the sufficient reduction property.

In contrast, little attention has been paid to the infeasible inexact PALM (PALM-I), where the subsolvers inexactly solve (1.3) but not necessarily give $x_i^{(k)} \in S_i$. However, when the constraints describing $\{S_i\}_{i=1}^{n}$ are complicated, infeasible subsolvers, such as (primal-)dual or penalty methods, are indispensable. To illustrate, we list two instances below, along with some state-of-the-art algorithms for computing the projections.
Example 1.1 (Linear constraints). The feasible region $S_i$ is the Birkhoff polytope

$$S_i := \{ W \in \mathbb{R}^{m_i \times m_i} : W1 = 1, W^T1 = 1, W \succeq 0 \},$$

where $1$ stands for the all-one vector in $\mathbb{R}^{m_i}$. This type of feasible region shows up frequently in applications such as optimal transport problems [24] and electronic structure calculations [12]. Since the number of constraints describing $S_i$ is much less than the underlying space dimension (given even moderate $m_i$), it is more reasonable to solve the subproblem (1.3) from the dual perspective. To this end, one may invoke the semismooth Newton method proposed in [18]. By exploiting the structure of $S_i$, high efficiency can be achieved [12]. Nevertheless, the recovered primal solution is infeasible.

Example 1.2 (Nonlinear constraints). The feasible region $S_i$ is an ellipsoid in $\mathbb{R}^{m_i}$, i.e.,

$$S_i := \left\{ w \in \mathbb{R}^{m_i} : \frac{1}{2}w^TA_iw + b_i^Tw \leq \alpha_i \right\},$$

where $I \neq A_i \in \mathbb{R}^{m_i \times m_i}$ is positive definite symmetric, $b_i \in \mathbb{R}^{m_i}$, and $\alpha_i > 0$. Projecting a point onto an ellipsoid emerges as one of the fundamental problems in convex analysis and numerical algorithms with applications in topology optimization [19] and 3-dimensional contact problems with anisotropic friction [15] as well as relations to polynomial optimization [11], just to mention a few. When $b_i = 0$, it is also related to the trust region subproblem in nonlinear optimization [26] and principle component analysis [33]. We refer interested readers to a recent work [14], where an alternating direction method of multipliers is proposed to solve a reformulated problem. The primal variables are then not necessarily feasible upon termination. The proposed method is reported to outperform the existing feasible one in [9].

Owing to the infeasibility, we see that the objective value sequence is not ensured to be monotonic, while the sufficient reduction of the objective value is presumably crucial in proving the stationarity of any accumulation point. The only work exploring the convergence properties of the PALM-I goes to [10]. The obtained results, however, might be of only theoretical values. Frankel et al. [10] imposed the following hypothesis: There exist $\beta_1, \beta_2 > 0$ such that for $i = 1, \ldots, n$ and $k \geq 0$,

$$\sum_{j=1}^{i-1} \| x_j^{(k+1)} - x_j^{(k)} \| + \sum_{j=i}^{n} \| x_j^{(k)} - x_j^{(k)} \| \leq \beta_1 \| x_i^{(k+1)} - x_i^{(k)} \|,$$

(1.5)

where $\bar{x}_i^{(k+1)}$ is the unique solution to (1.3), defined as

$$\bar{x}_i^{(k+1)} := \arg \min_{x_i \in S_i} \left\{ \nabla_i \left( f(x_i^{(k+1)}, x_{\neq i}^{(k)}) \right), x_i - x_i^{(k)} \right\} + \frac{\sigma_i^{(k)}}{2} \| x_i - x_i^{(k)} \|^2.$$

Based upon (1.5), they establish a sufficient reduction result over the objective value sequence $\{ f(\bar{x}^{(k)}) \}$, where $\bar{x}^{(k)} := (\bar{x}_1^{(k)}, \ldots, \bar{x}_n^{(k)})$. It is unclear how to fulfill (1.5) in practice for the reasons that (i) $\bar{x}_i^{(k)}$ and $\bar{x}_i^{(k+1)}$ cannot be computed, not to mention $\| x_i^{(k+1)} - x_i^{(k)} \|$; (ii) $\| x_i^{(k+1)} - x_i^{(k)} \|$ is needed for obtaining $\{ x_j^{(k)} \}_{j=1}^n$. Unfortunately, Frankel et al. [10] did not discuss these points. In consequence, the convergence properties of the PALM-I remain to be investigated, particularly with implementable inexact criteria. This is essential in providing a theoretical guarantee for the usage of efficient infeasible subsolvers.

Contributions. In this paper, we establish the convergence properties of the PALM-I for solving (1.1). In particular, we

(i) control the solution errors when solving (1.3) with a prescribed nonnegative sequence $\{ \varepsilon^{(k)} \}$ and an error bound that is computable for any subsolvers; our inexact criterion is thus much more pragmatic than that in [10];
(ii) construct a nonincreasing surrogate sequence \( \{ v^{(k)} \} \) to surmount the objective value nonmonotonicity issue; the objective value sequence is allowed to fluctuate, favoring more extensive flexibility than most existing works;

(iii) establish the convergence properties, including the iterate convergence to stationarity and the asymptotic iterate convergence rates, of the PALM-I with the help of the Lojasiewicz property of \( F \) in (1.2); these results are new to the best of our knowledge;

(iv) illustrate the considerable advantages of the PALM-I on CPU time over the PALM-E and PALM-F through numerical experiments on problems arising from quantum physics and 3-dimensional anisotropic frictional contact.

Before concluding this section, we gather some of the established asymptotic convergence rates in Table 1 to showcase the comparison with existing works, where \( \theta \) is the Lojasiewicz exponent of \( F \) associated with a compact set (defined later).

**Notations and organization.** This paper presents scalars, vectors and matrices in lower-case letters, bold lower-case letters and upper-case letters, respectively. The notation \( \mathbf{1} \) stands for the all-one vector with a proper dimension. The notations \( \langle \cdot, \cdot \rangle \) and \( \| \cdot \| \) calculate, respectively, the standard inner product and the norm of vectors in the ambient Euclidean space. We use \( \text{Diag}(\cdot) \) to form a diagonal matrix with the input vector.

We use subscripts to denote the components or blocks of vectors or matrices, e.g., \( x_i \) is the \( i \)-th variable block. Occasionally for brevity, we make abbreviation for the aggregation of variable blocks, e.g., \( x_{<i} := (x_1, \ldots, x_{i-1}) \) and \( x_{>i} := (x_{i+1}, \ldots, x_n) \) (clearly, \( x_{<0} \) and \( x_{>n} \) are null variable blocks, which may be used for notational ease). Likewise, we can define \( x_{\leq i}, x_{\geq i}, x_{(j,i]}, x_{[j,i]}, x_{[j,i)} \) and \( x_{[j,i]} \) (the latter four are also null if the index sets in the subscript are empty).

For a function \( h, \nabla h \) (resp. \( \partial h \)) is the gradient (resp. subdifferential) of \( h \) at a certain point where \( h \) is differentiable (resp. subdifferentiable). We add a subscript to indicate the block to which the derivative is taken with respect, e.g., \( \nabla_i h \). For a differentiable mapping \( h : \mathbb{R}^m \rightarrow \mathbb{R}^p \), we denote by \( \nabla h : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times p} \) its Jacobian. The notation \( \delta_S \) stands for the indicator function of a set \( S \), i.e., \( \delta_S(w) = 0 \) if \( w \in S \) and otherwise \( \infty \). We denote the effective domain of a function \( h \) by \( \text{dom}(h) := \{ y : h(y) < \infty \} \). With a slight abuse of notation, the domain of its subdifferential is \( \text{dom}(\partial h) := \{ y : \partial h(y) \neq \emptyset \} \).

Given a set \( S \) and a point \( w \),

\[
\text{dist}(w, S) := \inf_{w' \in S} \|w - w'\|
\]

stands for the distance from \( w \) to \( S \). If the set \( S \) is nonempty closed, we define the projection operator \( \mathcal{P}_S \) onto \( S \) as \( \mathcal{P}_S(w) \in \arg \min_{w' \in S} \|w - w'\| \). The notation “\( \otimes \)” denotes the Cartesian product of sets or spaces. The notation \( B_\eta(x) \) with \( \eta > 0 \) refers to the closed ball in the ambient Euclidean space centered at \( x \) with radius \( \eta \).

**Table 1** Asymptotic convergence rates of the PALM under different settings

| \( \theta \) | \( \varepsilon^{(k)} \) | Extra assumptions | Rates | References |
|---|---|---|---|---|
| 0 | \( \rho^k \) | \( \rho \in (0, 1) \) | \( \mathcal{O}(\rho^k) \), where \( \rho_1 \in (0, 1) \) | Theorem 5.4 |
| \( \frac{1}{(k+1)^2} \) | \( \ell \in (1, \infty) \) | \( \mathcal{O}(k^{-\ell-1}) \) | Theorem 5.6 |
| \( 0 \) | \( \rho^k \) | \( \rho \in (0, 1) \) | \( \mathcal{O}(\rho^k) \), where \( \rho_2 \in (0, 1) \) | [5, 30] |
| \( \frac{1}{(k+1)^2} \) | \( \ell \in (1, \infty) \) | \( \mathcal{O}(k^{-\ell-1}) \) | Theorem 5.4 |
| \( 0 \) | | | \( \mathcal{O}(k^{-\frac{3\theta}{2\theta-1}}) \) | [5, 30] |
| \( \frac{1}{(k+1)^2} \) | \( \ell \in (1, \infty) \) | \( \mathcal{O}(k^{-\frac{3\theta}{2\theta-1}}) \) if \( \ell \geq \frac{\theta}{2\theta-1} \) | Theorem 5.4 |
| | | | \( \mathcal{O}(k^{-\ell-1}) \) if \( \ell < \frac{\theta}{2\theta-1} \) | Theorem 5.6 |
The rest of this paper is organized as follows. In Section 2, we present some definitions used throughout this paper and introduce the Lojasiewicz property. We provide the complete description of the PALM-I in Section 3, including details on the inexact criterion in use. We establish the global convergence properties of the PALM-I in Section 4, including weak and strong forms. We analyze the asymptotic convergence rates of the PALM-I under different settings in Section 5. We give numerical experiments in Section 6. Finally in Section 7, we draw some concluding remarks.

2 Preliminaries

We collect several notions from convex analysis as well as the Lojasiewicz property in this section.

**Definition 2.1** (See [28]). Let $G : E \to (-\infty, \infty]$ be a proper closed function, where $E$ is an Euclidean space. For a given $x \in \text{dom}(G)$, the Fréchet subdifferential of $G$ at $x$, denoted by $\partial G(x)$, is defined as

$$\partial G(x) := \left\{ u \in E : \liminf_{y \neq x, y \to x} \frac{G(y) - G(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$ 

When $x \notin \text{dom}(G)$, we simply set $\partial G(x) = \emptyset$. When $\partial G(x)$ is a singleton, we say that $G$ is Fréchet differentiable at $x$ and denote the derivative by $\nabla G(x)$.

**Remark 2.2.**

(i) If $G : E \to (-\infty, \infty]$ is proper closed convex, then

$$\partial G(x) = \{ u \in E : G(y) - G(x) \geq \langle u, y - x \rangle, \forall y \in E \}, \quad \forall x \in \text{dom}(G).$$

(ii) If $G : E \to (-\infty, \infty]$ and $H : E \to (-\infty, \infty]$ are proper closed functions, and $G$ is Fréchet differentiable at $x$, then $\partial(G + H)(x) = \nabla G(x) + \partial H(x)$.

(iii) If $G : E \to (-\infty, \infty]$ is proper closed and $0 \in \partial G(x)$, we call $x$ a stationary point of $G$.

With the definition of the subdifferential in place, we recall the Lojasiewicz property given in [1]. The Lojasiewicz property is introduced first in [21] on the real analytic functions and then extended to the functions on the o-minimal structure in [16] and to the nonsmooth subanalytic functions in [4] under the name of Kurdyka-Lojasiewicz property afterward [2, 5, 30].

**Definition 2.3** (See [1]). Let $G : E \to (-\infty, \infty]$ be a proper closed function, where $E$ is an Euclidean space. The function $G$ is said to have the Lojasiewicz property at some stationary point $\bar{x}$ if there exist $\epsilon > 0$, $\theta \in [0, 1)$ and $\eta > 0$ such that for any $x \in B_\eta(\bar{x})$,

$$|G(x) - G(\bar{x})|^\theta \leq \epsilon \cdot \text{dist}(0, \partial G(x)),$$

where we adopt the convention $0^0 = 0$ if $\theta = 0$, and therefore, if $|G(x) - G(\bar{x})|^0 = 0$, we have $G(x) = G(\bar{x})$. We call $\theta$ the Lojasiewicz exponent of $G$ at $\bar{x}$.

**Remark 2.4.** Existing works have revealed some valid examples, for example, the real-analytic functions [20], the convex functions fulfilling certain growth conditions [4] and the semialgebraic functions [2]. We refer the readers to [2] for a comprehensive collection. Notably, the class of semialgebraic functions covers a wide range of functions commonly used by the optimization community.

In [1], Attouch and Bolte provided the following uniformized version of the Lojasiewicz property, which could be shown by using the Heine-Borel theorem.

