FULL SPLITTING ALGORITHMS FOR FRACTIONAL PROGRAMS WITH STRUCTURED NUMERATORS AND DENOMINATORS *

RADU IOAN BOŢ †, GUOYIN LI ‡, AND MIN TAO §

Abstract. In this paper, we consider a class of nonconvex and nonsmooth fractional programming problems that involve the sum of a convex, possibly nonsmooth function composed with a linear operator and a differentiable, possibly nonconvex function in the numerator and a convex, possibly nonsmooth function composed with a linear operator in the denominator. These problems have applications in various fields. We present a framework for a full-splitting proximal subgradient algorithm with two versions: (i) a smoothing-based version (S-FSPS) that uses carefully chosen smoothing parameters and step sizes; and (ii) an adaptive version (Adaptive FSPS) which incorporates extrapolation and backtracking to ensure the nonnegativity of the merit sequence. Both versions address the difficulty of decoupling the nonsmooth composition in the numerator. We prove that S-FSPS converges subsequentially to an exact lifted stationary point, and that Adaptive FSPS converges globally to an approximate lifted stationary point under the Kurdyka-Łojasiewicz property. Further discussions are provided on the tightness of the Adaptive FSPS convergence results and the reasoning behind aiming for an approximate lifted stationary point. We construct a series of counterexamples to demonstrate that the Adaptive FSPS algorithm may diverge when seeking exact solutions. We also developed practical versions incorporating a non-monotone line search to enhance performance. Our theoretical findings are validated through simulations involving limited-angle CT reconstruction and the robust sharp-ratio-type minimization problem.

Key words. structured fractional programs, full splitting algorithm, convergence analysis, lifted stationary points, Kurdyka-Łojasiewicz property, nonmonotone line search

MSC codes. 90C26, 90C32, 49M27, 65K05

1. Introduction.

In this paper, we consider the following class of nonsmooth and nonconvex fractional programs:

$$\min_{\mathbf{x}\in\mathcal{S}} F(\mathbf{x}) := \frac{g(A\mathbf{x}) + h(\mathbf{x})}{f(K\mathbf{x})}, \tag{1.1}$$

where $\mathcal{S}$ is a nonempty convex and compact subset of $\mathbb{R}^n$, $f : \mathbb{R}^p \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^s \to \overline{\mathbb{R}}$ are proper, nonsmooth convex and lower semicontinuous functions; $A : \mathbb{R}^n \to \mathbb{R}^s$ and $K : \mathbb{R}^n \to \mathbb{R}^p$ are linear operators; $h : \mathbb{R}^n \to \mathbb{R}$ is a (possibly nonconvex) differentiable function over an open set containing $\mathcal{S}$ and its derivative ∇h is Lipschitz continuous over this open set with a Lipschitz constant $L_{\nabla h}$. To ensure (1.1) is well-defined, we assume for the denominator that $K\mathbf{x} \in \mathrm{dom} f$ and $f(K\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{S}$. For more detailed assumptions, we direct the reader to our subsequent sections.

Problem (1.1) falls into the category of single-ratio fractional programming problems. However, its structure is more intricate than that of the problems discussed in [8] and [10]. When the linear operators A and K are identity mappings (represented

*Submitted to the editors DATE.

Funding: The research of RIB has been partially supported by the Austrian Science Fund (FWF), project number W1260-N35. The research of GL has been partially supported by the Australian Research Council (project number: DP190100555, DP250101112). The research of MT has been partially supported by the Natural Science Foundation of China (No. 12471289), by the China Scholarship Council (202006195015) and it has been conducted during her research visit at the University of Vienna in 2023.

†Faculty of Mathematics, University of Vienna, A-1090 Vienna, Austria, radu.bot@univie.ac.at.

‡Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia, g.li@unsw.edu.au.

§School of Mathematics, National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, Republic of China, taom@nju.edu.cn.

as I), the model (1.1) simplifies to the problem addressed in [21]. Model (1.1) encompasses a variety of optimisation problems, such as limited-angle CT reconstruction [19, 29], robust Sharpe ratio minimization [13], the single-period optimal portfolio selection problem [23], the sparse signal reconstruction problem [28, 21, 34], and so on. Subsequently, we offer two examples to demonstrate the nature of (1.1).

(a) The *limited-angle CT reconstruction problem* aims at reconstructing the true image from limited-angle scanning measurements. By representing an image as an $(n \times n)$ matrix, it can be mathematically formulated as

$$\min_{\mathbf{x}\in\mathcal{B}} \frac{\tau\|\nabla\mathbf{x}\|_1 + \frac{1}{2}\|P\mathbf{x}-\mathbf{f}\|^2}{\|\nabla\mathbf{x}\|_p}, \tag{1.2}$$

Let $1 < p < \infty$, P be the projection operator, $\mathbf{f}$ the oberved data, and $\tau > 0$ a regularization parameter. The linear operator $\nabla : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ denotes the discrete gradient operator, defined as $\nabla\mathbf{v} = (\nabla_{\mathbf{x}}\mathbf{v}, \nabla_{\mathbf{y}}\mathbf{v})$, where $\nabla_{\mathbf{x}}, \nabla_{\mathbf{y}} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ are the forward horizontal and vertical difference operators, respectively. Regarding the box constraint $\mathcal{B} := [\mathbf{l}, \mathbf{u}] \subseteq \mathbb{R}^{n\times n}$, which represents the range of pixel values of the true image [29], we assume that $\mathcal{B} \cap \operatorname{span}(\mathbf{E}) = \varnothing$, where $\mathbf{E}$ is the matrix with all entries equal to one. By identifying the matrix space $\mathbb{R}^{n\times n}$ as the Euclidean space $\mathbb{R}^{n^2}$, problem (1.2) can be written as a special case of (1.1) with $g(\mathbf{x}) := \tau\|\mathbf{x}\|_1$, $f(\mathbf{x}) := \|\mathbf{x}\|_p$, $A = K = \nabla$, $h(\mathbf{x}) := \frac{1}{2}\|P\mathbf{x}-\mathbf{f}\|^2$, and $\mathcal{S} := \mathcal{B}$. Here, $\|\cdot\|_p$ denotes the usual ℓ_p-norm for $1 < p < \infty$, while for $p = 2$ we will simply write $\|\cdot\|$ for $\|\cdot\|_2$.

(b) The *robust sharp-ratio-type optimization problem* under scenario data uncertainty, which arises in finance, takes the following form:

$$\min_{\mathbf{x}\in\Delta} \frac{\max_{1\leqslant i\leqslant m_1}\{r_i - \mathbf{a}_i^\top\mathbf{x}\}}{\max_{1\leqslant i\leqslant m_2} \mathbf{x}^\top C_i\mathbf{x}}, \tag{1.3}$$

where $\Delta = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{e}^\top\mathbf{x} = 1, \mathbf{x} \geqslant 0\}$ with $\mathbf{e} = (1,\ldots,1) \in \mathbb{R}^n$, $(\mathbf{a}_i, r_i) \in \mathbb{R}^n \times \mathbb{R}$, $i = 1,\ldots,m_1$, are such that $r_i - \mathbf{a}_i^\top\mathbf{x} \geqslant 0$ for all $\mathbf{x} \in \Delta$, and C_i, $i = 1,\ldots,m_2$, are positive definite matrices. The standard Sharpe ratio optimization problem without data uncertainty reads as (see [13]) $\max_{\mathbf{x}\in\Delta} \frac{\mathbf{a}^\top\mathbf{x}-r}{\sqrt{\mathbf{x}^\top C\mathbf{x}}}$. Another closely related equivalent model is $\max_{\mathbf{x}\in\Delta} \frac{\mathbf{a}^\top\mathbf{x}-r}{\mathbf{x}^\top C\mathbf{x}}$, where $C \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix and $(\mathbf{a}, r) \in \mathbb{R}^n \times \mathbb{R}$. Here, without loss of generality, we assume that $\mathbf{a}^\top\mathbf{x} - r \geqslant 0$ for all $\mathbf{x} \in \Delta$. Suppose that the data $(\mathbf{a}, r)$ and C are subject to scenario uncertainty, that is, $(\mathbf{a}, r) \in \mathcal{U}_1 = \{(\mathbf{a}_1, \overline{r}_1), \ldots, (\mathbf{a}_{m_1}, \overline{r}_{m_1})\}$ and $C \in \mathcal{U}_2 = \{C_1,\ldots,C_{m_2}\}$, where $(\mathbf{a}_i, \overline{r}_i) \in \mathbb{R}^n \times \mathbb{R}$, $i = 1,\ldots,m_1$, are such that $\mathbf{a}_i^\top\mathbf{x} - \overline{r}_i \geqslant 0$ for all $\mathbf{x} \in \Delta$ and C_i, $i = 1,\ldots,m_2$, are positive definite matrices. Then, the robust counterpart of the above Sharpe ratio optimization problem is