**Lemma 2.5** (See [1]). Let $G : E \to (-\infty, \infty]$ be a proper closed function, where $E$ is an Euclidean space. Let $\Omega \subseteq E$ be a connected compact set consisting of the stationary points of $G$. Assume that $G$ has the Lojasiewicz property at any stationary point. Then $G$ is a constant on $\Omega$ and there exist $c, \eta > 0$, and $\theta \in [0, 1)$ such that for any $\bar{x} \in \Omega$ and $x \in \{ y \in E : \text{dist}(y, \Omega) \leq \eta \}$,

$$|G(x) - G(\bar{x})|^\theta \leq c \cdot \text{dist}(0, \partial G(x)).$$

We call $\theta$ the Lojasiewicz exponent of $G$ (associated with $\Omega$).
The uniformized Lojasiewicz inequality with the constants \( \theta \) and \( c \) in Lemma 2.5 can be chosen in the Lojasiewicz inequality at any single stationary point belonging to \( \Omega \). In fact, there stands a chance that the achievable Lojasiewicz exponent at a single stationary point is smaller than \( \theta \), indicating a sharper local landscape.

When \( x \) satisfies both \( \text{dist}(x, \Omega) < \eta \) and \( |G(x) - G(\bar{x})| < 1 \), one could lift the Lojasiewicz exponent to a larger value, as observed in \([7,17]\).

**Corollary 2.6** (See \([7,17]\)). Let \( G : \mathbb{E} \to (-\infty, \infty] \) be a proper closed function, where \( \mathbb{E} \) is an Euclidean space. Let \( \Omega \subseteq \mathbb{E} \) be a connected compact set consisting of the stationary points of \( G \). Assume that \( G \) has the Lojasiewicz property at any stationary point. Let \( c, \eta > 0 \) and \( \theta \in [0, 1) \) be the constants associated with \( G \) and \( \Omega \) in Lemma 2.5. Then for any \( \theta \in [0, 1) \), for all \( \bar{x} \in \Omega \) and all \( x \in (y \in \mathbb{E} : \text{dist}(y, \Omega) < \eta) \cap \{ G(y) - G(\bar{x}) \} < 1 \), \( |G(x) - G(\bar{x})|^\theta \leq c \cdot \text{dist}(0, \partial G(x)) \). We call \( \theta \) the lifted Lojasiewicz exponent of \( G \) (associated with \( \Omega \)).

In the sequel, we distinguish the lifted exponents from the unlifted ones using overlines as above. We end this section with a list of inequalities for reference, whose proof is omitted.

**Lemma 2.7.** (i) For any \( a_i \geq 0 \) and \( i = 1, \ldots, n \),

\[
\sqrt{n} \prod_{i=1}^{n} a_i \leq \frac{1}{n} \sum_{i=1}^{n} a_i \leq \sqrt{n} \sum_{i=1}^{n} a_i^\frac{1}{2}.
\]

(ii) For any \( a, b \geq 0 \) and \( p \in (1, \infty) \), \( (a + b)^p \leq 2^{p-1}(a^p + b^p) \).

(iii) For any \( a, b \geq 0 \) and \( p \in (0, 1) \), \( (a + b)^p \leq a^p + b^p \).

### 3 PALM-I

We give the complete description of the PALM-I in this section (see Algorithm 2). In comparison with Framework 1, we specify the inexact criterion for subsolvers as well as some additional parameters for determining \( \{G^{(k)}_i\} \). The constants \( M_i \) and \( M_u \), defined later in Section 4, are associated with \( f \), \( \{S_i\}_{i=1}^{n} \) and \( \{\varepsilon^{(k)}\} \). For \( i = 1, \ldots, n \), the residual function \( r_i : \mathbb{R}^m_n \times \mathbb{R} \times \mathbb{R}^m_+ \to \mathbb{R}_+ \) is defined as

\[
r_i(x_i, \lambda_i, \bar{x}_i) := \max\{\langle x_i, x_i - \bar{x}_i + \nabla h_i(x_i)\lambda_i, 0 \rangle, 0\} + \|x_i - \bar{x}_i + \nabla h_i(x_i)\lambda_i\|_{\infty} + \|\max\{h_i(x_i), 0\}\|_{\infty} + \|\lambda_i - h_i(x_i)\|_0.
\]  

(3.1)

**Algorithm 2** PALM-I for solving (1.1)

**Require:** Initial point \( x^{(0)}_i = (x^{(0)}_i)_{i=1}^{n} \in \mathbb{R}^n_+ \), \( \varepsilon > 0 \), nonnegative sequence \( \{\varepsilon^{(k)} \leq \varepsilon\} \), \( M_e \geq M_e > 0 \), initial proximal parameters \( \{\sigma_i^{(0)} \in [M_i, M_u]\}_{i=1}^{n} \).

**Ensure:** An approximate solution \( x^{(k)} := (x^{(k)}_i)_{i=1}^{n} \in \mathbb{R}^n_+ \).

1: Set \( k := 0 \).
2: while certain conditions are not satisfied do
3: \hspace{1em} for \( i = 1, \ldots, n \) do
4: \hspace{2em} Solve the \( i \)-th proximal linearized subproblem

\[
\min_{x_i \in S_i} \langle \nabla f(x^{(k)}_i), x_i - x^{(k)}_i \rangle + \frac{\sigma_i^{(k)}}{2} \|x_i - x^{(k)}_i\|^2
\]

(3.2)

5: \hspace{2em} to obtain \( x^{(k+1)}_i \in \mathbb{R}^m_i \), such that there exists a \( \lambda^{(k+1)} \in \mathbb{R}_+^n \) fulfilling

\[
\sqrt{r_i(x^{(k+1)}_i, \lambda^{(k+1)}_i, \bar{x}^{(k)}_i) \leq \varepsilon^{(k)},
\]

where \( \bar{x}^{(k)}_i \) is computed via (1.4).
6: \hspace{2em} Update the \( i \)-th proximal parameter \( \sigma_i^{(k)} \) to \( \sigma_i^{(k+1)} \in [M_i, M_u] \) if necessary.
7: \hspace{1em} Set \( k := k + 1 \).
8: end while
It is not difficult to observe that $r_i$ measures the violation of Karush-Kuhn-Tucker (KKT) conditions when solving the subproblem (3.2). Hence, the inexact criterion can always be met if (3.2) is solved to sufficient accuracy.

**Remark 3.1.** If we employ primal-dual subsolvers to solve (3.2), the dual variables can just be taken as $\lambda_i^{(k+1)}$ in the PALM-I. Otherwise, one could solve the linear programming

$$
\min_{\lambda \in \mathbb{S}^n_i} 0 \quad \text{s.t.} \quad \begin{cases}
(x_i^{(k+1)}, x_i^{(k+1)} - \tilde{x}_i^{(k)} + \nabla h_i(x_i^{(k+1)})\lambda_i) \leq \left(\frac{\varepsilon(k)}{4}\right)^2,

-\frac{1}{4} - \frac{\varepsilon(k)}{4} \leq x_i^{(k+1)} - \tilde{x}_i^{(k)} + \nabla h_i(x_i^{(k+1)})\lambda_i \leq \left(\frac{\varepsilon(k)}{4}\right)^2, \\
\langle \lambda, h_i(x_i^{(k+1)}) \rangle \leq \left(\frac{\varepsilon(k)}{4}\right)^2
\end{cases}
$$

at any sub-iteration, where $\|\max\{h_i(x_i^{(k+1)}), 0\}\|_\infty \leq \frac{\varepsilon(k)}{4}$.

Below, we explain the inexact criterion in detail based on Example 1.1 in Section 1.

**Example 3.2** (Revisiting Example 1.1). Suppose that $S_i$ is the Birkhoff polytope defined in Example 1.1 and we invoke the semismooth Newton method [18] for the dual of (3.2). In each sub-iteration, the subsolver mainly proceeds as follows:

- **(S1)** Inexactly solve a Newton equation to update trial dual solutions $\lambda_i^{(k+1)}, \lambda_i^{(k+1)} \in \mathbb{R}^{n_i}$ corresponding to the constraints $X_i^1 = 1$ and $X_i^1 1 = 1$, respectively.

- **(S2)** Generate the trial primal solution 

$$X_i^{(k+1)} := \max\{X_i^{(k+1)} + \lambda_i^{(k+1)T} 1 + 1\lambda_i^{(k+1)T}, 0\} \in \mathbb{R}^{n_i \times m_i}$$

and check the stopping criterion.

Now let us discuss the stopping criterion. Let

$$\Phi_i^{(k+1)} := X_i^{(k+1)} - \tilde{X}_i^{(k+1)} - \lambda_i^{(k+1)T} 1 - 1\lambda_i^{(k+1)T} \in \mathbb{R}^{m_i \times m_i},$$

which plays the trial dual solution associated with the nonnegative constraints. From the definition of $\Phi_i^{(k+1)}$, one has 

$$X_i^{(k+1)} - \tilde{X}_i^{(k+1)} - \lambda_i^{(k+1)T} 1 - 1\lambda_i^{(k+1)T} - \Phi_i^{(k+1)} = 0,$$

i.e., the substationarity of (3.2) is fulfilled and the former two terms in (3.1) vanish. As a result, we obtain the following simplified residual function:

$$r_i(X_i^{(k+1)}, \lambda_i^{(k+1)}, \lambda_i^{(k+1)}, \Phi_i^{(k+1)}, \tilde{X}_i^{(k)}) = \|X_i^{(k+1)T} 1 - 1\|_\infty + \|X_i^{(k+1)T} 1 - 1\|_\infty + \max\{\lambda_i^{(k+1)}, X_i^{(k+1)} 1 - 1, 0\} + \max\{\lambda_i^{(k+1)}, X_i^{(k+1)T} 1 - 1, 0\}.$$

One can terminate the subsolver once $\sqrt{r_i}$ falls below the prescribed $\varepsilon(k)$.

The inexact criterion adopted in the PALM-I guarantees an error bound for (3.2) under certain conditions.

**Lemma 3.3.** Suppose that $f$ is continuously differentiable with respect to each variable block over $\mathcal{S}_i := \{w_i \in \mathbb{R}^{m_i} : \text{dist}(w_i, S_i) \leq \varepsilon\}$, where for $i = 1, \ldots, n,

$$S_i := \{w_i \in \mathbb{R}^{m_i} : \text{dist}(w_i, S_i) \leq \varepsilon\}.$$

For $i = 1, \ldots, n$, assume that $S_i$ is convex compact and $h_i$ fulfills one of the following:

- **(i)** $h_i$ is a linear mapping.

- **(ii)** $h_i$ satisfies the Slater constraint qualification, i.e., $h_i(\tilde{x}_i) < 0$ for some $\tilde{x}_i \in \mathbb{R}^{m_i}$, and the Hoffman-like bound

$$\text{dist}(x_i, S_i) \leq c_i \max\{h_i(x_i), 0\}, \quad \forall x_i \in S_i := \left\{w_i \in \mathbb{R}^{m_i} : \text{dist}(w_i, S_i) \leq \frac{\mathcal{M}_i}{M_i}\right\}$$

(3.4)
holds for some constant \( \tilde{c}_i \geq 0 \), where \( M_i := \sup_{x \in \mathbb{R}^{n_i}} \| \nabla_i f(x) \| \).

Let \( \{x^{(k)}\} \) be the iterate sequence generated by the PALM-I. Then there exists a constant \( \omega \geq 0 \) such that \( \|x^{(k+1)} - x^{(k+1)}\| \leq \omega e^{(k)} \) holds for any \( k \geq 0 \).

**Proof.** By [22, Theorem 2.2] and the assumptions on \( \{h_i\}_{i=1}^n \), it holds, for \( i = 1, \ldots, n \) and \( k \geq 0 \), that
\[
\|x_i^{(k+1)} - x_i^{(k+1)}\|^2 \\
\leq \max\{\langle \lambda_i^{(k+1)}, h_i(x_i^{(k+1)}) \rangle, 0\} + \omega_{i,1} \|x_i^{(k+1)} - x_i^{(k)} + \nabla h_i(x_i^{(k+1)})\|_\infty \\
+ \max\{\langle x_i^{(k+1)}, x_i^{(k+1)} - x_i^{(k)} + \nabla h_i(x_i^{(k+1)})\rangle, 0\} + \omega_{i,2} \max\{h_i(x_i^{(k+1)}), 0\}\|_\infty,
\]
where \( \omega_{i,1} := \max_{x_i \in S_i} \| \lambda_i \|_1 \), \( \omega_{i,2} := \min_{\lambda_i \in W_i} \| \lambda_i \|_1 \), and \( W_i \subseteq \mathbb{R}^{n_i} \) is the set of optimal Lagrange multipliers of (3.2). By the Hoffman-like bound (3.4), we obtain from [3, Proposition 3] that \( \sup_k \omega_{i,2}^{(k)} < \infty \). Let \( \omega_i := \max\{1, \omega_{i,1}, \sup_k \omega_{i,2}^{(k)}\} \). The proof is then completed after letting \( \omega := \max_i \sqrt{\omega_i} \) and noticing (3.1).

**Remark 3.4.** Compared with [10, the hypothesis (1.5)], the inexact criterion in Algorithm 2 is much more implementable. From [3], we know that when \( h_i \) is linear (see, e.g., Example 1.1) or satisfies an enhanced version of the Slater constraint qualification (see, e.g., Example 1.2), i.e.,

\[
\exists \tilde{x}_i \in \mathbb{R}^{n_i}, \quad \text{s.t.} \quad h_i(\tilde{x}_i) < 0,
\]

\[
\exists \zeta > 0, \quad \text{s.t.} \quad \left\| y_i - \tilde{x}_i \right\| - \text{dist}(y_i, S_i) \leq \zeta, \quad \forall y_i \in \tilde{S}_i,
\]

the Hoffman-like bound in Lemma 3.3 readily holds. We then bound the solution errors without computing \( \{\tilde{x}^{(k)}\} \).

**Remark 3.5.** Since the inexact criterion described in Algorithm 2 also covers the feasible inexactness, the theoretical results in this paper apply to the PALM-F as well.

## 4 Global convergence properties of the PALM-I

In this section, we investigate the global convergence properties of the PALM-I, including the stationarity of any accumulation point and the iterate convergence.

In the beginning, we state some assumptions and conditions for \( f, \{S_i\}_{i=1}^n, \{h_i\}_{i=1}^n \) and \( \{e^{(k)}\} \).

**Assumption 4.1.** The objective function \( f \) in (1.1) is Lipschitz continuously differentiable with respect to each variable block over \( \mathbb{S}_{i=1}^n S_i \), i.e., for \( i = 1, \ldots, n \), there exists a modulus \( L_i > 0 \) such that for any \( x, x' \in \mathbb{S}_{i=1}^n S_i \),

\[
\|\nabla f(x) - \nabla f(x')\| \leq L_i \|x - x'\|. 
\]

**Assumption 4.2.** For \( i = 1, \ldots, n \), \( S_i \) is convex and compact, and one of the following holds for \( h_i \):

(a) \( h_i \) is linear;

(b) \( h_i \) satisfies the Slater constraint qualification and the Hoffman-like error bound (3.4).

**Condition 4.3.** (a) The sequence \( \{e^{(k)}\} \) is nonnegative square summable.

(b) The sequence \( \{e^{(k)}\} \) is nonnegative summable and there exists a \( \bar{\theta} \in (0, 1) \) such that \( \{e^{(k)}\}^{\bar{\theta}} \) is summable, where \( e^{(k)} := \sum_{k=0}^{\infty} \theta^k (e^{(t)})^2 \) for any \( k \geq 0 \).

**Remark 4.4.** One may find Condition 4.3(b) pretty restrictive at first glance. In fact, given \( \ell > 1 \), the sequence \( \{\theta^k\} \) just meets the demand. Note that for this choice, \( e^{(k)} \) decays as \( O(k^{-2(\ell - 1)}) \). To ensure the summability of \( \{e^{(k)}\}^{\bar{\theta}} \), it then suffices to choose \( \theta > 1 - \frac{1}{2(\ell - 1)} \). We emphasize that what we only require is the existence of such a \( \bar{\theta} \) rather than its explicit value.

One more restrictive but more intuitive alternative for \( \sum_{k=0}^{\infty} \{e^{(k)}\}^{\bar{\theta}} < \infty \) is \( \sum_{k=0}^{\infty} k \{e^{(k)}\}^{2\bar{\theta}} < \infty \). Nonetheless, to retain potential flexibility, we use the one stated in Condition 4.3(b).
Let $L := \max_i L_i$. We specify in the PALM-I that $M_i = \gamma L$, where $\gamma > 1$, and $M_u$ is any scalar not smaller than $M_i$. We further rewrite $M_u$ as $M$ for brevity. Some constants are defined beforehand: $\alpha^{(k)} := \min_i \alpha^{(k)}_i$, $\bar{\alpha}^{(k)} := \max_i \alpha^{(k)}_i$, $\nu := \frac{12}{\sqrt{27}}$, $M := \sqrt{3} (M + L \sqrt{n})$, 

$$C_0^{(k)} := \frac{\alpha^{(k)} - L [1 + \frac{6}{\bar{\alpha}^{(k)}}]}{2}, \quad C_1^{(k)} := \frac{\bar{\alpha}^{(k)} + L [\frac{3}{2} n^2 + (2 + \frac{2}{\bar{\alpha}^{(k)}}) n + 2 \nu + \frac{4}{\bar{\alpha}^{(k)}} + 3]}{2},$$

(4.1)

and $\bar{C}_1 := 2 \omega^2 \max_k C_1^{(k)}$, where $\omega$ is defined in Lemma 3.3. We use the notation “$\Delta$” for the difference between the optimal iterate and the real iterate, e.g., $\Delta x_{(k+1)} := x_{(k+1)} - x_{(k+1)}$ and $\Delta x_{(k+1)} := x_{(k+1)} - x_{(k+1)}$. The proof sketch is as follows:

(i) deduce the approximate sufficient reduction on the objective value sequence;

(ii) deduce the sufficient reduction on a surrogate sequence;

(iii) deduce the approximate relative error bound for the subdifferential;

(iv) show the stationarity of any accumulation point;

(v) show the iterate convergence with the help of the Lojasiewicz property.

We begin with a block-wise lemma.

**Lemma 4.5.** Suppose that Assumption 4.1 holds. Let $\{x^{(k)}\}$ be the iterate sequence generated by the PALM-I. Then for $i = 1, \ldots, n$ and $k \geq 0$,

$$f(x^{(k+1)}_{<i}), x^{(k)}_{i}, x^{(k)}_{>i}) - f(x^{(k+1)}_{<i}), x^{(k+1)}_{i}, x^{(k)}_{>i}) \geq \frac{\alpha^{(k)} - L [1 + \frac{6}{\bar{\alpha}^{(k)}}]}{2} \|x^{(k+1)}_{i} - x^{(k)}_{i}\|^2 - \frac{\bar{\alpha}^{(k)} + L [\frac{3}{2} n^2 + (2 + \frac{2}{\bar{\alpha}^{(k)}}) n + 2 \nu + \frac{4}{\bar{\alpha}^{(k)}} + 3]}{2} \|\Delta x^{(k)}_{i}\|^2.$$ 

(4.2)

**Proof.** The proof mainly leverages Assumption 4.1 and the optimality of $x^{(k+1)}$ in (1.6). We first note that the expression on the left-hand side of (4.2) can be split into five telescoping summations below:

$$I_1 := \sum_{j=1}^{i-1} f(x^{(k+1)}_{<j}, x^{(k+1)}_{j}, x^{(k)}_{>i}) - f(x^{(k+1)}_{<i}), x^{(k+1)}_{i}, x^{(k)}_{>i}),$$

$$I_2 := \sum_{j=i+1}^{n} f(x^{(k+1)}_{<j}, x^{(k)}_{i}, x^{(k)}_{>j}) - f(x^{(k+1)}_{<i}), x^{(k)}_{i}, x^{(k)}_{>j}),$$

$$I_3 := f(x^{(k+1)}_{<i}, x^{(k)}_{i}, x^{(k)}_{>i}) - f(x^{(k+1)}_{<i}, x^{(k+1)}_{i}, x^{(k)}_{>i}),$$

$$I_4 := \sum_{j=1}^{i-1} f(x^{(k+1)}_{<j}, x^{(k+1)}_{j}, x^{(k)}_{>i}) - f(x^{(k+1)}_{<j}, x^{(k)}_{i}, x^{(k)}_{>i}),$$

$$I_5 := \sum_{j=i+1}^{n} f(x^{(k+1)}_{<i}, x^{(k)}_{i}, x^{(k)}_{>j}) - f(x^{(k)}_{<i}, x^{(k)}_{i}, x^{(k)}_{>j}).$$

By Assumption 4.1 and the optimality of $x^{(k+1)}$ in (3.2), we readily have a lower bound for $I_3$:

$$I_3 \geq -\nabla_i f(x^{(k+1)}_{<i}, x^{(k)}_{i}, x^{(k)}_{>i}), x^{(k+1)}_{i} - x^{(k)}_{i}) - \frac{L_i}{2} \|x^{(k+1)}_{i} - x^{(k)}_{i}\|^2$$

$$= -\nabla_i f(x^{(k+1)}_{<i}, x^{(k)}_{i}, x^{(k)}_{>i}), x^{(k)}_{i} - x^{(k)}_{i}) - \frac{L_i}{2} \|x^{(k+1)}_{i} - x^{(k)}_{i}\|^2$$

$$+ \nabla_i f(x^{(k+1)}_{<i}, x^{(k)}_{i}, x^{(k)}_{>i}), x^{(k)}_{i} - x^{(k)}_{i}) - \nabla_i f(x^{(k+1)}_{<i}, x^{(k)}_{i}, x^{(k)}_{>i}), x^{(k)}_{i} - x^{(k)}_{i})$$

$$\geq \frac{\alpha^{(k)} - L [1 + \frac{6}{\bar{\alpha}^{(k)}}]}{2} \|x^{(k+1)}_{i} - x^{(k)}_{i}\|^2 - \frac{\bar{\alpha}^{(k)} + L [\frac{3}{2} n^2 + (2 + \frac{2}{\bar{\alpha}^{(k)}}) n + 2 \nu + \frac{4}{\bar{\alpha}^{(k)}} + 3]}{2} \|\Delta x^{(k)}_{i}\|^2$$

$$- \frac{L [\nu (i - 1) + 2 + \frac{2}{\bar{\alpha}^{(k)}}]}{2} \|\Delta x^{(k+1)}_{i}\|^2.$$
The second inequality also invokes Lemma 2.7(i) and the definition of $L$.

Besides the analyses for the remaining differences are analogous, we merely demonstrate in detail for $I_1$ and $I_4$. By Assumption 4.1,

\[
I_1 \geq \sum_{j=1}^{i-1} \left[ \nabla_j f(x^{(k)}) - \nabla_i f(x^{(k)}) \right] \cdot \Delta x^{(k+1)} - \frac{L}{2} \| \Delta x^{(k+1)} \|_2^2,
\]

\[
I_4 \geq \sum_{j=1}^{i-1} \left[ \nabla_j f(x^{(k)}) - \nabla_i f(x^{(k)}) \right] \cdot \Delta x^{(k+1)} - \frac{L}{2} \| \Delta x^{(k+1)} \|_2^2.
\]

Combining the above two inequalities implies

\[
I_1 + I_4 \geq -L \sum_{j=1}^{i-1} \| \Delta x^{(k+1)} \|_2^2 - L \sum_{j=1}^{i-1} \left\| \begin{pmatrix} x^{(k)}_j - x^{(k)}_i \\ x^{(k+1)}_j - x^{(k+1)}_i \\ x^{(k)}_i - x^{(k+1)}_i \\ x^{(k)}_i - x^{(k)}_i \end{pmatrix} \right\|_2 \| \Delta x^{(k+1)} \|_2 - L \left( i - 1 \right) \sum_{j=1}^{i-1} \| \Delta x^{(k+1)} \|_2^2.
\]

where the first inequality comes from Assumption 4.1 and the definition of $L$, and the second and the last inequalities follow from Lemma 2.7(i). Similar arguments yield a lower bound for $I_2 + I_5$:

\[
I_2 + I_5 \geq -L \left[ \nu(n-i) + 2 \right] \sum_{j=1}^{i-1} \| \Delta x^{(k+1)} \|_2^2 - L \left[ \nu \| x^{(k+1)} - x^{(k)} \|_2^2 + \| \Delta x^{(k)} \|_2^2 \right],
\]

where the first inequality comes from Assumption 4.1 and the definition of $L$, and the second and the last inequalities follow from Lemma 2.7(i). Similar arguments yield a lower bound for $I_2 + I_5$:

\[
I_2 + I_5 \geq -L \left[ \nu(n-i) + 2 \right] \sum_{j=1}^{i-1} \| \Delta x^{(k+1)} \|_2^2 - L \left[ \nu \| x^{(k+1)} - x^{(k)} \|_2^2 + \| \Delta x^{(k)} \|_2^2 \right],
\]

which completes the proof after noticing the definitions of $x^{(k+1)}$ and $x^{(k)}$. 

\[\Box\]
The following approximate sufficient reduction is then a direct corollary.

**Corollary 4.6.** Suppose that Assumption 4.1 holds. Let \( \{x^{(k)}\} \) be the iterate sequence generated by the PALM-I. Then for any \( k \geq 0 \),

\[
    f(\tilde{x}^{(k)}) - f(x^{(k+1)}) \geq C_0^k \| x^{(k+1)} - x^{(k)} \|^2 - C_1^k \| \Delta x^{(k)} \|^2 - C_1^{k+1} \| \Delta x^{(k+1)} \|^2. \tag{4.4}
\]

**Proof.** From Lemma 4.5, we have that by the telescoping summation,

\[
    f(\tilde{x}^{(k)}) - f(x^{(k+1)}) \\
    \geq \sum_{i=1}^{n} \left[ \sigma_i^{(k)} - L \left| 1 + \frac{\nu}{i} \right| \| x^{(k)} - x^{(k)} \|^2 - \frac{\sigma_i^{(k)} + L(2\nu + 3 + \frac{1}{i})}{2} \| \Delta x^{(k)} \|^2 \right] \\
    \geq \sum_{i=1}^{n} \left[ \sigma_i^{(k)} - L \left| 1 + \frac{\nu}{i} \right| \| x^{(k)} - x^{(k)} \|^2 - \frac{\sigma_i^{(k)} + L(2\nu + 3 + \frac{1}{i})}{2} \| \Delta x^{(k)} \|^2 \right] \\
    \geq \min \sigma_i^{(k)} - \frac{\sigma_i^{(k)} + L(2\nu + 3 + \frac{1}{i})}{2} \| x^{(k)} - x^{(k)} \|^2 - \frac{\sigma_i^{(k)} + L(2\nu + 3 + \frac{1}{i})}{2} \| \Delta x^{(k)} \|^2,
\]

which completes the proof by noting the definitions of \( \sigma_i^{(k)} \), \( \tilde{x}^{(k)} \), \( C_0^k \) and \( C_1^k \) in (4.1). \( \square \)

The infeasibility brings additional error terms, in particular, the error term from the last step to the right-hand side of (4.4). Consequently, the objective nonmonotonicity of the PALM-I appears to be inevitable.