$$\max_{\mathbf{x}\in\Delta} \min_{\substack{(a,r)\in\mathcal{U}_1,\\ C\in\mathcal{U}_2}} \frac{\mathbf{a}^\top\mathbf{x}-r}{\mathbf{x}^\top C\mathbf{x}} = \max_{\mathbf{x}\in\Delta} \frac{\min_{1\leqslant i\leqslant m_1}\{\mathbf{a}_i^\top\mathbf{x} - \overline{r}_i\}}{\max_{1\leqslant i\leqslant m_2} \mathbf{x}^\top C_i\mathbf{x}},$$

which can be further equivalently rewritten as $\min_{\mathbf{x}\in\Delta} \frac{\max_{1\leqslant i\leqslant m_1}\{\overline{r}_i - \mathbf{a}_i^\top\mathbf{x}\}}{\max_{1\leqslant i\leqslant m_2} \mathbf{x}^\top C_i\mathbf{x}}$. By adding a positive constant if necessary (without affecting the solutions), we obtain (1.3). The problem of (1.3) is a special case of (1.1) with $f(\mathbf{x_1},\cdots,\mathbf{x}_{m_2}) := \max_{1\leqslant i\leqslant m_2}\|\mathbf{x}_i\|_2^2$, $K : \mathbf{x} \mapsto (C_1^{1/2}\mathbf{x},\cdots,C_{m_2}^{1/2}\mathbf{x})$, $g(\mathbf{x}) := \|\mathbf{r}-\mathbf{x}\|_\infty$ with $\mathbf{r} := (r_1,\ldots,r_{m_1})$, $A : \mathbf{x} \mapsto (\mathbf{a}_1^\top\mathbf{x}, \mathbf{a}_2^\top\mathbf{x},\cdots,\mathbf{a}_{m_1}^\top\mathbf{x})^\top$, $h(\mathbf{x}) = 0$ and $\mathcal{S} := \Delta$.

The conventional approach to tackling single ratio fractional programming problems commonly involves utilizing Dinkelbach's method or its variants [15, 18]. The recent monograph [14] comprehensively explores Dinkelbach's algorithm, incorporating surrogation mechanism to overcome the inherent nonconvexity of the resultant subproblems. For solving simple single ratio problems, where compositions of nonsmooth functions with linear operators do not occur, various splitting algorithms have been proposed in recent works [8, 9, 10, 21, 35]. These methods share the feature that instead of invoking an inner loop aimed at solving the resulting Dinkelbach's scalarization of the fractional program, they execute only one iteration of a suitable splitting algorithm and update the sequence of function values. On the other hand, direct adaptations of these techniques for solving (1.1) often lead to double-loop algorithms.

Our goal is to develop a single-loop, full-splitting algorithm with convergence guarantees for efficiently solving problem (1.1). By *fully splitting*, we mean that the algorithm relies solely on the proximity operators of either g or g^*, and either f or f^*. To address this challenge, inspired by [12, 21], we propose a framework for the *Fully Splitting Proximal Subgradient* (FSPS) algorithm. Specifically, we introduce two iterative schemes. The first one is based on a smoothing approach with carefully selected step sizes and smoothing parameters to ensure (subsequential) convergence to an exact lifted stationary point. The second one is an adaptive algorithm with an extrapolated step that enjoys global convergence guarantees, albeit with the trade-off of convergence to an approximate lifted stationary point.

The smoothing-based algorithm, called S-FSPS, uses a smooth approximation of the nonsmooth function g through the Moreau envelope g_γ (defined in (2.2)) as $\gamma \downarrow 0$ [5]. By carefully choosing the step sizes and the smoothing parameters, we show that a cluster point of S-FSPS is an exact lifted stationary point. On the other hand, this scheme need not exhibit global convergence for the whole sequence. To address this issue, we propose an adaptive algorithm, called Adaptive FSPS, that approximates $g \circ A$ from below using the conjugate function of g. Adaptive FSPS incorporates a backtracking strategy to maintain the positivity of the augmented function sequence — a crucial property for convergence analysis. Additionally, an extrapolated step [36, 37] is introduced to maintain positivity in the augmented function values. We establish sequential convergence to an exact lifted stationary point for Adaptive FSPS when g is smooth and satisfies the KL property. When g is nonsmooth, we demonstrate sequential convergence to an approximate lifted stationary point under the KL property. We justify the convergence of the adaptive FSPS to an approximate lifted stationary point when g is nonsmooth. The approximation error can be set to an arbitrarily small value. Counterexamples are constructed to demonstrate that Adaptive FSPS may diverge to an exact stationary point, regardless of whether the smoothing parameter γ_k tends to zero as $k \to +\infty$ or is set to zero. Unlike existing splitting methods for nonconvex problems [11, 17, 20, 24, 30], global convergence for Adaptive FSPS is guaranteed without requiring full-rank assumptions on the linear operators. Furthermore, we propose practical versions of these algorithms by incorporating a nonmonotone line search [32, 35] to improve performance.

The remainder is organized as follows. Section 2 presents the necessary notions and results. Section 3 introduces the stationarity concepts and investigates their interrelationships. In Section 4, we develop a framework of fully splitting proximal subgradient (FSPS) algorithm, propose a smoothing-based version (S-FSPS), and establish its subsequential convergence to an exact lifted stationary point—although without a guarantee of global convergence. Section 6 is devoted to the development of an adaptive FSPS algorithm and the establishment of its global convergence to an approximate lifted stationary point under the Kurdyka–Łojasiewicz (KL) assumption. In Section 7, we discuss several important aspects related to the conceptual

FSPS algorithm and its variants. Section 8 introduces practical adaptations via the integration of a nonmonotone line search strategy, and presents numerical results demonstrating their effectiveness. Finally, Section 9 concludes the paper.

2. Preliminaries and calculus rules.

Finite-dimensional spaces within the paper will be equipped with the Euclidean norm, denoted by $\|\cdot\|$, while $\langle\cdot,\cdot\rangle$ will represent the Euclidean scalar product. Given a set $\mathcal{C} \subseteq \mathbb{R}^n$, $\mathrm{ri}(\mathcal{C})$, $\mathrm{int}(\mathcal{C})$ and $\mathrm{cl}(C)$ denote its *relative interior*, *interior* and *closure*, respectively. The function $\iota_{\mathcal{C}} : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, defined by $\iota_{\mathcal{C}}(\mathbf{x}) = 0$, for $\mathbf{x} \in \mathcal{C}$, and $\iota_{\mathcal{C}}(\mathbf{x}) = +\infty$, otherwise, denotes the *indicator function* of the set $\mathcal{C}$.

For a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we denote by $\mathrm{dom}\, f := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < +\infty\}$ its *effective domain* and say that it is *proper* if $\mathrm{dom}\, f \neq \emptyset$. For $\overline{\mathbf{x}} \in \mathrm{dom}\, f$, the set

$$\hat{\partial} f(\overline{\mathbf{x}}) := \left\{ \mathbf{v} \in \mathbb{R}^n : \liminf_{\mathbf{x}\to\overline{\mathbf{x}}\ \mathbf{x}\neq\overline{\mathbf{x}}} \frac{f(\mathbf{x}) - f(\overline{\mathbf{x}}) - \langle \mathbf{v}, \mathbf{x} - \overline{\mathbf{x}}\rangle}{\|\mathbf{x} - \overline{\mathbf{x}}\|} \geqslant 0 \right\}$$

is the so-called *Fréchet subdifferential* of f at $\overline{\mathbf{x}}$. The *limiting subdifferential* of f at $\overline{\mathbf{x}}$ is defined as

$$\partial f(\overline{\mathbf{x}}) := \left\{ \mathbf{v} \in \mathbb{R}^n : \exists\ \{\mathbf{x}^k\} \to \overline{\mathbf{x}},\ f(\mathbf{x}^k) \to f(\overline{\mathbf{x}}), \{\mathbf{v}^k\} \to \mathbf{v} \text{ as } k \to +\infty,\ \mathbf{v}^k \in \hat{\partial} f(\mathbf{x}^k) \right\}.$$

If f is proper, convex and lower semicontinuous function and $\varepsilon \geqslant 0$, we denote by

$$\partial_\varepsilon f(\overline{\mathbf{x}}) := \{\mathbf{v} \in \mathbb{R}^n : f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}}) + \langle \mathbf{v}, \mathbf{x} - \overline{\mathbf{x}}\rangle - \varepsilon\ \forall \mathbf{x} \in \mathbb{R}^n\} \tag{2.1}$$

the ε-*subdifferential* of f at $\overline{\mathbf{x}}$. It holds $\mathbf{v} \in \partial_\varepsilon f(\overline{\mathbf{x}})$ if and only if $f^*(\mathbf{v}) + f(\overline{\mathbf{x}}) - \langle \mathbf{v}, \overline{\mathbf{x}}\rangle \leqslant \varepsilon$, where $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$, $f^*(\mathbf{v}) = \sup_{\mathbf{x}\in\mathbb{R}^n}\{\langle \mathbf{v}, \mathbf{x}\rangle - f(\mathbf{x})\}$, denotes the *(Fenchel) conjugate function* of f. The *convex subdifferential* of f at $\overline{\mathbf{x}}$ is defined by $\partial f(\overline{\mathbf{x}}) := \partial_0 f(\overline{\mathbf{x}})$. The *domain* of the convex subdifferential is defined as $\mathrm{dom}(\partial f) := \{\mathbf{x} \in \mathbb{R}^n : \partial f(\mathbf{x}) \neq \emptyset\}$. For a proper, convex lower semicontinuous function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, its *proximal operator of modulus* $\gamma > 0$ is defined as

$$\mathrm{Prox}_{\gamma f} : \mathbb{R}^n \to \mathbb{R}^n,\ \mathrm{Prox}_{\gamma f}(\mathbf{x}) = \arg\min_{\mathbf{y}\in\mathbb{R}^n} \left\{ f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{x}\|^2 \right\}.$$

The *Moreau envelope of* f *with modulus* $\gamma > 0$ is defined as

$$f_\gamma : \mathbb{R}^n \to \mathbb{R}, \quad f_\gamma(\mathbf{x}) := \min_{\mathbf{y}\in\mathbb{R}^n} \left\{ f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{x}\|^2 \right\}. \tag{2.2}$$

For all $\mathbf{x} \in \mathbb{R}^n$, it holds that

$$f_\gamma(\mathbf{x}) = \left(f^* + \frac{\gamma}{2}\|\cdot\|^2 \right)^* (\mathbf{x}) = \sup_{\mathbf{v}\in\mathbb{R}^n} \left\{ \langle \mathbf{x}, \mathbf{v}\rangle - f^*(\mathbf{v}) - \frac{\gamma}{2}\|\mathbf{v}\|^2 \right\}. \tag{2.3}$$

The Moreau envelope of f with modulus $\gamma > 0$ is Fréchet differentiable on $\mathbb{R}^n$, and its gradient satisfies, for all $\mathbf{x} \in \mathbb{R}^n$, $\nabla(f_\gamma)(\mathbf{x}) = \frac{1}{\gamma}\left(\mathbf{x} - \mathrm{Prox}_{\gamma f}(\mathbf{x})\right) = \mathrm{Prox}_{f^*/\gamma}\left(\frac{\mathbf{x}}{\gamma}\right)$. A proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *essentially strictly convex* if it is strictly convex on every convex subset of $\mathrm{dom}(\partial f)$. For a proper, convex and lower semicontinuous function f, f is essentially strictly convex if and only if its conjugate f^* is essentially smooth [26]. Given a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$, we denote by $A^* : \mathbb{R}^m \to \mathbb{R}^n$ its *adjoint operator*. We also use $\sigma_A := \|A\| = \sup\{\|A\mathbf{x}\| : \|\mathbf{x}\| = 1\}$ to denote its norm. Given $r > 0$ and $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{B}(\mathbf{x}, r)$ denotes the closed ball centered at $\mathbf{x}$ with radius r. Next, we review the Kurdyka-Łojasiewicz (KL) property [3, 6] and the concept of calmness [27].

DEFINITION 2.1. *A proper and lower semicontinuous function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is said to satisfy the Kurdyka-Łojasiewicz (KL) property at a point* $\hat{\mathbf{x}} \in \mathrm{dom}(\partial f)$ *if there exist a constant* $\mu \in (0, +\infty]$*, an open neighborhood* U *of* $\hat{\mathbf{x}}$*, and a desingularization function* $\phi : [0, \mu) \to [0, +\infty)$*, which is continuous and concave, and continuously differentiable on* $(0, \mu)$ *with* $\phi(0) = 0$ *and* $\phi' > 0$ *on* $(0, \mu)$*, such that for every* $\mathbf{x} \in U$ *with* $f(\hat{\mathbf{x}}) < f(\mathbf{x}) < f(\hat{\mathbf{x}}) + \mu$ *it holds* $\phi'(f(\mathbf{x}) - f(\hat{\mathbf{x}}))\mathrm{dist}(\mathbf{0}, \partial f(\mathbf{x})) \geqslant 1$.

DEFINITION 2.2. *A proper function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is said to be calm at* $\mathbf{x} \in \mathrm{dom} f$ *if there exist* $\varepsilon > 0$ *and* $\kappa > 0$ *such that* $|f(\mathbf{y}) - f(\mathbf{x})| \leqslant \kappa \|\mathbf{y} - \mathbf{x}\|$ *for all* $\mathbf{y} \in B(\mathbf{x}, \varepsilon) := \{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{z} - \mathbf{x}\| < \varepsilon\}$.

LEMMA 2.3. *Let* $O \subseteq \mathbb{R}^n$ *be an open set, and* $f_1 : O \to \overline{\mathbb{R}}$ *and* $f_2 : O \to \mathbb{R}$ *be two functions which are finite at* $\mathbf{x} \in O$ *with* $f_2(\mathbf{x}) > 0$*. Suppose that* f_1 *is continuous at* $\mathbf{x}$ *relative to* $\mathrm{dom} f_1$*, that* f_2 *is calm at* $\mathbf{x}$*, and denote* $\alpha_i := f_i(\mathbf{x})$*,* $i = 1, 2$*.*
(i) Then

$$\hat{\partial}\left(\frac{f_1}{f_2}\right)(\mathbf{x}) = \frac{\hat{\partial}(\alpha_2 f_1 - \alpha_1 f_2)(\mathbf{x})}{f_2(\mathbf{x})^2}. \tag{2.4}$$

(ii) If, in addition, f_2 *is convex and* $\alpha_1 \geqslant 0$*, then*

$$\hat{\partial}\left(\frac{f_1}{f_2}\right)(\mathbf{x}) \subseteq \frac{\hat{\partial}(\alpha_2 f_1)(\mathbf{x}) - \alpha_1 \hat{\partial} f_2(\mathbf{x})}{f_2(\mathbf{x})^2}. \tag{2.5}$$

Proof. (i) The proof is similar to [35, Proposition 2.2]. (ii) If f_2 is convex and $\alpha_1 \geqslant 0$, then $\hat{\partial}(\alpha_1 f_2)(\mathbf{x}) \neq \varnothing$ thanks to $\mathbf{x} \in \mathrm{int}(\mathrm{dom} f_2)$. According to [22, eq. (1.6)], this further leads to $\hat{\partial}(\alpha_2 f_1 - \alpha_1 f_2)(\mathbf{x}) \subseteq \hat{\partial}(\alpha_2 f_1)(\mathbf{x}) - \hat{\partial}(\alpha_1 f_2)(\mathbf{x}) = \hat{\partial}(\alpha_2 f_1)(\mathbf{x}) - \alpha_1 \hat{\partial} f_2(\mathbf{x})$. ∎

Next, we present a lemma that will be useful in establishing approximate stationarity.