Instead of striving to achieve monotonicity, in the spirit of [17, 29, 31], we explicitly include the error terms in an surrogate sequence for which a sufficient reduction result is obtained.

**Proposition 4.7.** Suppose that Assumptions 4.1 and 4.2 hold. Let \( \{x^{(k)}\} \) be the iterate sequence generated by the PALM-I, where \( \{\tilde{x}^{(k)}\} \) fulfills Condition 4.3(a). Then the following assertions hold:

(i) The sequence \( \{v^{(k)} := F(\tilde{x}^{(k)}) + u^{(k)} + u^{(k+1)} \} \) is well defined, where \( u^{(k)} := \sum_{t=k}^{\infty} C_1^{(t)} \| \Delta x^{(t)} \|^2 \).

(ii) For any \( k \geq 0 \), \( v^{(k)} - v^{(k+1)} \geq C_0^k \| x^{(k+1)} - x^{(k)} \|^2 \geq 0 \).

(iii) The sequence \( \{v^{(k)}\} \) converges monotonically to some \( \tilde{F} \), which is attainable for \( F \) over \( \bigotimes_{i=1}^{n} S_i \).

In particular, \( F(\tilde{x}^{(k)}) \to \tilde{F} \) as \( k \to \infty \).

(iv) If there exists an integer \( k \in \mathbb{N} \) such that \( v^{(k)} = \tilde{F} \), then \( v^{(k)} = \tilde{F} \) and \( x^{(k)} = \tilde{x}^{(k+1)} \) for any \( k \geq k \). Moreover, if there exists an integer \( k \geq k \) such that \( x^{(k)} = \tilde{x}^{(k)} \), then one further has \( x^{(k)} = \tilde{x}^{(k+1)} = \tilde{x}^{(k+1)} \) for any \( k \geq k \).

**Proof.** (i) Since \( \{\tilde{e}^{(k)}\} \) is square summable, for each \( k \), we have

\[
    v^{(k)} = \sum_{t=k}^{\infty} C_1^{(t)} \| \Delta x^{(t)} \|^2 \leq \omega^2 \sum_{t=k}^{\infty} C_1^{(t)} (\tilde{e}^{(t)})^2 \leq \tilde{C}_1^k \| \Delta x^{(k)} \|^2 < \infty,
\]

where the first inequality follows from Lemma 3.3. The well-definedness then follows.

(ii) Simply plugging the definition of \( \{v^{(k)}\} \) into Corollary 4.6 leads to the first inequality. The second inequality is due to \( \sigma_i^{(k)} \geq \gamma L \) in the PALM-I, which yields \( C_0^k \geq \frac{\gamma - 1}{4} L \).

(iii) Since \( \{\tilde{e}^{(k)}\} \) is square summable, we have \( u^{(k)} \to 0 \) as \( k \to \infty \). The desired result is then obtained from the sufficient reduction given by (ii), the fact that \( C_0^k \geq \frac{\gamma - 1}{4} L \) for any \( k \geq 0 \), and the lower boundedness and continuity of \( F \) over \( \bigotimes_{i=1}^{n} S_i \).

(iv) The former part follows directly from the statements (ii) and (iii) and the fact that \( C_0^k \geq \frac{\gamma - 1}{4} L \) for any \( k \geq 0 \). To show the latter part, recalling the definition of \( v^{(k)} \), we obtain from \( v^{(k)} = v^{(k+1)} \) that

\[
    v^{(k)} = F(\tilde{x}^{(k)}) + u^{(k)} + u^{(k+1)} = F(\tilde{x}^{(k+1)}) + u^{(k+1)} + u^{(k+2)} = v^{(k+1)},
\]

where the first inequality follows from Lemma 3.3. The well-definedness then follows.
which, combined with \( \bar{x}^{(k+1)} = x^{(k)} = \bar{x}^{(k)} \), yields
\[
0 = u^{(k)} - u^{(k+2)} = C_0^{(k)} \| \Delta x^{(k)} \| + C_0^{(k+1)} \| \Delta x^{(k+1)} \|
\]
Since \( C_0^{(k+1)} \geq \frac{\gamma-1}{\gamma} L > 0 \), we have \( \bar{x}^{(k+1)} = x^{(k+1)} \). By induction, we derive the desired relation for any \( k \geq \bar{k} \).

Next, we seek to prove the approximate relative error bound for the subdifferential.

**Proposition 4.8.** Suppose that Assumption 4.1 holds. Let \( \{x^{(k)}\} \) be the iterate sequence generated by the PALM-I. Then for any \( k \geq 0 \), there exists a \( w^{(k+1)} \in \partial F(x^{(k+1)}) \) such that
\[
\|w^{(k+1)}\| \leq \tilde{M} \|\bar{x}^{(k+1)} - x^{(k)}\| + \|\Delta x^{(k+1)}\|.
\]

**Proof.** For each \( i \in \{1, \ldots, n\} \), it follows from the optimality of \( x_i^{(k+1)} \) in (1.6) and Remark 2.2(iii) that there exists an \( \alpha_i^{(k+1)} \in \partial \delta_{S_i}(x_i^{(k+1)}) \) such that
\[
0 = \nabla_i f(x_{<i}^{(k+1)}, x_i^{(k+1)}, x_{>i}^{(k)}) + \alpha_i^{(k+1)}(x_i^{(k+1)} - x_i^{(k)}) + \sigma_i^{(k)}(x_i^{(k+1)} - x_i^{(k)}) + \sigma_i^{(k)}(x_i^{(k+1)} - x_i^{(k)}).
\]
Using the above relation and the calculus of Fréchet subdifferential, we have
\[
\partial F(x^{(k+1)}) \supseteq \nabla f(x^{(k+1)}) + (\alpha_i^{(k+1)})_{i=1}^n
\]
\[
= (\nabla f(x^{(k+1)}) - \nabla f(x_{<i}^{(k+1)}, x_i^{(k+1)}, x_{>i}^{(k)}) - \sigma_i^{(k)}(x_i^{(k+1)} - x_i^{(k)}) + \sigma_i^{(k)}(x_i^{(k+1)} - x_i^{(k)}))_{i=1}^n.
\]
Let \( w^{(k+1)} := \nabla f(x^{(k+1)}) + (\alpha_i^{(k+1)})_{i=1}^n \). We have that for \( i = 1, \ldots, n, \)
\[
\|\nabla f(x^{(k+1)}) - \nabla f(x_{<i}^{(k+1)}, x_i^{(k+1)}, x_{>i}^{(k)}) - \sigma_i^{(k)}(x_i^{(k+1)} - x_i^{(k)}) + \sigma_i^{(k)}(x_i^{(k+1)} - x_i^{(k)})\|
\]
\[
\leq L \left\| \left( x_{<i}^{(k+1)} - x_{<i}^{(k)} \right) + \sigma_i^{(k)}(x_i^{(k+1)} - x_i^{(k)}) \right\|
\]
\[
\leq L \|\Delta x^{(k+1)}\| + \|\Delta x^{(k)}\| + \sigma_i^{(k)}\|\Delta x^{(k)}\| - \sigma_i^{(k)}\|\Delta x^{(k)}\|,
\]
where the first inequality follows from Assumption 4.1 and the definition of \( L \). Therefore, it follows again from Lemma 2.7(i) that
\[
\|w^{(k+1)}\|^2 = \sum_{i=1}^n [L\|\Delta x^{(k+1)}\| + L\|\Delta x^{(k+1)} - x^{(k)}\| + \sigma_i^{(k)}\|\Delta x^{(k)} - x^{(k)}\|]^2
\]
\[
\leq 3nL^2\|\Delta x^{(k+1)}\|^2 + \|\Delta x^{(k+1)} - x^{(k)}\|^2 + 3\|\Delta x^{(k)} - x^{(k)}\|^2 + 3\|\Delta x^{(k+1)} - x^{(k)}\|^2,
\]
which completes the proof by recalling the definition of \( \tilde{M} \) ahead of (4.1).

In what follows, we prove a weak result, i.e., the stationarity of any accumulation point of \( \{x^{(k)}\} \), assuming that \( \{\varepsilon^{(k)}\} \) is square summable. The accumulation point set \( \Omega(x^{(0)}) \) is defined as
\[
\Omega(x^{(0)}) := \left\{ x = (x_i)_{i=1}^n \in \bigotimes_{i=1}^n \mathbb{R}^{m_i} : \exists K \subseteq N, \text{s.t. } x^{(k)} \to x \text{ in } K \right\}.
\]
Given \( \varepsilon^{(k)} \to 0 \), one has from Lemma 3.3 that \( \|\Delta x^{(k)}\| \to 0 \), which yields
\[
\Omega(x^{(0)}) = \left\{ x = (x_i)_{i=1}^n \in \bigotimes_{i=1}^n \mathbb{R}^{m_i} : \exists K \subseteq N, \text{s.t. } x^{(k)} \to x \text{ in } K \right\}.
\]

**Proposition 4.9.** Suppose that Assumptions 4.1 and 4.2 hold. Let \( \{x^{(k)}\} \) be the iterate sequence generated by the PALM-I, where \( \varepsilon^{(k)} \) fulfills Condition 4.3(a). Then we have
(i) \( \Omega(x^{(0)}) \subseteq \bigotimes_{i=1}^n S_i \) is nonempty and each member is a stationary point of \( F \); (ii) \( \Omega(x^{(0)}) \) is compact connected and \( \text{dist}(x^{(k)}, \Omega(x^{(0)})) \to 0 \); (iii) \( F \) is finite and constant on \( \Omega(x^{(0)}) \).
Proof. (i) Since \( \{ \tilde{x}^{(k)} \} \subseteq \bigotimes_{i=1}^{n} S_i \) and \( S_i \) is bounded (by Assumption 4.2), there exist a subsequence \( \{ \tilde{x}^{(k)} \}_{k \in K} \) and \( \tilde{x} \) such that \( \tilde{x}^{(k)} \to \tilde{x} \) in \( K \), which gives \( \Omega(\tilde{x}^{(0)}) \neq \emptyset \) in view of (4.5). Also note that \( \bigotimes_{i=1}^{n} S_i \) is closed (by Assumption 4.2), thus \( \tilde{x} \in \bigotimes_{i=1}^{n} S_i \) and hence \( \Omega(\tilde{x}^{(0)}) \subseteq \bigotimes_{i=1}^{n} S_i \).

From Proposition 4.7(ii) and the fact that \( C_0^{(k)} \geq \frac{1}{4} L \) for any \( k \geq 0 \), we obtain that for any \( k \geq 1, \)

\[
\frac{\gamma - 1}{4} L \sum_{k=r}^{s} \| \tilde{x}^{(k)} - \tilde{x}^{(k+1)} \|^2 \leq \sum_{k=r}^{s} (v^{(k)} - v^{(k+1)}) = v^{(r)} - v^{(s+1)} \leq v^{(r)} - F \\
\leq (F(\tilde{x}^{(r)}) - F) + 2 \sum_{t=r}^{s} C_1^{(t)} \| \Delta x^{(t)} \|^2 \\
\leq (F(\tilde{x}^{(r)}) - F) + C_1 \sum_{t=r}^{s} (\varepsilon^{(k)})^2,
\]

where the second inequality follows from Proposition 4.7(iii) and the last one is due to (4.1) and Lemma 3.3. By the square summability of \( \{ \varepsilon^{(k)} \} \), the term on the right-hand side is finite for any \( s \). Therefore, \( \| \tilde{x}^{(k+1)} - \tilde{x}^{(k)} \| \to 0 \). Proposition 4.8, combined with Lemma 3.3 and the square summability of \( \{ \varepsilon^{(k)} \} \), further implies \( w^{(k+1)} \to 0 \).

Now pick \( \tilde{x} \in \Omega(\tilde{x}^{(0)}) \) and the associated converging subsequence \( \{ \tilde{x}^{(k+1)} \}_{k \in K} \). Since the subsequence \( \{ (\tilde{x}^{(k+1)}, w^{(k+1)}) \}_{k \in K} \) converges to \( (\tilde{x}, 0) \) and \( F(\tilde{x}^{(k+1)}) \to F(\tilde{x}) \) (by Proposition 4.7(iii) and the continuity of \( F \) in its domain), we deduce from Remark 2.2 that \( 0 \in \partial F(\tilde{x}) \). We complete this statement by the arbitrariness of \( \tilde{x} \).

(ii)-(iii) directly follow from [5, Lemma 5].

To obtain the iterate convergence of the PALM-I when solving the general nonconvex problem (1.2), we assume the Łojasiewicz property for \( F \) in (1.2) and impose stronger assumptions on \( \{ \varepsilon^{(k)} \} \).

Assumption 4.10. The Łojasiewicz property holds for the objective function \( F \) in (1.2) at any stationary point.

By Lemma 2.5, Proposition 4.9 and Assumption 4.10, \( F \) satisfies the uniformized Łojasiewicz property over \( \Omega(\tilde{x}^{(0)}) \), given the square summability of \( \{ \varepsilon^{(k)} \} \). Let \( c, \eta > 0 \) and \( \theta \in [0, 1) \) be the scalars associated with \( \Omega = \Omega(\tilde{x}^{(0)}) \) and \( G = F \).