LEMMA 2.4. *Let* $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a proper, convex and lower semicontinuous function, and* $\mathbf{w} \in \mathrm{int}(\mathrm{dom} g)$*. Let* $\varepsilon > 0$ *and* $\mathcal{K}$ *be a compact set such that* $B(\mathbf{w}, \varepsilon) \subseteq \mathcal{K} \subseteq \mathrm{int}(\mathrm{dom} g)$ *and* g *is Lipschitz continuous on* $\mathcal{K}$ *with constant* $\kappa > 0$*. Further, let* $\mathbf{z} \in \mathbb{R}^n$ *be such that* $\mathrm{dist}(\mathbf{w}, \partial g^*(\mathbf{z})) := \inf\{\|\mathbf{w} - \boldsymbol{\eta}\| : \boldsymbol{\eta} \in \partial g^*(\mathbf{z})\} \leqslant \varepsilon$*. Then, one has* $\mathbf{z} \in \partial_{\hat{\varepsilon}} g(\mathbf{w})$*, where* $\hat{\varepsilon} := 2\kappa\varepsilon$*.*

Proof. First, as $B(\mathbf{w}, \varepsilon) \subseteq \mathcal{K}$ and g is Lipschitz continuous on $\mathcal{K}$ with constant $\kappa > 0$, we observe that $\sup\{\|\boldsymbol{\xi}\| : \boldsymbol{\xi} \in \partial g(\mathbf{w} + \mathbf{u}), \|\mathbf{u}\| \leqslant \varepsilon\} \leqslant \kappa$. For $\overline{\boldsymbol{\eta}} := \mathrm{Proj}_{\partial g^*(\mathbf{z})}(\mathbf{w})$, the projection of $\mathbf{w}$ on $\partial g^*(\mathbf{z})$ (which exists and is unique), it holds $\|\overline{\boldsymbol{\eta}} - \mathbf{w}\| \leqslant \varepsilon$. Therefore, for $\overline{\mathbf{u}} := \overline{\boldsymbol{\eta}} - \mathbf{w}$, we have $\|\overline{\mathbf{u}}\| \leqslant \varepsilon$, $\mathbf{w} + \overline{\mathbf{u}} \in \partial g^*(\mathbf{z})$ or, equivalently, $\mathbf{z} \in \partial g(\mathbf{w} + \overline{\mathbf{u}})$. Next, we claim that

$$g(\mathbf{w}) - g(\mathbf{w} + \overline{\mathbf{u}}) + \langle \overline{\mathbf{u}}, \mathbf{z} \rangle \leqslant \hat{\varepsilon}, \tag{2.6}$$

where $\hat{\varepsilon}$ is defined in the statement of the lemma. Since $\langle \mathbf{w} + \overline{\mathbf{u}}, \mathbf{z} \rangle - g^*(\mathbf{z}) = g(\mathbf{w} + \overline{\mathbf{u}})$, thanks to the fact $\mathbf{z} \in \partial g(\mathbf{w} + \overline{\mathbf{u}})$, this is equivalent to $g(\mathbf{w}) + g^*(\mathbf{z}) \leqslant \hat{\varepsilon} + \langle \mathbf{w}, \mathbf{z} \rangle$, and so, the conclusion follows. Now we will prove that (2.6) is true. By direct calculations, we have $g(\mathbf{w}) - g(\mathbf{w} + \overline{\mathbf{u}}) + \langle \overline{\mathbf{u}}, \mathbf{z} \rangle \leqslant \kappa \|\overline{\mathbf{u}}\| + \kappa \|\overline{\mathbf{u}}\| \leqslant 2\kappa\varepsilon = \hat{\varepsilon}$. ∎

3. Basic assumptions and stationary points of fractional programs.

We introduce the basic assumptions and present notions of stationary points.

ASSUMPTION 3.1. *Throughout this paper, we assume that*
(a) $\mathcal{S} \subseteq \mathbb{R}^n$ *is a nonempty convex and compact set;*

(b) g is a proper, convex and lower semicontinuous function;
(c) h is differentiable with Lipschitz continuous gradient over an open set containing the compact set $\mathcal{S}$ with a Lipschitz constant $L_{\nabla h}$;
(d) f is a proper, convex and lower semicontinuous function with $K(\mathcal{S}) \subseteq \mathrm{int}(\mathrm{dom} f)$ and $f(K\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{S}$;
(e) $\mathcal{S} \cap A^{-1}(\mathrm{dom} g) \neq \varnothing$ and $\alpha := \inf_{\mathbf{x}\in\mathcal{S}}\{g(A\mathbf{x}) + h(\mathbf{x})\} > 0$;
(f) It holds that $A(\mathcal{S}) \subseteq \mathrm{dom}(\partial g)$ and there exists a constant $\ell > 0$ such that $\mathrm{dist}(0, \partial g(A\mathbf{x})) \leqslant \ell$ for all $\mathbf{x} \in \mathcal{S}$.

The assumption $\mathcal{S} \cap A^{-1}(\mathrm{dom} g) \neq \varnothing$ ensures that the objective function F is not identically $+\infty$. The second condition in Assumption 3.1(e) can be satisfied by augmenting the objective with a suitable positive constant[1], noting from Assumption 3.1(a)-(d) that $\inf_{\mathbf{x}\in\mathcal{S}} F(\mathbf{x}) > -\infty$. Assumption 3.1(f) is automatically satisfied when the compact set $A(\mathcal{S})$ is a subset of the interior of $\mathrm{dom}\, g$.

Remark 3.2. In the illustrative examples of (a) and (b) provided in Section 1, the functions f and g have a full domain — therefore Assumption 3.1(a)-(f) are fulfilled. In example (a), it holds $\alpha := \inf_{\mathbf{x}\in\mathcal{S}}\{g(A\mathbf{x})+h(\mathbf{x})\} > 0$ owing the assumption $\mathcal{B} \cap \mathrm{span}(\mathbf{E}) = \varnothing$. In example (b), it holds $\alpha := \inf_{\mathbf{x}\in\mathcal{S}}\{g(A\mathbf{x}) + h(\mathbf{x})\} \geqslant 0$, however, one could then augment the objective by adding a positive constant in order to make the inequality strict.

DEFINITION 3.3. *For the optimization problem (1.1), we say that $\overline{\mathbf{x}} \in \mathbb{R}^n$ is*
(i) a Fréchet stationary point if $0 \in \hat{\partial}\left(\frac{g\circ A+h+\iota_{\mathcal{S}}}{f\circ K}\right)(\overline{\mathbf{x}})$;
(ii) a limiting lifted stationary point if

$$0 \in \left(A^*\partial g(A\overline{\mathbf{x}}) + \nabla h(\overline{\mathbf{x}}) + \partial\iota_{\mathcal{S}}(\overline{\mathbf{x}})\right) f(K\overline{\mathbf{x}}) - (g(A\overline{\mathbf{x}}) + h(\overline{\mathbf{x}}))K^*\partial f(K\overline{\mathbf{x}}).$$

Any local minimizer $\overline{\mathbf{x}} \in \mathbb{R}^n$ of (1.1) is a Fréchet stationary point. If $\overline{\mathbf{x}} \in \mathbb{R}^n$ is a Fréchet stationary point of (1.1) such that $K\overline{\mathbf{x}} \in \mathrm{int}(\mathrm{dom} f)$, and either $\overline{\mathbf{x}} \in \mathrm{ri}(\mathcal{S}) \cap A^{-1}\mathrm{ri}(\mathrm{dom} g)$ or $\mathcal{S}$ is polyhedral and $\overline{\mathbf{x}} \in \mathcal{S} \cap A^{-1}\mathrm{ri}(\mathrm{dom} g)$, then, according to Lemma 2.3, $\overline{\mathbf{x}} \in \mathbb{R}^n$ is also a limiting lifted stationary point of (1.1). Example 3.1 in [8] also illustrates that a limiting lifted stationary point may not be a Fréchet stationary point.

Next, we introduce the notion of an approximate lifted stationary point for problem (1.1).

DEFINITION 3.4. *Given $\epsilon_1, \epsilon_2 \geqslant 0$, we say that $\overline{\mathbf{x}} \in \mathbb{R}^n$ is a limiting (ϵ_1, ϵ_2)-lifted stationary point of the problem (1.1) if there exists $\overline{\Psi} \in \mathbb{R}$ with $|\overline{\Psi} - (g(A\overline{\mathbf{x}}) + h(\overline{\mathbf{x}}))| \leqslant \epsilon_2$ such that $0 \in \left(A^*\partial_{\epsilon_1} g(A\overline{\mathbf{x}}) + \nabla h(\overline{\mathbf{x}}) + \partial\iota_{\mathcal{S}}(\overline{\mathbf{x}})\right) f(K\overline{\mathbf{x}}) - \overline{\Psi}\, K^*\partial f(K\overline{\mathbf{x}})$.*

If $\epsilon_1 = \epsilon_2 = 0$, then this notion reduces to the limiting lifted stationary point. Below, we provide a lemma stating that there are positive uniform lower/upper bounds on the denominator values of (1.1) under Assumption 3.1. The proof is omitted due to its simplicity.

LEMMA 3.5. *Suppose Assumption 3.1 holds. Then, there exist two positive scalars m and M such that $m < f(K\mathbf{x}) \leqslant M$ for all $\mathbf{x} \in \mathcal{S}$.*

4. Full splitting proximal subgradient algorithm.

We first propose a *conceptual* algorithmic framework for solving (1.1) which we call *full splitting proximal subgradient algorithm with an extrapolated step* (FSPS).

[1] Different choices of the constant α may affect numerical performance.

Let $0 < \beta < 2$, the sequences of scalars $\{\gamma_k\}$ and $\{\delta_k\}$ such that $\gamma_k \geqslant 0$ and $\delta_k > 0$ for all $k \geqslant 0$, $\theta_0 > 0$ and a given starting point $(\mathbf{x}^0, \mathbf{z}^0, \mathbf{u}^0)$ with $\mathbf{x}^0 \in \mathcal{S}$. For all $k \geqslant 0$, we consider the following update rule:

$$
\left\{\begin{array}{lcl}
\mathbf{y}^{k+1} & \in & \partial f(K\mathbf{x}^k) \\
\mathbf{x}^{k+1} & := & \mathrm{Proj}_{\mathcal{S}}\left(\mathbf{u}^k + \frac{\theta_k}{\delta_k}K^*\mathbf{y}^{k+1} - \frac{1}{\delta_k}\nabla h(\mathbf{x}^k) - \frac{1}{\delta_k}A^*\mathbf{z}^k\right), \\
\mathbf{u}^{k+1} & := & (1-\beta)\mathbf{u}^k + \beta\mathbf{x}^{k+1}, \\
\mathbf{z}^{k+1} & := & \arg\min_{\mathbf{z}}\left[g^*(\mathbf{z}) - \langle A\mathbf{x}^{k+1}, \mathbf{z}\rangle + \frac{\gamma_k}{2}\|\mathbf{z}\|^2\right], \\
\theta_{k+1} & := & \dfrac{\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}; \delta_k, \gamma_k)}{f(K\mathbf{x}^{k+1})},
\end{array}\right. \tag{4.1}
$$

where $\Psi(\mathbf{x}, \mathbf{z}, \mathbf{u}, \delta, \gamma) := \langle \mathbf{z}, A\mathbf{x}\rangle - g^*(\mathbf{z}) + h(\mathbf{x}) + \iota_{\mathcal{S}}(\mathbf{x}) + \frac{\delta}{2}\|\mathbf{x}-\mathbf{u}\|^2 - \frac{\gamma}{2}\|\mathbf{z}\|^2$.

4.1. Smoothing-based FSPS algorithm. In this section, we consider a variant of FSPS, which we refer to as the S-FSPS algorithm. This algorithm assumes that $\beta = 1$ in (4.1), and therefore uses

$$\Psi(\mathbf{x}, \mathbf{z}; \gamma) := \langle \mathbf{z}, A\mathbf{x}\rangle - g^*(\mathbf{z}) + h(\mathbf{x}) + \iota_{\mathcal{S}}(\mathbf{x}) - \frac{\gamma}{2}\|\mathbf{z}\|^2.$$

ALGORITHM 4.1 (S-FSPS algorithm). *Let $\{\gamma_k\}$ be a positive and nonincreasing sequence with $\lim_{k\to+\infty}\gamma_k = 0$ and $\sum_{k\geqslant 0}\gamma_k = +\infty$, $\chi > 1$, and $\delta_k = \chi\left(L_{\nabla h} + \frac{\sigma_A^2}{\gamma_k}\right)$ for all $k \geqslant 0$, $\theta_0 > 0$, and a given starting point $(\mathbf{x}^0, \mathbf{z}^0)$. For all $k \geqslant 0$, we consider the following update rule:*

Choose $\mathbf{y}^{k+1} \in \partial f(K\mathbf{x}^k)$.

Update $\mathbf{x}^{k+1} := \mathrm{Proj}_{\mathcal{S}}\left(\mathbf{x}^k + \dfrac{\theta_k}{\delta_k}K^*\mathbf{y}^{k+1} - \dfrac{1}{\delta_k}\nabla h(\mathbf{x}^k) - \dfrac{1}{\delta_k}A^*\mathbf{z}^k\right)$.

Update $\mathbf{z}^{k+1} := \mathrm{Prox}_{g^*/\gamma_k}\left(\dfrac{A\mathbf{x}^{k+1}}{\gamma_k}\right)$.

Update $\theta_{k+1} := \dfrac{\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}; \gamma_k)}{f(K\mathbf{x}^{k+1})}$.

Remark 4.2. For the sequence $\{(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)\}$ generated by Algorithm 4.1, we have for all $k \geqslant 0$

$$
\begin{aligned}
\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}; \gamma_k) &= \langle \mathbf{z}^{k+1}, A\mathbf{x}^{k+1}\rangle - g^*(\mathbf{z}^{k+1}) - \frac{\gamma_k}{2}\|\mathbf{z}^{k+1}\|^2 + h(\mathbf{x}^{k+1}) + \iota_{\mathcal{S}}(\mathbf{x}^{k+1}) \\
&= g_{\gamma_k}(A\mathbf{x}^{k+1}) + h(\mathbf{x}^{k+1}) + \iota_{\mathcal{S}}(\mathbf{x}^{k+1}),
\end{aligned} \tag{4.2}
$$

where the last equality is due to (2.3) and $\mathbf{z}^{k+1} = \mathrm{Prox}_{g^*/\gamma_k}\left(\frac{A\mathbf{x}^{k+1}}{\gamma_k}\right)$.

In addition, the update of $\mathbf{x}^{k+1}$ can be equivalently written as, for all $k \geqslant 0$,

$$\mathbf{x}^{k+1} = \mathrm{Proj}_{\mathcal{S}}\left(\mathbf{x}^k + \frac{\theta_k}{\delta_k}K^*\mathbf{y}^{k+1} - \frac{1}{\delta_k}\nabla h(\mathbf{x}^k) - \frac{1}{\delta_k}\nabla\left(g_{\gamma_{k-1}} \circ A\right)(\mathbf{x}^k)\right).$$

This shows that Algorithm 4.1 can be reformulated using the Moreau envelope of g. As a result, one can interpret it as a smoothing-based proximal-subgradient method.

4.2. Convergence analysis of S-FSPS. We provide a convergence analysis for Algorithm 4.1 and denote for simplicity $\boldsymbol{V}^k := (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)$ for all $k \geqslant 0$.

THEOREM 4.3. *Suppose Assumption 3.1 holds. Let Ω be the set of the accumulation points of the sequence $\{\boldsymbol{V}^k\}$ generated by Algorithm 4.1. Then, the following statements hold:*

(i) For all $k \geqslant 1$ it holds

$$\begin{aligned} &\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}; \gamma_k) + \theta_k \left[f(K\mathbf{x}^k) - \left(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle - f^*(\mathbf{y}^{k+1}) \right) \right] \\ &\leqslant \Psi(\mathbf{x}^k, \mathbf{z}^k; \gamma_{k-1}) - c_k \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \Xi^{k+1}, \end{aligned} \tag{4.3}$$

where

$$\Xi^{k+1} := \frac{\gamma_{k-1} - \gamma_k}{2} \|\mathbf{z}^{k+1}\|^2 \geqslant 0 \quad \textit{and} \quad c_k := \frac{(\chi - 1)}{2} \left(L_{\nabla h} + \frac{\sigma_A^2}{\gamma_k} \right) > 0.$$

(ii) The sequence $\{\boldsymbol{V}^k\}$ is bounded.
(iii) There exists an index $K_1 \geqslant 1$ such that $\theta_k \geqslant 0$ for all $k \geqslant K_1$.
(iv) $\lim_{k\to+\infty} \theta_k = \overline{\theta}$ for some $\overline{\theta} \geqslant 0$.
(v) It holds that $\liminf_{k\to+\infty} \delta_k \|\mathbf{x}^{k+1} - \mathbf{x}^k\| = 0$.