Theorem 4.11. Suppose that Assumptions 4.1, 4.2 and 4.10 hold. Let \( \{ x^{(k)} \} \) be the iterate sequence generated by the PALM-I, where \( \{ \varepsilon^{(k)} \} \) fulfills Condition 4.3(b). Then the sequence \( \{ x^{(k)} \} \) converges to a stationary point of \( F \).

Proof. Without loss of generality, we assume that \( \tilde{\theta} \geq \theta \) in Condition 4.3(b), i.e., \( \tilde{\theta} \) is lifted from \( \theta \). To show the convergence of \( \{ x^{(k)} \} \), it suffices to prove that \( \{ \| x^{(k)} - x^{(k+1)} \| \} \) has finite length. Note that for any \( s, t \in \mathbb{N} \) \((s \geq t)\),

\[
\sum_{k=s}^{t} \| x^{(k)} - x^{(k+1)} \| \leq \sum_{k=s}^{t} \| x^{(k+1)} - x^{(k)} \| + \sum_{k=s}^{t} \| \Delta x^{(k+1)} \|,
\]

and the assumption

\[
\sum_{k=1}^{\infty} \varepsilon^{(k)} < \infty
\]

already guarantees the finiteness of the second term regardless of \( t \). Hence what remains is to show the summability of \( \{ \| x^{(k+1)} - x^{(k)} \| \} \). We divide the discussion into two cases.

Case I. There exists an integer \( k \in \mathbb{N} \) such that \( v^{(k)} - F = 0 \).
By Proposition 4.7(iv), we deduce that \( v^{(k)} = F \) and \( x^{(k)} = \bar{x}^{(k+1)} \) for any \( k \geq \hat{k} \), giving rise directly to the summability of \( \{\|x^{(k+1)} - x^{(k)}\|\} \). Based upon (4.6) and Proposition 4.9(i), \( \{x^{(k)}\} \) converges to a stationary point of \( F \).

In this case, if there exists an integer \( \hat{k} \in \mathbb{N} (\hat{k} \geq k) \) such that \( x^{(k)} = \bar{x}^{(k)} \), we infer something better. Indeed, by Proposition 4.7(iv), one has \( x^{(k)} = \bar{x}^{(k+1)} = x^{(k+1)} \) for any \( k \geq \hat{k} \). Proposition 4.8 then implies \( u^{(k+1)} = 0 \) for all \( k \geq \hat{k} \). That is to say, \( \{x^{(k)}\} \) finitely terminates at a stationary point of \( F \).

**Case II.** For any \( k \geq 0 \), \( v^{(k)} - F > 0 \).

We prove by showing a recursive relationship for \( \{\|x^{(k+1)} - x^{(k)}\|\} \). Let

\[
\varphi(s) := \frac{c}{1 - \theta} s^{1 - \theta}.
\]

Since \( v^{(k)} - F \) is positive, \( \varphi'(v^{(k)} - F) \) is well defined. By the concavity of \( \varphi \), we derive that

\[
D^{(k,k+1)} := \varphi(v^{(k)} - F) - \varphi(v^{(k+1)} - F) \\
\geq \varphi'(v^{(k)} - F)(v^{(k)} - v^{(k+1)}) \\
\geq c \frac{F(\bar{x}^{(k+1)} - x^{(k)})^2}{\gamma\phi(1)} \\
= \frac{c(\gamma - 1) L}{\gamma\phi(1)} \|x^{(k+1)} - x^{(k)}\|^2 \\
\geq \frac{c(\gamma - 1) L}{\gamma\phi(1)} \|x^{(k+1)} - x^{(k)}\|^2,
\]

where the second inequality is due to Proposition 4.7(ii) and the definition of \( v^{(k)} \), and the last one follows from \( C_0^{(k)} \geq \frac{\gamma - 1}{\gamma\phi(1) L} \) for any \( k \geq 0 \). The above inequality further provides an upper bound, i.e.,

\[
\|x^{(k+1)} - x^{(k)}\| \\
\leq 2 \sqrt{\frac{1}{c(\gamma - 1) L}} \sqrt{|F(\bar{x}^{(k)} - F + u^{(k)} + u^{(k+1)}|\theta D^{(k,k+1)}} \\
\leq \sqrt{\frac{1}{c(\gamma - 1) L}} \left[ \frac{1}{p}\|F(\bar{x}^{(k)} - F + u^{(k)} + u^{(k+1)}|\theta + pD^{(k,k+1)} \right],
\]

where the second inequality comes from Lemma 2.7(i). The constant \( p \) is chosen such that

\[
p \geq \sqrt{\frac{c}{(\gamma - 1) L M}}.
\]

On the other hand, by Lemma 2.7(iii), we obtain

\[
\|F(\bar{x}^{(k)} - F + u^{(k)} + u^{(k+1)}|\theta \cdot 2u^{(k)}|^\theta \\
= |F(\bar{x}^{(k)} - F|\theta + \left( 2 \sum_{t=k}^{\infty} C_1^{(t)} \|x^{(t)}\| \theta \right)^\theta \\
\leq |F(\bar{x}^{(k)} - F|^\theta + C_1^\theta \left( \sum_{t=k}^{\infty} \theta(s)^2 \right)^\theta \\
= |F(\bar{x}^{(k)} - F|^\theta + C_1^\theta \|x^{(k)}\|^\theta,
\]

where the second inequality follows from (4.1) and Lemma 3.3. From the summability of \( \{e^{(k)}\} \), Proposition 4.7(iii) and Proposition 4.9(ii), we know that there exists an integer \( k_1 \in \mathbb{N} \) such that for all \( k \geq k_1 \),

\[
x^{(k)} \in \{ x : \text{dist}(x, \Omega(x^{(0)})) < \eta \} \cap \{ x : |F(x) - F| < 1 \}.
\]

Lemma 2.5, Corollary 2.6 and Proposition 4.8 then yield that for any \( k \geq k_1 \),

\[
c\bar{M} \|x^{(k)} - x^{(k-1)}\| + \|\Delta x^{(k)}\| \geq c \|w^{(k)}\| \geq c \cdot \text{dist}(0, \partial F(x^{(k)})) \geq |F(x^{(k)}) - F|\theta.
\]
Plugging (4.10) into (4.9) and invoking Proposition 4.8, one has that for any $k \geq k_1$,
\[
[F(\bar{x}^{(k)}) - F + u^{(k)} + u^{(k+1)}]^{\bar{\theta}} \leq \frac{cM}{p} \left[ \left\| \bar{x}^{(k)} - x^{(k-1)} \right\| + \left\| \Delta x^{(k)} \right\| + \frac{C^{\bar{\theta}}}{p} (e^{(k)})^{\bar{\theta}} + pD_{k,k+1} \right].
\]  
(4.11)
Inserting (4.11) into (4.7) gives that for any $k \geq k_1$,
\[
\sqrt{c(\gamma - 1) L} \left\| x^{(k+1)} - x^{(k)} \right\| \leq \frac{cM}{p} \left[ \left\| x^{(k)} - x^{(k-1)} \right\| + \left\| \Delta x^{(k)} \right\| + \frac{C^{\bar{\theta}}}{p} (e^{(k)})^{\bar{\theta}} + pD_{k,k+1} \right].
\]
Summing the above inequality over $k$ from $s$ to $t$ ($t \geq s \geq k_1$) yields
\[
\sqrt{c(\gamma - 1) L} \sum_{k=s}^{t} \left\| x^{(k+1)} - x^{(k)} \right\| \leq \frac{cM}{p} \left[ \sum_{k=s}^{t} \left\| x^{(k+1)} - x^{(k)} \right\| + \sum_{k=s}^{t} \left\| \Delta x^{(k)} \right\| + \frac{C^{\bar{\theta}}}{p} \sum_{k=s}^{t} (e^{(k)})^{\bar{\theta}} + pD_{s,t+1} \right].
\]
where the second inequality follows from Lemma 3.3, the assumptions on $\{\epsilon^{(k)}\}$ and $\tilde{\varphi} > 0$ over $(0, \infty)$. Hence,
\[
\sqrt{c(\gamma - 1) L} \sum_{k=s}^{t} \left\| x^{(k+1)} - x^{(k)} \right\| \leq \frac{cM}{p} \left[ \sum_{k=s}^{t} \left\| x^{(s)} - x^{(s-1)} \right\| + \sum_{k=1}^{\infty} \epsilon^{(k)} + \sum_{k=1}^{\infty} \epsilon^{(k)} \phi^{(k)} \right] + \frac{C^{\bar{\theta}}}{p} \sum_{k=1}^{\infty} (e^{(k)})^{\bar{\theta}} + pD_{s,t} \right].
\]  
(4.12)
Note that due to (4.8), the coefficient on the left-hand side of (4.12) is positive. Therefore, in view of the assumptions on $\{\epsilon^{(k)}\}$, (4.12) in fact shows the finite length of $\left\{ \left\| x^{(k+1)} - x^{(k)} \right\| \right\}$. Based upon (4.6) and Proposition 4.9(i), we complete the proof. □

**Remark 4.12.** Lifting the Łojasiewicz exponent $\theta$ is crucial in proving the iterate convergence when the unlifted one equals 0. Just take a look at (4.7). If we keep using $\theta = 0$, we would merely obtain the square summability of $\left\{ \left\| x^{(k+1)} - x^{(k)} \right\| \right\}$ using telescoping-summation arguments. This point is fairly different from the PALM-E because, in the latter case, $u^{(k)} \equiv 0$ and (4.7) would then give rise to a finite termination at a stationary point [5,30].

5 Asymptotic convergence rates of the PALM-I

In this section, we investigate the asymptotic convergence rates of the PALM-I on the basis of Theorem 4.11; henceforth, $\theta$ refers to the Łojasiewicz exponent of $F$ at the unique limit point of $\{x^{(k)}\}$. To derive specific rates, we consider both exponentially and sublinearly decreasing $\{\epsilon^{(k)}\}$.

For notational convenience, let
\[
S^{(t)} := \sum_{k=t}^{\infty} \left\| x^{(k+1)} - x^{(k)} \right\|.
\]
Under the assumptions made in Theorem 4.11,
\[
S^{(0)} < \infty, \quad \sum_{k=1}^{\infty} \left\| \Delta x^{(k)} \right\| < \infty,
\]
and $\{x^{(k)}\}$ converges to some stationary point $\bar{x}$ of $F$ in (1.2). It is then easy to get that for any $t \geq 0$,
\[
\left\| x^{(t)} - \bar{x} \right\| \leq \sum_{k=t}^{\infty} \left\| x^{(k+1)} - x^{(k)} \right\| + \left\| x^{(k+1)} - x^{(k+1)} \right\| \leq S^{(t)} + \sum_{k=t}^{\infty} \epsilon^{(k+1)},
\]  
(5.1)
where the second inequality follows from Lemma 3.3.

If there exists a $K \in \mathbb{N}$ such that $v^{(K)} = \bar{F}$, the asymptotic convergence rates depend only on the choices of $\{\varepsilon^{(k)}\}$ because $S^{(t)} \leq 0$ for all sufficiently large $t$ by Proposition 4.7(iv). We then readily have the following result, whose proof is omitted.

**Theorem 5.1.** Suppose that the assumptions in Theorem 4.11 hold. Let $\bar{x}$ be the unique limit point of the sequence $\{x^{(t)}\}$ generated by the PALM-I. Assume that there exists a $K \in \mathbb{N}$ such that $v^{(K)} = \bar{F}$.

(i) If $\varepsilon^{(k)} = \varepsilon \bar{p}^{k}$ for any $k \geq 0$, where $\bar{p} \in (0, 1)$, then $\|x^{(k)} - \bar{x}\| \leq O(\bar{p}^{k})$ for any $k \geq K$.

(ii) If $\varepsilon^{(k)} = \varepsilon \frac{k}{(k+1)^{1}}$ for any $k \geq 0$, where $\ell > 1$, then $\|x^{(k)} - \bar{x}\| \leq O(k^{-(\ell-1)})$ for any $k \geq K$.

Due to the errors in solving (3.2), $v^{(k)} = \bar{F}$ can hardly happen in implementation. In the remainder of this section, we focus on the cases where $v^{(k)} - \bar{F} > 0$ for any $k \geq 0$. We first derive a universal upper bound on $S^{(t)}$ in this setting.