Below, we further assume that $A(\mathcal{S}) \subseteq \operatorname{int}(\operatorname{dom} g)$. [2]

(vi) For every $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}) \in \Omega$, it holds that $\frac{g(A\overline{\mathbf{x}}) + h(\overline{\mathbf{x}}) + \iota_{\mathcal{S}}(\overline{\mathbf{x}})}{f(K\overline{\mathbf{x}})} = \overline{\theta}$, where $\overline{\theta}$ is given as in (iv).
(vii) Let $\{\mathbf{x}^{k_j}\}$ be a subsequence of $\mathbf{x}^k$ such that $\lim_{j\to+\infty} \delta_{k_j} \|\mathbf{x}^{k_j+1} - \mathbf{x}^{k_j}\| = 0$ (whose existence is guaranteed by (vi)). Then, any accumulation point $\overline{\mathbf{x}}$ of it is a limiting lifted stationary point for the optimization problem (1.1).

Proof. (i) Let $k \geqslant 1$. According to the properties of the projection, the $\mathbf{x}$-update in Algorithm 4.1 gives us that

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\mathcal{S}} \left[\langle \mathbf{z}^k, A\mathbf{x} \rangle - \theta_k \langle K\mathbf{x}, \mathbf{y}^{k+1} \rangle + \langle \nabla h(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k \rangle + \frac{\delta_k}{2} \|\mathbf{x} - \mathbf{x}^k\|^2 \right].$$

The objective function of the above optimization problem is strongly convex with modulus δ_k, therefore,

$$\begin{aligned} &\langle A\mathbf{x}^{k+1}, \mathbf{z}^k \rangle + \langle \mathbf{x}^{k+1} - \mathbf{x}^k, \nabla h(\mathbf{x}^k) \rangle - \theta_k \langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle \\ &\leqslant \langle A\mathbf{x}^k, \mathbf{z}^k \rangle - \theta_k \langle K\mathbf{x}^k, \mathbf{y}^{k+1} \rangle - \frac{\delta_k}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2. \end{aligned}$$

Combined this with $\langle K(\mathbf{x}^k - \mathbf{x}^{k+1}), \mathbf{y}^{k+1} \rangle = f(K\mathbf{x}^k) - \left(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle - f^*(\mathbf{y}^{k+1}) \right)$, leads to

$$\begin{aligned} &\langle A\mathbf{x}^{k+1}, \mathbf{z}^k \rangle + \langle \mathbf{x}^{k+1} - \mathbf{x}^k, \nabla h(\mathbf{x}^k) \rangle + \theta_k \left[f(K\mathbf{x}^k) - (\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1} \rangle - f^*(\mathbf{y}^{k+1})) \right] \\ &\leqslant \langle A\mathbf{x}^k, \mathbf{z}^k \rangle - \frac{\delta_k}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2. \end{aligned} \tag{4.4}$$

Since ∇h is Lipschitz continuous with constant $L_{\nabla h}$, it holds that

$$h(\mathbf{x}^{k+1}) - h(\mathbf{x}^k) \leqslant \langle \mathbf{x}^{k+1} - \mathbf{x}^k, \nabla h(\mathbf{x}^k) \rangle + \frac{L_{\nabla h}}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \tag{4.5}$$

[2]We note that the conditions $A(\mathcal{S}) \subseteq \operatorname{int}(\operatorname{dom} g)$ is satisfied with our motivation examples. Also, it ensures that Assumption 3.1(f) holds.

Combining (4.5) with (4.4), we obtain

$$\Psi(\mathbf{x}^{k+1},\mathbf{z}^k;\gamma_{k-1})+\theta_k\left[f(K\mathbf{x}^k)-(\langle K\mathbf{x}^{k+1},\mathbf{y}^{k+1}\rangle-f^*(\mathbf{y}^{k+1}))\right]$$

$$\leqslant \Psi(\mathbf{x}^k,\mathbf{z}^k;\gamma_{k-1})-\frac{1}{2}(\delta_k-L_{\nabla h})\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2. \tag{4.6}$$

From the $\mathbf{z}$-update in (4.1) it follows that $A\mathbf{x}^k-\gamma_{k-1}\mathbf{z}^k\in\partial g^*(\mathbf{z}^k)$, therefore

$$-g^*(\mathbf{z}^{k+1})\leqslant -g^*(\mathbf{z}^k)-\langle A\mathbf{x}^k-\gamma_{k-1}\mathbf{z}^k,\mathbf{z}^{k+1}-\mathbf{z}^k\rangle.$$

Combining this inequality with the identity

$$-\frac{\gamma_{k-1}}{2}\|\mathbf{z}^{k+1}\|^2=-\frac{\gamma_{k-1}}{2}\|\mathbf{z}^k\|^2-\gamma_{k-1}\langle\mathbf{z}^{k+1}-\mathbf{z}^k,\mathbf{z}^k\rangle-\frac{\gamma_{k-1}}{2}\|\mathbf{z}^{k+1}-\mathbf{z}^k\|^2,$$

it yields

$$\begin{aligned}&\langle A\mathbf{x}^{k+1},\mathbf{z}^{k+1}\rangle-g^*(\mathbf{z}^{k+1})-\frac{\gamma_{k-1}}{2}\|\mathbf{z}^{k+1}\|^2\\ \leqslant&\langle A\mathbf{x}^{k+1},\mathbf{z}^k\rangle-g^*(\mathbf{z}^k)-\frac{\gamma_{k-1}}{2}\|\mathbf{z}^k\|^2+\langle A\mathbf{x}^{k+1}-A\mathbf{x}^k,\mathbf{z}^{k+1}-\mathbf{z}^k\rangle-\frac{\gamma_{k-1}}{2}\|\mathbf{z}^k-\mathbf{z}^{k+1}\|^2\\ \leqslant&\langle A\mathbf{x}^{k+1},\mathbf{z}^k\rangle-g^*(\mathbf{z}^k)-\frac{\gamma_{k-1}}{2}\|\mathbf{z}^k\|^2+\frac{\sigma_A^2}{2\gamma_k}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2,\end{aligned} \tag{4.7}$$

where the last estimate follows from the Cauchy-Schwarz inequality and $\gamma_k\leqslant\gamma_{k-1}$. Therefore,

$$\Psi(\mathbf{x}^{k+1},\mathbf{z}^{k+1};\gamma_k)\leqslant\Psi(\mathbf{x}^{k+1},\mathbf{z}^k;\gamma_{k-1})+\frac{\sigma_A^2}{2\gamma_k}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2+\frac{\gamma_{k-1}-\gamma_k}{2}\|\mathbf{z}^{k+1}\|^2.$$

By combining the above inequality with (4.6) and recalling the choice of δ_k, assertion (i) follows.

[(ii) & (iii)] Since $\mathcal{S}$ is a compact set, the sequence $\{\mathbf{x}^k\}\subseteq\mathcal{S}$ is bounded. The sequence $\{\mathbf{y}^k\}$ is bounded due to Assumption 3.1(d) and the local boundedness property of the convex subdifferential. Let $k\geqslant 1$. According to Assumption 3.1(f), there exists $\tilde{\mathbf{z}}^k\in\partial g(A\mathbf{x}^{k+1})$ with $\|\tilde{\mathbf{z}}^k\|\leqslant\ell+1$. Invoking the definition of $\mathbf{z}^{k+1}$, we have

$$\begin{aligned}&g^*(\mathbf{z}^{k+1})-\langle A\mathbf{x}^{k+1},\mathbf{z}^{k+1}\rangle+\frac{\gamma_k}{2}\|\mathbf{z}^{k+1}\|^2\\ \leqslant\ &g^*(\tilde{\mathbf{z}}^k)-\langle A\mathbf{x}^{k+1},\tilde{\mathbf{z}}^k\rangle+\frac{\gamma_k}{2}\|\tilde{\mathbf{z}}^k\|^2=-g(A\mathbf{x}^{k+1})+\frac{\gamma_k}{2}\|\tilde{\mathbf{z}}^k\|^2.\end{aligned}$$

In particular, we see that $\|\mathbf{z}^{k+1}\|\leqslant\|\tilde{\mathbf{z}}^k\|\leqslant\ell+1$ for all $k\geqslant 0$. So, the sequence $\{\mathbf{z}^k\}$ is also bounded, and hence, (ii) follows. Now, according to Assumption 3.1(e),

$$\langle A\mathbf{x}^{k+1},\mathbf{z}^{k+1}\rangle-g^*(\mathbf{z}^{k+1})+h(\mathbf{x}^{k+1})-\frac{\gamma_k}{2}\|\mathbf{z}^{k+1}\|^2\geqslant\underbrace{g(A\mathbf{x}^{k+1})+h(\mathbf{x}^{k+1})}_{\geqslant\alpha}-\frac{\gamma_k}{2}\|\tilde{\mathbf{z}}^k\|^2.$$

Since $\lim_{k\to+\infty}\gamma_k=0$, there exists an index $K_1\geqslant 1$ such that, for all $k\geqslant K_1$, we have $\frac{\gamma_k}{2}\|\tilde{\mathbf{z}}^k\|\leqslant\frac{\alpha}{2}$. Therefore, (iii) holds by combining the above inequality with Assumption 3.1(d) and (e).

(iv) Invoking (4.3) and $\theta_k f(K\mathbf{x}^k)=\Psi(\mathbf{x}^k,\mathbf{z}^k,\gamma_{k-1})$, for all $k\geqslant K_1$ we have

$$\Psi(\mathbf{x}^{k+1},\mathbf{z}^{k+1};\gamma_k)\leqslant\theta_k\left(\langle K\mathbf{x}^{k+1},\mathbf{y}^{k+1}\rangle-f^*(\mathbf{y}^{k+1})\right)-c_k\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2+\Xi^{k+1}$$

$$\leqslant \theta_k f(K\mathbf{x}^{k+1}) - \frac{(\chi-1)\delta_k}{2\chi}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \Xi^{k+1}.$$

From here it follows that for all $k \geqslant K_1$

$$\theta_{k+1} \leqslant \theta_k - \frac{(\chi-1)\delta_k}{2\chi M}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \frac{\Xi^{k+1}}{M}, \tag{4.8}$$

where $M > 0$ is given by Lemma 3.5. The boundedness of $\{\mathbf{z}^k\}$ guarantees the summability of $\{\Xi^k\}$, therefore, from [4, Lemma 5.31] it yields $\lim_{k\to+\infty} \theta_k := \overline{\theta} \geqslant 0$ for some $\overline{\theta} \geqslant 0$, and $\sum_{k=0}^{+\infty} \delta_k \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 < +\infty$.

(v) In the proof of statement (iv), we have seen that $\sum_{k=0}^{+\infty} \frac{1}{\delta_k}(\delta_k\|\mathbf{x}^k - \mathbf{x}^{k+1}\|)^2 < +\infty$. On the other hand, $\sum_{k=0}^{+\infty} \frac{1}{\delta_k} = +\infty$, thus $\liminf\limits_{k\to+\infty} \delta_k \|\mathbf{x}^{k+1} - \mathbf{x}^k\| = 0$.

(vi) Let $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}) \in \Omega$, $\{(\mathbf{x}^{k_j}, \mathbf{y}^{k_j}, \mathbf{z}^{k_j})\}$ be a subsequence of $\{(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)\}$ such that $(\mathbf{x}^{k_j}, \mathbf{y}^{k_j}, \mathbf{z}^{k_j}) \to (\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}})$ as $j \to +\infty$. Clearly, $\{\mathbf{x}^{k_j}\} \subseteq \mathcal{S}$ and $\overline{\mathbf{x}} \in \mathcal{S}$.

For convenience, we denote $\lambda_j := \gamma_{k_j-1} \to 0$ as $j \to +\infty$, and write $[A\mathbf{x}^{k_j}, A\overline{\mathbf{x}}] = \{tA\mathbf{x}^{k_j} + (1-t)A\overline{\mathbf{x}} : t \in [0,1]\}$. We claim that, for all $\mathbf{w}^{k_j} \in [A\mathbf{x}^{k_j}, A\overline{\mathbf{x}}] \subseteq A(\mathcal{S})$, one has $\operatorname{Prox}_{\lambda_j g}(\mathbf{w}^{k_j}) \to A\overline{\mathbf{x}}$ as $j \to +\infty$. To see this, we observe from our assumption that $\mathbf{w}^{k_j} \in A(\mathcal{S}) \subseteq \operatorname{int}(\operatorname{dom} g)$, $A\overline{\mathbf{x}} \in A(\mathcal{S}) \subseteq \operatorname{int}(\operatorname{dom} g)$ and

$$\begin{aligned}
\|\operatorname{Prox}_{\lambda_j g}(\mathbf{w}^{k_j}) - A\overline{\mathbf{x}}\| &\leqslant \|\operatorname{Prox}_{\lambda_j g}(\mathbf{w}^{k_j}) - \operatorname{Prox}_{\lambda_j g}(A\overline{\mathbf{x}})\| + \|\operatorname{Prox}_{\lambda_j g}(A\overline{\mathbf{x}}) - A\overline{\mathbf{x}}\| \\
&\leqslant \|\mathbf{w}^{k_j} - A\overline{\mathbf{x}}\| + \|\operatorname{Prox}_{\lambda_j g}(A\overline{\mathbf{x}}) - A\overline{\mathbf{x}}\| \\
&\leqslant \|A\mathbf{x}^{k_j} - A\overline{\mathbf{x}}\| + \|\operatorname{Prox}_{\lambda_j g}(A\overline{\mathbf{x}}) - A\overline{\mathbf{x}}\|,
\end{aligned}$$

where the second inequality follows from the non-expansiveness of the proximal operator of convex functions. Then, the claim follows by noting that, as $j \to +\infty$, $\mathbf{x}^{k_j} \to \overline{\mathbf{x}}$ and $\operatorname{Prox}_{\lambda_j g}(A\overline{\mathbf{x}}) \to A\overline{\mathbf{x}}$ (thanks to [25, Proposition 2.2] and $A\overline{\mathbf{x}} \in \operatorname{int}(\operatorname{dom} g)$).

From the claim and the assumption $A(\mathcal{S}) \subseteq \operatorname{int}(\operatorname{dom} g)$, it follows that there exist an index j_0 and a bounded set U with $A(\mathcal{S}) \subseteq \operatorname{cl}(U) \subseteq \operatorname{int}(\operatorname{dom} g)$ such that

$$\operatorname{Prox}_{\lambda_j g}(\mathbf{w}^{k_j}) \in U \quad \text{for all } j \geqslant j_0 \text{ and for all } \mathbf{w}^{k_j} \in [A\mathbf{x}^{k_j}, A\overline{\mathbf{x}}].$$

Note that the function g is Lipschitz continuous (with some Lipschitz constant $L_g > 0$) on $\operatorname{cl}(U)$. It follows that for all $j \geqslant j_0$ and for all $\mathbf{w}^{k_j} \in [A\mathbf{x}^{k_j}, A\overline{\mathbf{x}}]$, $\sup\{\|u\| : u \in \partial g(\operatorname{Prox}_{\lambda_j g}(\mathbf{w}^{k_j}))\|\} \leqslant L_g$. As $\nabla g_{\lambda_j}(\mathbf{w}^{k_j}) \in \partial g\big(\operatorname{Prox}_{\lambda_j g}(\mathbf{w}^{k_j})\big)$, we further deduce that $\|\nabla g_{\lambda_j}(\mathbf{w}^{k_j})\| \leqslant L_g$ for all $j \geqslant j_0$ and for all $\mathbf{w}^{k_j} \in [A\mathbf{x}^{k_j}, A\overline{\mathbf{x}}]$. This, together with the mean value theorem, implies that, for all $j \geqslant j_0$,

$$\begin{aligned}
|g_{\lambda_j}(A\mathbf{x}^{k_j}) - g(A\overline{\mathbf{x}})| &\leqslant |g_{\lambda_j}(A\mathbf{x}^{k_j}) - g_{\lambda_j}(A\overline{\mathbf{x}})| + |g_{\lambda_j}(A\overline{\mathbf{x}}) - g(A\overline{\mathbf{x}})| \\
&\leqslant L_g\|A\mathbf{x}^{k_j} - A\overline{\mathbf{x}}\| + |g_{\lambda_j}(A\overline{\mathbf{x}}) - g(A\overline{\mathbf{x}})|.
\end{aligned}$$