**Lemma 5.2.** Suppose that the assumptions in Theorem 4.11 hold and $\theta \in [0, 1)$. Assume that $v^{(k)} - \bar{F} > 0$ for any $k \geq 0$. Then we have that for any $t \geq 1$,

$$S^{(t)} \leq [S^{(t-1)} - S^{(t)}] + C_{2}[S^{(t-1)} - S^{(t)} + \omega \varepsilon^{(t)}]^{\frac{1-\theta}{\bar{F}}} + C_{3}E_{\theta}^{(t)},$$

where

$$E_{\theta}^{(t)} := \sum_{k=t}^{\infty} \varepsilon^{(k)} + \sum_{k=t}^{\infty} \frac{c^{(k)} \theta^{1}}{p q (1-\theta)} + (e^{(t)})^{1-\theta},$$

$$p := 2 \sqrt{c \frac{c}{(\gamma-1) M}}, \quad q := \sqrt{c (\gamma-1) L},$$

$$C_{2} := \frac{cp(c \bar{M})^{\frac{1-\theta}{\bar{F}}}}{q (1-\theta)}, \quad C_{3} := \max \left\{ \omega, \frac{c \theta^{1}}{p q (1-\theta)} \right\}.$$

**Proof.** In view of (4.12) in the proof of Theorem 4.11, one has the following upper bound on $S^{(t)}$:

$$S^{(t)} \leq \|x^{(t)} - x^{(t-1)}\| + \omega \sum_{k=t}^{\infty} \varepsilon^{(k)} + \frac{c \theta^{1}}{p q} \sum_{k=t}^{\infty} \varepsilon^{(k)} \theta^{1} \left[F(x^{(t)}) - F^{\bar{F}} + \frac{2}{q (1-\theta)} \sum_{k=t}^{\infty} c^{(k)} \theta^{1} \|\Delta x^{(k)}\|^{2} \right]^{1-\theta} \leq [S^{(t-1)} - S^{(t)}] + \omega \sum_{k=t}^{\infty} \varepsilon^{(k)} + \frac{c \theta^{1}}{p q} \sum_{k=t}^{\infty} \varepsilon^{(k)} \theta^{1} \left[F(x^{(t)}) - F^{\bar{F}} + \frac{C_{1} \theta^{1}}{q (1-\theta)} \right]^{1-\theta},$$

where the second inequality follows from (4.1), Lemma 2.7(iii), Lemma 3.3 and the summability of $\{\varepsilon^{(k)}\}$. By Corollary 2.6 and Proposition 4.8, we obtain

$$|F(x^{(t)}) - F^{\bar{F}}|^{\theta} \leq c \cdot \text{dist}(0, \partial F(x^{(t)})) \leq c \|u^{(t)}\| \leq c \bar{M} \|x^{(t)} - x^{(t-1)}\| + \|\Delta x^{(t)}\| \leq c \bar{M} [S^{(t-1)} - S^{(t)} + \omega \varepsilon^{(t)}],$$

where the last inequality uses Lemma 3.3. Therefore, since $\frac{1-\theta}{\bar{F}} > 0$, we have

$$|F(x^{(t)}) - F^{\bar{F}}|^{1-\theta} \leq (c \bar{M})^{\frac{1-\theta}{\bar{F}}} \left[S^{(t-1)} - S^{(t)} + \omega \varepsilon^{(t)}\right]^{\frac{1-\theta}{\bar{F}}}.$$  

(5.4)

Plugging (5.4) into (5.3), we conclude that

$$S^{(t)} \leq [S^{(t-1)} - S^{(t)}] + \frac{c p(c \bar{M})^{\frac{1-\theta}{\bar{F}}}}{q (1-\theta)} [S^{(t-1)} - S^{(t)} + \omega \varepsilon^{(t)}]^{\frac{1-\theta}{\bar{F}}}.$$
This completes the proof by noting the definitions of $C_2$, $C_3$ and $E^{(t)}_\theta$ in (5.2).

Below, given $\nu^{(k)} - \mathcal{F} > 0$ for any $k \geq 0$, we present the asymptotic convergence rates of the PALM-I under different values of $\theta$ and choices of $\{\varepsilon^{(k)}\}$. Before that, we give a technical lemma, whose proof could be found in [25, Chapter 2, Lemma 4].

**Lemma 5.3.** Let $\{a^{(k)}\}$ be a nonnegative sequence. If

$$a^{(k+1)} \leq \left(1 - \frac{b}{k}\right)a^{(k)} + \frac{d}{k^{s+1}},$$

where $b$, $d$ and $s$ are positive scalars, and $b > s$, then

$$a^{(k)} \leq \frac{d}{b-s} \frac{1}{k^s} + o\left(\frac{1}{k^s}\right).$$

We begin with exponentially decreasing $\{\varepsilon^{(k)}\}$.

**Theorem 5.4.** Suppose that the assumptions in Theorem 4.11 hold with $\varepsilon^{(k)} = \bar{\nu}^k$ for any $k \geq 0$, where $\bar{\nu} \in (0, 1)$. Let $\mathbf{x}$ be the unique limit point of the sequence $\{\mathbf{x}^{(k)}\}$ generated by the PALM-I and $\theta \in [0, 1)$ be the Łojasiewicz exponent of $F$ at $\mathbf{x}$. Assume that $\nu^{(k)} > \mathcal{F}$ for any $k \geq 0$.

(i) If $\theta = 0$, then there exists a $\rho_1 \in (0, 1)$ such that $\|\mathbf{x}^{(k)} - \mathbf{x}\| \leq C_1 \rho_1^k$ for all sufficiently large $k$.

(ii) If $\theta \in (0, \frac{1}{2})$, then there exists a $\rho_2 \in (0, 1)$ such that $\|\mathbf{x}^{(k)} - \mathbf{x}\| \leq C_2 \rho_2^k$ for all sufficiently large $k$.

(iii) If $\theta \in (\frac{1}{2}, 1)$, then $\|\mathbf{x}^{(k)} - \mathbf{x}\| \leq C_3 \rho_3^k$ for all sufficiently large $k$.

**Proof.** (i) and (ii) By the choice of $\{\varepsilon^{(k)}\}$, any $\bar{\nu} \in (0, \frac{1}{2})$ satisfies Condition 4.3(b). Hence, Lemma 5.2 is valid for any $\bar{\nu} \in (0, \frac{1}{2})$ : $\bar{\nu} \geq \theta$. From the proof of Proposition 4.9(i) and the choice of $\{\varepsilon^{(k)}\}$, there exists a $k_3 \in \mathbb{N}$ such that $S^{(t-1)} - S^{(t)} + \omega \varepsilon^{(t)} \in [0, 1)$ for any $t \geq k_3$. Since $\bar{\nu}_t^\bar{\nu} \geq 1$ for any $\bar{\nu} \in (0, \frac{1}{2})$, one has

$$[S^{(t-1)} - S^{(t)} + \omega \varepsilon^{(t)}]^{\frac{1-\theta}{\bar{\nu}} \frac{1}{\bar{\nu}}} \leq S^{(t-1)} - S^{(t)} + \omega \varepsilon^{(t)}, \quad \forall t \geq k_3.$$  

Combining the last inequality and Lemma 5.2, we obtain

$$S^{(t)} \leq \bar{\nu} S^{(t-1)} + \bar{C}_2 \varepsilon^{(t)} + \bar{C}_3 E^{(t)}_\theta, \quad \forall t \geq k_3,$$

where

$$\bar{\nu} := \frac{1 + C_2}{2 + C_2} \in (0, 1), \quad \bar{C}_2 := \frac{\omega C_2}{2 + C_2}, \quad \bar{C}_3 := \frac{C_3}{2 + C_2}.$$  

Invoking the above recursion repeatedly, together with the choice of $\{\varepsilon^{(k)}\}$, yields that for any $t \geq k_3$,

$$S^{(t)} \leq \bar{\nu}^{t-k_3+1} S^{(k_3-1)} + \bar{C}_2 \sum_{k=k_3}^{t} \bar{\nu}^{t-k} \bar{\nu}^{k} + \bar{C}_3 \sum_{k=k_3}^{t} \bar{\nu}^{t-k} E^{(k)}_\theta$$

$$\leq \bar{\nu}^{t-k_3+1} S^{(k_3-1)} + \bar{C}_2 t \max\{\bar{\nu}, \bar{\nu}_t\}^t + \bar{C}_3 \sum_{k=k_3}^{t} \bar{\nu}^{t-k} E^{(k)}_\theta. \quad (5.5)$$

We then proceed with calculations on $E^{(t)}_\theta$: For any $t \geq 1$,

$$E^{(t)}_\theta = \sum_{k=1}^{\infty} \varepsilon^{(k)} + \sum_{k=1}^{\infty} (\varepsilon^{(k)})^{\bar{\nu}} + (\varepsilon^{(t)})^{1-\bar{\nu}} = \frac{\bar{\nu}}{1 - \bar{\nu}} \varepsilon^{\bar{\nu}} + \frac{\varepsilon^{\bar{\nu}} \bar{\nu}^{2\theta t}}{(1 - \bar{\nu})^{\bar{\nu}(1-\bar{\nu}) t}} + \frac{\varepsilon^{(1-\bar{\nu}) \bar{\nu}^{2\theta(1-\bar{\nu})}}}{{(1 - \bar{\nu})^{1-\bar{\nu}}}}.$$

Since $\bar{\nu} \in (0, \frac{1}{2}]$, we have

$$2\bar{\nu} \leq \min\{2(1 - \bar{\nu}), 1\}.$$
Moreover, for any $	heta > 1$, we have \( \frac{\theta}{\theta - 1} < 1 \). Therefore, invoking Lemma 2.7(iii), we get

\[
[S(t) - S(t)]^{\frac{1}{\theta}} + \omega_c(\epsilon(t)]^{\frac{1}{\theta}} 
\leq (S(t) - S(t)]^{\frac{1}{\theta}} + \omega_c(\epsilon(t)]^{\frac{1}{\theta}}.
\]

Moreover, for any $t \geq k_1$, we have

\[
S(t) - S(t) \leq [S(t) - S(t)]^{\frac{1}{\theta}}.
\]

Combining the above two relations with Lemma 5.2, one has that for any $t \geq k_4$,

\[
S(t) - S(t) \leq (S(t) - S(t)]^{\frac{1}{\theta}} + M_2\bar{\rho}^{\frac{1}{\theta}} + M_3\bar{E}_2(t).
\]

With simple calculations, we reach

\[
\omega_c(\epsilon(t)]^{\frac{1}{\theta}} + C_3\bar{E}_2(t) = \omega_c(\epsilon(t)]^{\frac{1}{\theta}} + C_3\left[\frac{\epsilon(t)}{1 - \bar{\rho}} + \frac{\bar{\rho}^{2\theta_t} \bar{\rho}^{2\theta_t}}{(1 - \bar{\rho}^{2\theta})(1 - \bar{\rho}^{2\theta})} + \frac{\epsilon(t)}{(1 - \bar{\rho}^{2\theta})}\right].
\]

Putting the last equality into (5.6), we have the existence of some $\bar{M}_2 > 0$ for which

\[
S(t) - S(t) \leq (2 + C_2)(S(t) - S(t)]^{\frac{1}{\theta}} + \bar{M}_2\bar{\rho}^{\frac{1}{\theta}}t, \quad \forall t \geq k_4.
\]

Since $\frac{\theta}{\theta - 1} > 1$, invoking Lemma 2.7(ii), we obtain from the last inequality that

\[
(S(t) - S(t)]^{\frac{1}{\theta}} \leq 2\frac{\theta - 1}{\theta - 2}[(2 + C_2)(S(t) - S(t)]^{\frac{1}{\theta}} + \bar{M}_2\bar{\rho}^{\frac{1}{\theta}}t], \quad \forall t \geq k_4,
\]

which further yields

\[
S(t) - S(t) \leq C_4(S(t) - S(t)]^{\frac{\theta}{\theta - 1}} + C_5\bar{\rho}^{\frac{1}{\theta}}t, \quad \forall t \geq k_4,
\]

where

\[
C_4 := 2\frac{\theta - 1}{\theta - 2} (2 + C_2)^{\frac{\theta}{\theta - 1}}, \quad C_5 := \bar{M}_2^{\frac{\theta}{\theta - 1}} (2 + C_2)^{-\frac{\theta}{\theta - 1}}.
\]

Let $h_\theta : \mathbb{R}_+ \to \mathbb{R}$ be defined as $h_\theta(x) := x^{\frac{\theta}{\theta - 1}}$. Since $\frac{\theta}{\theta - 1} > 1$, $h_\theta$ is convex on $\mathbb{R}_+$. Hence, for a fixed $\xi > 0$,

\[
(S(t) - S(t)]^{\frac{\theta}{\theta - 1}} - (\xi t^{-\frac{\theta}{\theta - 1}})^{\frac{\theta}{\theta - 1}} = h_\theta(S(t) - S(t)] - h_\theta(\xi t^{-\frac{\theta}{\theta - 1}})]
\]

\[
\geq h_\theta(\xi t^{-\frac{\theta}{\theta - 1}})[S(t) - S(t)] - (\xi t^{-\frac{\theta}{\theta - 1}})]^{\frac{\theta}{\theta - 1}} = \frac{\theta}{(1 - \theta)t}[S(t) - S(t)] - (\xi t^{-\frac{\theta}{\theta - 1}})]^{\frac{\theta}{\theta - 1}}.
\]

Plugging the last inequality into (5.7), one has that for any $t \geq k_4$ (possibly after enlargement),

\[
S(t) - S(t) \leq C_4[(S(t) - S(t)]^{\frac{\theta}{\theta - 1}} - (\xi t^{-\frac{\theta}{\theta - 1}})]^{\frac{\theta}{\theta - 1}} - C_4(\xi t^{-\frac{\theta}{\theta - 1}})]^{\frac{\theta}{\theta - 1}} + C_5\bar{\rho}^{\frac{1}{\theta}}t.
\]
Thus, after choosing $\xi$ such that $C_4 \theta \xi^{\frac{2t+1}{t-\theta}} > \frac{1-\theta}{2\theta-1}$, we conclude from Lemma 5.3 that

$$S(t) \leq O(t^{-\frac{1-\theta}{2\theta-1}}), \quad \forall t \geq k_4,$$

which completes the proof of the statement (iii) after combination with (5.1).

Due to the solution errors in solving (3.2), the finite termination of the PALM-E when $\theta = 0$ seems to go beyond the reach of the PALM-I, no matter how fast $\{\xi(k)\}$ decreases. Note that we separate the statements (i) and (ii) in Theorem 5.4 to indicate that $\rho_1$ and $\rho_2$ can take different values.

When $\{\xi(k)\}$ decays sublinearly, only sublinear rates are achievable, regardless of the value of $\theta$. Below, we first give an auxiliary lemma, whose proof is straightforward and thus omitted.

Lemma 5.5. Let $\theta \in (\frac{1}{2}, 1)$ and $\ell > \frac{\theta+1}{2\theta}$. Define

$$\tau(\theta, \ell) := \min\left\{\frac{1-\theta}{\theta} \ell, \ell - 1, (2\ell - 1)\theta - 1, (2\ell - 1)(1-\theta)\right\}.$$

Then $\tau(\theta, \ell)$ has the following closed-form expression:

$$\tau(\theta, \ell) = \begin{cases} \frac{1-\theta}{\theta} \ell, & \text{if } \ell \in \left[\frac{\theta}{2\theta-1}, \infty\right), \\ (2\ell - 1)\theta - 1, & \text{if } \ell \in \left(\frac{\theta+1}{2\theta}, \frac{\theta}{2\theta-1}\right]. \end{cases}$$

Theorem 5.6. Suppose that the assumptions in Theorem 4.11 hold with $\xi(k) = \frac{\varepsilon}{(k+1)^{\frac{1}{\theta}}}$ for any $k \geq 0$, where $\ell > 1$. Let $\mathbf{x}$ be the unique limit point of the sequence $\{x^{(k)}\}$ generated by the PALM-I and $\theta \in [0, 1)$ be the Lojasiewicz exponent of $F$ at $\mathbf{x}$. Assume that $\omega(\ell) > \bar{F}$ for any $k \geq 0$. Then for all sufficiently large $k$,

$$\|x^{(k)} - \mathbf{x}\| \leq \begin{cases} O(k^{-(\frac{1-\theta}{2\theta-1})}), & \text{if } \ell \in \left[\frac{\theta}{2\theta-1}, \infty\right) \text{ and } \theta \in \left(\frac{1}{2}, 1\right), \\ O(k^{-(\ell-1)}), & \text{otherwise}. \end{cases}$$

Proof. The proof of this theorem is analogous to that of the statement (iii) in Theorem 5.4. By Remark 4.4, any $\bar{\theta} \in [0, 1) : \bar{\theta} > \max\left\{\sqrt{\frac{1-\theta}{2\theta-1}}, \frac{1}{2}\right\}$ complies with Condition 4.3(b). For such $\bar{\theta}$, Lemma 5.2 is valid. From the proof of Proposition 4.9(i) and the choice of $\{\xi(k)\}$, there exists a $k_5 \in \mathbb{N}$ such that

$$S^{(t-1)} - S^{(t)} + \omega\varepsilon^{(t)} \in [0, 1)$$

for any $t \geq k_5$. By the choice of $\{\xi(k)\}$,

$$\omega C_2(\bar{F}^{(t)})^{1-\theta} < C_3 E_\theta^{(t)}$$

$$= \omega C_2(\bar{F}^{(t)})^{1-\theta} + C_3 \left[\sum_{k=1}^{\infty} \varepsilon^{(k)} + \sum_{k=1}^{\infty} \varepsilon^{(k)} + (\varepsilon^{(t)})^{1-\theta}\right]$$

$$\leq \omega C_2(\bar{F}^{(t)})^{1-\theta} + C_3 \left[\sum_{k=1}^{\infty} \frac{\varepsilon}{(k+1)^{\frac{1}{\theta}}} + \sum_{k=1}^{\infty} \frac{\varepsilon^{2\bar{\theta}}}{(2\ell - 1)^{\bar{\theta}}k^{(2\ell-1)\bar{\theta}}} + \left(\sum_{k=1}^{\infty} \frac{\varepsilon^{2}}{(k+1)^{2}}\right)^{1-\theta}\right]$$
with the minimizer in (5.8). Suppose otherwise. It is easy to check that the optimal values of both (A) and (B) are
\[
\min \left\{ \frac{\bar{\theta}}{2\theta - 1}, \tau(\bar{\theta}, \ell) \frac{\bar{\theta}}{1 - \theta} \right\} > 1
\]
due to $\bar{\theta} > \sqrt{\frac{1}{2\theta - 1}}$. Thus, after choosing $\xi$ such that
\[
C_4\xi \frac{2^{\bar{\theta}-1}}{1 - \theta} > \min \left\{ \frac{\bar{\theta}}{2\theta - 1}, \tau(\bar{\theta}, \ell) \frac{\bar{\theta}}{1 - \theta} \right\} - 1,
\]
we conclude from Lemma 5.3 that
\[
S^{(t)} \leq O(t^{-\min\left(\frac{\bar{\theta}}{1 - \theta}, \tau(\bar{\theta}, \ell) \frac{\bar{\theta}}{1 - \theta}\right)-1}), \quad \forall t \geq k_5.
\]
Invoking Lemma 5.5 and combining (5.1), we derive
\[
\|\hat{x}^{(t)} - \bar{x}\| \leq \begin{cases} O(t^{-\frac{\bar{\theta}}{1 - \theta}}), & \text{if } \ell \in \left[\frac{\bar{\theta}}{2\theta - 1}, \infty\right), \\ O(t^{-\frac{2\bar{\theta} - 1}{2\theta - 1}}), & \text{if } \ell \in \left(\frac{\bar{\theta}^2 + 1}{2\bar{\theta}^2}, \frac{\bar{\theta}}{2\theta - 1}\right), \end{cases} \quad \forall t \geq k_5. \tag{5.9}
\]
Since (5.9) holds for any $\bar{\theta} \in [\theta, 1) : \bar{\theta} > \max\left\{\sqrt{\frac{1}{2\theta - 1}}, \frac{1}{2}\right\}$ and $k_5$ does not rely on its value, the best rate exponent must be attained at one of the following two:
\[
\begin{align*}
(A) \quad & \max_{\bar{\theta} \in [\sqrt{\frac{1}{2\theta - 1}}, \frac{1}{2}], \bar{\theta} \geq \theta} \frac{1 - \bar{\theta}}{2\theta - 1} ; \\
(B) \quad & \max_{\bar{\theta} \in \left(\sqrt{\frac{1}{2\theta - 1}}, \frac{1}{2}\right], \bar{\theta} \geq \theta} \frac{2\bar{\theta}^2 - \ell - \bar{\theta}^2 - 1}{2\theta - 1}. 
\end{align*}
\]
Note that $\bar{\theta} \geq \frac{\ell}{2\theta - 1}$ holds for any $\bar{\theta} \geq \theta$ if and only if $\ell \geq \frac{\theta}{2\theta - 1}$ and $\theta \in (\frac{1}{2}, 1)$. Suppose $\ell \geq \frac{\theta}{2\theta - 1}$ and $\theta \in (\frac{1}{2}, 1)$. Then the best rate exponent is achieved at (A) with just $\bar{\theta} = \theta$, establishing the first line in (5.8). Suppose otherwise. It is easy to check that the optimal values of both (A) and (B) are $\ell - 1$ with the minimizer $\bar{\theta} = \frac{\ell}{2\theta - 1}$, leading to the second line of (5.8). The proof is completed.

The first line of (5.8) recovers the result for the PALM-E [5,30]. In view of this and the statement (iii) in Theorem 5.4, it appears that the solution errors in solving (3.2) do not affect the asymptotic rates of the PALM-I at all if $\varepsilon^{(k)}$ decreases fast enough.

6 Numerical experiments

In this section, we use numerical results to validate the convergence of the PALM-I and demonstrate its merits over the PALM-E and PALM-F. All the numerical experiments presented are run in a platform with Intel(R) Xeon(R) Gold 6242R CPU @ 3.10GHz and 510GB RAM running MATLAB R2018b under Ubuntu 20.04.
6.1 Optimization with linear constraints

The first class of problems under consideration arises from quantum physics [12]. More specifically, strongly correlated quantum systems can be well described by multi-marginal optimal transport problems with Coulomb cost (MMOT). Since the number of unknowns in MMOT scales exponentially with the number of electrons, one could adopt the so-called Monge-like ansatz, which is of significant physical interest, to transform the MMOT into a mathematical programming with generalized complementarity constraints, waiving the curse of dimensionality. Penalizing the troublesome generalized complementarity constraints finally leads to

\[
\min_{\{X_i\}_{i=1}^N} \sum_{i=2}^N \langle X_i, \Lambda C \rangle + \sum_{i<j} \left( \langle X_i, \Lambda X_j C \rangle + \beta \langle X_i, X_j \rangle \right)
\]

s.t. \( X_i \in \mathcal{S} := \{W \in \mathbb{R}^{K \times K} : W1 = 1, W^T \mathbf{g} = \mathbf{g}, \text{Tr}(W) = 0, W \succeq 0\}, \quad \forall i \).

(6.1)

Here, \( \beta > 0 \) is the penalty parameter, and \( N, K \in \mathbb{N} \) refer, respectively, to the number of electrons in the system and finite elements \( \mathcal{T} := \{\{e_k\}_{k=1}^K \subseteq \mathbb{R}^d \} \) discretizing a bounded domain. The vector \( \mathbf{g} := [g_1, \ldots, g_K]^T \in \mathbb{R}^K \) is defined as \( g_k := \int_{\mathcal{T}} \rho(r) \text{d}r \) for any \( k \in \{1, \ldots, K\} \), where \( \rho : \mathbb{R}^d \to \mathbb{R}_+ \) is the single-electron density of the system. The diagonal matrix \( \Lambda := \text{Diag}(\mathbf{g}) \) and \( C = (C_{ij}) \) denotes the discretized Coulomb cost matrix whose diagonal elements are all set to zero to avoid numerical instability, i.e.,

\[
C_{ij} := \begin{cases} 
\|d_i - d_j\|^{-1}, & \text{if } i \neq j, \\
0, & \text{otherwise}
\end{cases}
\]

with \( \{d_k\}_{k=1}^K \subseteq \mathbb{R}^d \) being the barycenters of elements \( \{e_k\}_{k=1}^K \). For brevity, let \( \mathcal{B} : \mathbb{R}^{K \times K} \to \mathbb{R}^{2K+1} \) be a linear operator defined as

\[
\mathcal{B}(W) := [W^T, \mathbf{g}^T W, \text{Tr}(W)]^T \in \mathbb{R}^{2K+1}, \quad \forall W \in \mathbb{R}^{K \times K}
\]

and \( \mathbf{b} := [1^T, \mathbf{g}^T, 0]^T \in \mathbb{R}^{2K+1} \). Then the set \( \mathcal{S} \) can be expressed simply as \( \mathcal{S} := \{W : \mathcal{B}(W) = \mathbf{b}, W \succeq 0\} \).

This type of constraint has been mentioned in Example 1.1. Solving (6.1) is of great importance for describing correlation among electrons. In our experiments, we consider a 1-dimensional system with \( N = 3 \) electrons and the domain \([-1, 1]\); the density is a normalized Gaussian, i.e.,

\[
\rho(x) \propto \exp(-x^2/\sqrt{\pi}), \quad \forall x \in \mathbb{R}.
\]

We adopt an equi-mass discretization so that all the entries in \( \mathbf{g} \) are identical.

We compare the performances of the PALM-E and PALM-I when solving (6.1) with \( K = 36 \); that is to say, the number of variables equals \( 36^2 	imes 2 = 2,592 \). The proximal parameter is fixed at \( \sigma_0(k) \equiv \sigma = 10^{-2} \) for \( i = 1, \ldots, n \) and \( k \geq 0 \). In both the PALM-E and PALM-I, we adapt the semismooth Newton-CG (ssncg) proposed in [18] to efficiently compute the projection \( \mathcal{P}_\mathcal{S} \). As noted in Example 1.1, the infeasibility is inevitable. In light of (3.3), the residual function \( r_i \) becomes

\[
\max\{\langle \lambda_i^{(k+1)}, \mathcal{B}(X_i^{(k+1)}) - \mathbf{b}\rangle, 0\} + \|\mathcal{B}(X_i^{(k+1)}) - \mathbf{b}\|_\infty,
\]

where \( \lambda_i^{(k+1)} \in \mathbb{R}^{2K+1} \) is an approximate dual solution given by ssncg. In the PALM-E, we set \( \{\epsilon(k) \equiv 10^{-7}\}_{k \geq 0} \) such that all the subproblems are solved to high accuracy, whereas in the PALM-I, we pick a nonincreasing sequence \( \{\epsilon(k) = \max\{10^{-1}, 10^{-7}\}\}_{k \geq 0} \) with \( \ell = 0.75 \). Note that by Assumption 4.2, the stationarity point of (6.1) can be characterized by KKT conditions. The outer PALM-E or PALM-I framework is therefore stopped once the relative KKT violation is smaller than \( 10^{-6} \).

We first compare the performances of the PALM-E and PALM-I with random initializations. The built-in “rand” function in MATLAB is invoked to generate 100 initial points, and then we plot out the averaged history of the relative KKT violation for both the PALM-E and PALM-I with respect to the CPU time and iteration number (see (a) and (c) of Figure 1). Although both algorithms require similar
iteration numbers, the average CPU time used by the PALM-I is about 0.46 seconds, in sharp contrast to approximately 15.97 seconds for the PALM-E. The considerable superiority of the PALM-I in terms of CPU time is evidently gained by inexact solving subproblems while maintaining convergence. Since (6.1) is nonconvex, it is interesting and necessary to inspect the differences between the terminating objective values of the PALM-E and PALM-I. We plot in (b) and (d) of Figure 1 the averaged history of relative objective difference from the optimal value [8] for both the PALM-E and PALM-I with respect to the CPU time and iteration number. We can observe that starting with randomly generated initial points, the PALM-E and PALM-I often stop at points of similar qualities.

We then conduct a performance comparison between the PALM-E and PALM-I with good initializations. The good initial points could be generated by random perturbation on the discretized optimal solution $X^\star$ supplied in [8]. The built-in “rand” function is invoked again to generate 100 good initial points, whose entry-wise deviations from $X^\star$ are at most $10^{-3}$. Analogously, we plot out the averaged history of the relative KKT violation and objective difference from the optimal value for both the PALM-E and PALM-I with respect to the CPU time and iteration number (see Figure 2). The average CPU time used by the PALM-E is about 0.85 seconds, while that of the PALM-I is only approximately 0.03 seconds. These reflect the time advantage of the PALM-I over the PALM-E in a neighborhood of the optimal solution. Incidentally, the infeasible nature of the PALM-I does not ruin much the solution quality; the maximum absolute difference between the terminating objective value and the optimal one is merely $2.39 \times 10^{-4}$, and it is less than $10^{-5}$ on over 75% samples.

Figure 1 (Color online) With random initializations, the averaged history of relative KKT violation and relative objective difference from the optimal value for the PALM-E and PALM-I with respect to the CPU time and iteration number when solving (6.1)
Figure 2 (Color online) With good initializations, the averaged history of relative KKT violation and relative objective difference from the optimal value for the PALM-E and PALM-I with respect to the CPU time and iteration number when solving (6.1).

The numerical results in this subsection reflect that the PALM-I converges well and is clearly more efficient than the PALM-E even with infeasibility. The efficiency is brought by the infeasible subsolver ssncg, whose usage is ensured by our theoretical results.

6.2 Optimization with nonlinear constraints

The second class of testing problems involves nonconvex quadratic objective functions and multiple ellipsoidal constraints, having the form

$$\begin{align*}
\min_x & \frac{1}{2} \langle x, Ax \rangle + \langle b, x \rangle \\
\text{s.t.} & \frac{1}{2} \langle x_i, B_i x_i \rangle + \langle c_i, x_i \rangle \leq 1, \quad i = 1, \ldots, n, 
\end{align*}$$

(6.2)

where for $i = 1, \ldots, n$, $m_i = m \in \mathbb{N}$; $A \in \mathbb{S}^{m \times n}$, while $\{B_i\}_{i=1}^n \subseteq \mathbb{S}^{m \times m}$; $b \in \mathbb{R}^{mn}$ and $\{c_i\}_{i=1}^n \subseteq \mathbb{R}^m$.

This problem class is related to several domains [11, 15, 19], as noted in Example 1.2. The ellipsoidal constraints can model the correlated uncertainty range in topology optimization [19] and can arise in the dual variational formulation of 3-dimensional contact problems with anisotropic frictions [15]. Solving the problem (6.2) is thus meaningful for these applications. In our implementation, $A$ and $b$ are generated by the built-in function “randn” in MATLAB. To form $\{B_i = (b_{i,j,k})\}_{i=1}^n$, we adopt the construction in [9, 14]:

$$b_{i,j,k} = 10^{-\frac{i-1}{\text{cond}}}, \quad \text{if } j = k; \quad b_{i,j,k} = 0, \quad \text{otherwise.}$$
It is easy to see that \(n_{\text{cond}} \in \mathbb{R}_{++}\) controls the condition number and the spectrum of each \(B_i\) is spread in \([1, 10^{n_{\text{cond}}}]\).

We compare the performances of the PALM-E, PALM-F and PALM-I when solving (6.2) with \(n = 5\) and \(m = 500\), i.e., the number of variables is 2,500. We select
\[
\{n_{\text{cond}}\}_{i=1}^n = \{3.00, 3.25, 3.50, 3.75, 4.00\}.
\]
The vectors \(\{\mathbf{c}_i\}_{i=1}^n\) are set to be all-zero so that all the ellipsoids are concentric. The proximal parameter is fixed at \(\sigma_i^{(k)} = \sigma = 1\) for \(i = 1, \ldots, n\) and \(k \geq 0\). The three algorithms are armed with different subsolvers. Specifically, both the PALM-E and PALM-I invoke the self-adaptive alternating direction methods of the multiplier (sadmm) proposed in [14]; the PALM-F uses the feasible hybrid projection (hp) algorithm in [9]. Note that the iterates produced by sadmm are not necessarily feasible. For \(i = 1, \ldots, n\) and \(k \geq 0\), we terminate hp within the PALM-F if
\[
\|\mathbf{x}_i^{(k+1)} - \bar{x}_i^{(k)} + \mathbf{\lambda}_i^{(k+1)} (B_i \mathbf{x}_i^{(k+1)} + \mathbf{c}_i)\| \leq \frac{\eta}{2}\|\mathbf{x}_i^{(k+1)} - \mathbf{x}_i^{(k)}\|_{\infty},
\]
where \(\eta = 0.99\sigma\) and \(\mathbf{\lambda}_i^{(k+1)} \geq 0\) estimates the multiplier associated with the ellipsoidal constraint. It is not difficult to verify that the above inexact criteria help produce iterates fulfilling the assumption in [13]. Regarding sadmm in the PALM-E and PALM-I, we see that the residual function \(r_i\) becomes
\[
\max\{\langle \mathbf{x}_i^{(k+1)} - \bar{x}_i^{(k)}, B_i \mathbf{x}_i^{(k+1)} + c_i \rangle, 0\}
+ \|\mathbf{x}_i^{(k+1)} - \bar{x}_i^{(k)} + \mathbf{\lambda}_i^{(k+1)} (B_i \mathbf{x}_i^{(k+1)} + \mathbf{c}_i)\|_{\infty}
+ \mathbf{\lambda}_i^{(k+1)} \max \left\{ - \frac{1}{2} \langle \mathbf{x}_i^{(k+1)}, B_i \mathbf{x}_i^{(k+1)} \rangle + \langle \mathbf{c}_i, \mathbf{x}_i^{(k+1)} \rangle - \alpha_i, 0 \right\}
+ \max \left\{ \frac{1}{2} \langle \mathbf{x}_i^{(k+1)}, B_i \mathbf{x}_i^{(k+1)} \rangle + \langle \mathbf{c}_i, \mathbf{x}_i^{(k+1)} \rangle - \alpha_i, 0 \right\},
\]
where \(\mathbf{\lambda}_i^{(k+1)} \in \mathbb{R}_{+}\) is an approximate dual solution given by sadmm. In the PALM-E, we set \(\varepsilon^{(k)} \equiv 10^{-6}\) for the PALM-I, we choose
\[
\varepsilon^{(k)} = \max \left\{ \frac{10^{-1}}{(k+1)\ell}, 10^{-6} \right\},
\]
with \(\ell = 0.75\). As in the previous subsection, the three outer frameworks are stopped once the relative KKT violation is smaller than \(10^{-5}\).

We invoke the built-in “randn” function in MATLAB to generate 100 random initial points and then draw the averaged history of the relative KKT violation and relative objective difference from \(\{f_1^*\}_{t=1}^{100}\) for the three algorithms, where \(f_1^*\) refers to the minimum of the three converged values in the \(t\)-th trial \((t = 1, \ldots, 100)\) (see Figure 3). The respective average CPU times used by the PALM-E, PALM-F and PALM-I are approximately 30.15 seconds, 2.96 seconds and 1.74 seconds. One can observe that the PALM-I takes the strengths of the infeasible subsolver sadmm and stands out with the best performance. Moreover, it appears that even equipped with an infeasible solver, the PALM-I is capable of yielding objective values much closer than the PALM-F to those of the PALM-E.

\section{Conclusions}

We recognize by examples the indispensability of infeasible subsolvers in the PALM whenever constraints are complicated and illustrate through numerical simulations that the PALM-I can be far more efficient than the PALM-E and PALM-F. The shortage of existing works on the PALM-I motivates us to analyze its convergence properties, particularly in the presence of objective value nonmonotonicity. We achieve this by constructing a monotonically decreasing surrogate sequence. Moreover, an implementable inexact criterion for subsolvers is devised for practical usage.
Future improvements can be anticipated in several lines. For example, one could incorporate nonsmooth regularization terms into the objective function and handle infeasibility and nonsmoothness simultaneously. Besides, the assumptions, such as the Hoffman-like error bound, may appear restrictive and call for further relaxation. Last but not least, it is worth investigating the convergence properties of the PALM-I on problems with nonconvex constraints and designing implementable inexact criteria for those contexts.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 12125108, 11971466, 11991021, 11991020, 12021001 and 12288201), Key Research Program of Frontier Sciences, Chinese Academy of Sciences (Grant No. ZDBS-LY-7022) and CAS (the Chinese Academy of Sciences) AMSS (Academy of Mathematics and Systems Science)-PolyU (The Hong Kong Polytechnic University) Joint Laboratory of Applied Mathematics. The authors thank the referees for their helpful suggestions and comments.

References

1 Attouch H, Bolte J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math Program, 2009, 116: 5–16
2 Attouch H, Bolte J, Redont P, et al. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. Math Oper Res, 2010, 35: 438–457
3 Bertsekas D P. A note on error bounds for convex and nonconvex programs. Comput Optim Appl, 1999, 12: 41–51
4 Bolte J, Danilidis A, Lewis A. The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J Optim, 2007, 17: 1205–1223
5 Bolte J, Sabach S, Teboulle M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems.
6 Bonettini S, Prato M, Rebegoldi S. A block coordinate variable metric linesearch based proximal gradient method. Comput Optim Appl, 2018, 71: 5–52
7 Chill R. On the Lojasiewicz-Simon gradient inequality. J Funct Anal, 2003, 201: 572–601
8 Colombo M, De Pascale L, Di Marino S. Multimarginal optimal transport maps for one-dimensional repulsive costs. Canad J Math, 2015, 67: 350–368
9 Dai Y H. Fast algorithms for projection on an ellipsoid. SIAM J Optim, 2006, 16: 986–1006
10 Frankel P, Garrigos G, Peypouquet J. Splitting methods with variable metric for Kurdyka-Lojasiewicz functions and general convergence rates. J Optim Theory Appl, 2015, 165: 874–900
11 He S M, Li Z N, Zhang S Z. Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. Math Program, 2010, 125: 353–383
12 Hu Y K, Chen H J, Liu X. A global optimization approach for multimarginal optimal transport problems with Coulomb cost. SIAM J Sci Comput, 2023, in press
13 Hua X Q, Yamashita N. Block coordinate proximal gradient methods with variable Bregman functions for nonsmooth separable optimization. Math Program, 2016, 160: 1–32
14 Jia Z H, Cai X J, Han D R. Comparison of several fast algorithms for projection onto an ellipsoid. J Comput Appl Math, 2017, 319: 320–337
15 Kučera R. Convergence rate of an optimization algorithm for minimizing quadratic functions with separable convex constraints. SIAM J Optim, 2008, 19: 846–862
16 Kurdyka K. On gradients of functions definable in o-minimal structures. Ann Inst Fourier (Grenoble), 1998, 48: 769–783
17 Li X, Milzarek A, Qiu J W. Convergence of random reshuffling under the Kurdyka-Lojasiewicz inequality. SIAM J Optim, 2023, in press
18 Li X D, Sun D F, Toh K-C. On the efficient computation of a generalized Jacobian of the projector over the Birkhoff polytope. Math Program, 2020, 179: 419–446
19 Liu B S, Jiang C, Li G Y, et al. Topology optimization of structures considering local material uncertainties in additive manufacturing. Comput Methods Appl Mech Engrg, 2020, 360: 112786
20 Lojasiewicz S. Une propriété topologique des sous-ensembles analytiques réels. In: Les équations aux dérivées partielles, vol. 117. Paris: Éditions du Centre National de la Recherche Scientifique, 1963, 87–89
21 Lojasiewicz S. Sur la géométrie semi-et sous-analytique. Ann Inst Fourier (Grenoble), 1993, 43: 1575–1595
22 Mangasarian O L, De Leone R. Error bounds for strongly convex programs and (super)linearly convergent iterative schemes for the least 2-norm solution of linear programs. Appl Math Optim, 1988, 17: 1–14
23 Ochs P. Unifying abstract inexact convergence theorems and block coordinate variable metric iPiano. SIAM J Optim, 2019, 29: 541–570
24 Peyré G, Cuturi M. Computational optimal transport: With applications to data science. Found Trends Mach Learn, 2019, 11: 355–607
25 Polyak B T. Introduction to Optimization. New York: Optimization Software, 1987
26 Powell M J D, Yuan Y. A trust region algorithm for equality constrained optimization. Math Program, 1991, 49: 189–211
27 Razaviyayn M, Hong M Y, Luo Z-Q. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. SIAM J Optim, 2013, 23: 1126–1153
28 Rockafellar R T, Wets R J B. Variational Analysis. Heidelberg-Berlin: Springer, 1998
29 Sun T, Jiang H, Cheng L Z, et al. A convergence framework for inexact nonconvex and nonsmooth algorithms and its applications to several iterations. arXiv:1709.04072, 2017
30 Xu Y Y, Yin W T. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM J Imaging Sci, 2013, 6: 1758–1789
31 Yang L. Proximal gradient method with extrapolation and line search for a class of nonconvex and nonsmooth problems. arXiv:1711.06831, 2017
32 Yang Y, Pesavento M, Luo Z-Q, et al. Inexact block coordinate descent algorithms for nonsmooth nonconvex optimization. IEEE Trans Signal Process, 2019, 68: 947–961
33 Zhou S Y, Bai Y Q. Convergence analysis of Oja's iteration for solving online PCA with nonzero-mean samples. Sci China Math, 2021, 64: 849–868