Since $g_{\lambda_j}(A\overline{\mathbf{x}}) \to g(A\overline{\mathbf{x}})$ (see [4, Proposition 12.33(ii)]), it implies that

$$g_{\gamma_{k_j-1}}(A\mathbf{x}^{k_j}) = g_{\lambda_j}(A\mathbf{x}^{k_j}) \to g(A\overline{\mathbf{x}}) \text{ as } j \to +\infty. \tag{4.9}$$ [3]

Now, recall that $\theta_{k_j} = \frac{\Psi(\mathbf{x}^{k_j}, \mathbf{z}^{k_j}; \gamma_{k_j-1})}{f(K\mathbf{x}_{k_j})}$. Letting $j \to \infty$, and using (4.2), (4.9) and Assumption 3.1(d), assertion (vi) follows.

[3] In the case where g is a finite-valued convex function, the assertion (4.9) follows directly by [1, Proposition 1(d)] . Here, we establish this under the weaker assumption that $A(\mathcal{S}) \subseteq \operatorname{int}(\operatorname{dom} g)$. We also note that, if g is a proper lower semicontinuous (possibly) nonconvex function with the additional assumption that $\inf g > -\infty$, then this also follows from [16, Lemma 1].

(vii) From the $\mathbf{x}$-update in (4.1) we also have for all $k \geqslant 1$

$$\begin{aligned} 0 &\in \partial\iota_{\mathcal{S}}(\mathbf{x}^{k+1}) + \mathbf{x}^{k+1} - \mathbf{x}^k + \frac{A^*\mathbf{z}^k + \nabla h(\mathbf{x}^k) - \theta_k K^*\mathbf{y}^{k+1}}{\delta_k} \\ &= \partial\iota_{\mathcal{S}}(\mathbf{x}^{k+1}) + \mathbf{x}^{k+1} - \mathbf{x}^k + \frac{\nabla(g_{\gamma_{k-1}} \circ A)(\mathbf{x}^k) + \nabla h(\mathbf{x}^k) - \theta_k K^*\mathbf{y}^{k+1}}{\delta_k}. \end{aligned} \tag{4.10}$$

Let $\{\mathbf{x}^{k_j}\}$ be a subsequence of $\{\mathbf{x}^k\}$ such that $\lim_{j\to+\infty} \delta_{k_j}\|\mathbf{x}^{k_j+1} - \mathbf{x}^{k_j}\| = 0$ and let $\overline{\mathbf{x}} \in \mathcal{S}$ be an accumulation point of it. Then there exists a further subsequence $\{\mathbf{x}^{k_s}\}$ of $\{\mathbf{x}^{k_j}\}$ that converges to $\overline{\mathbf{x}}$ as $s \to +\infty$. Since $\delta_k \geqslant \chi L_{\nabla h}$ for all $k \geqslant 0$, we have $\lim_{s\to+\infty} \|\mathbf{x}^{k_s+1} - \mathbf{x}^{k_s}\| = 0$, thus $\mathbf{x}^{k_s+1} \to \overline{\mathbf{x}}$ as $s \to +\infty$. By passing to a further subsequence if necessary, without loss of generality, we assume that $\mathbf{y}^{k_s+1} \to \overline{\mathbf{y}}$ as $s \to +\infty$, for some $\overline{\mathbf{y}}$. From (4.10) and $\nabla(g_{\gamma_{k_s-1}} \circ A)(\mathbf{x}^{k_s}) = A^*\nabla g_{\gamma_{k_s-1}}(A\mathbf{x}^{k_s})$, for all $s \geqslant 0$, there exists $\xi^{k_s+1} \in \partial\iota_{\mathcal{S}}(\mathbf{x}^{k_s+1})$ such that

$$0 = \xi^{k_s+1} + \delta_{k_s}(\mathbf{x}^{k_s+1} - \mathbf{x}^{k_s}) + A^*\nabla g_{\gamma_{k_s-1}}(A\mathbf{x}^{k_s}) + \nabla h(\mathbf{x}^{k_s}) - \theta_{k_s} K^*\mathbf{y}^{k_s+1}. \tag{4.11}$$

Next, we see that $\{\xi^{k_s+1}\}$ is bounded. To see this, using a similar proof as in (vi) and the assumption $A(\mathcal{S}) \subseteq \operatorname{int}(\operatorname{dom} g)$, one can deduce that there exist an index K and $L_g > 0$ such that $\|\nabla g_{\gamma_{k_s-1}}(A\mathbf{x}^{k_s})\| \leqslant L_g$ for all $s \geqslant K$. So, $\|A^*\nabla g_{\gamma_{k_s-1}}(A\mathbf{x}^{k_s})\| \leqslant \sigma_A L_g$ for all $s \geqslant K$. This together with (4.11) implies that the sequence $\{\xi^{k_s+1}\}$ is bounded.

Now, from the boundness of $\{\xi^{k_s+1}\}$ and $\{A^*\nabla g_{\gamma_{k_s-1}}(A\mathbf{x}^{k_s})\}$, by further passing to subsequence, we can assume that $\xi^{k_s+1} \to \overline{\xi}$ and $A^*\nabla g_{\gamma_{k_s-1}}(A\mathbf{x}^{k_s}) \to \overline{a}$ for some $\overline{\xi}$ and $\overline{a}$. Using $\xi^{k_s+1} \in \partial\iota_{\mathcal{S}}(\mathbf{x}^{k_s+1})$, $\nabla g_{\gamma_{k_s-1}}(A\mathbf{x}^{k_s}) \subseteq \partial g\big(\operatorname{Prox}_{(\gamma_{k_s-1})g}(A\mathbf{x}^{k_s})\big)$, $\operatorname{Prox}_{(\gamma_{k_s-1})g}(A\mathbf{x}^{k_s}) \to A\overline{\mathbf{x}}$ and the outer semicontinuity (OSC) of subdifferentials [27, Proposition 8.7], we have $\overline{\xi} \in \partial\iota_{\mathcal{S}}(\overline{\mathbf{x}})$ and $\overline{a} \in A^*\partial g(A\overline{\mathbf{x}})$. By passing to the limit in (4.11), and noting that $\nabla h(\mathbf{x}^{k_s}) \to \nabla h(\overline{\mathbf{x}})$ and $\mathbf{y}^{k_s+1} \to \overline{\mathbf{y}}$ as $s \to +\infty$, we conclude that $0 \in \partial\iota_{\mathcal{S}}(\overline{\mathbf{x}}) + A^*\partial g(A\overline{\mathbf{x}}) + \nabla h(\overline{\mathbf{x}}) - \overline{\theta} K^*\overline{\mathbf{y}}$, as desired. ∎

In the previous theorem, we have derived a subsequential convergence for the S-FSPS algorithm in the sense that there exists a subsequence whose cluster point is a lifted stationary point of the problem. On the other hand, there is no guarantee of the convergence of the full sequence. Indeed, to the best of our knowledge, obtaining convergence of the full sequence generated by smoothing-based algorithms is non-trivial in general. It has been recently derived for some special structured non-fractional optimization problems involving cardinality functions (see, for example, [5]).

This motivates us to develop an alternative algorithm in the next section, which enjoys global convergence guarantees under some commonly used and mild assumptions, such as the KL property.

5. Adaptive FSPS algorithm.

We present an adaptive version of FSPS, called the Adaptive FSPS algorithm, which determines the parameter sequences $\{\gamma_k\}$ and $\{\delta_k\}$ in a self-adapting manner and ensures the positivity of the sequence $\{\theta_k\}$.

ALGORITHM 5.1 (Adaptive FSPS algorithm). *Let $0 < \beta < 2$, $\chi > 1$, $0 < q < 1$, $\delta_0, \theta_0 > 0$, $\gamma_0 = 1$ and $\varepsilon > 0$, and given a starting point $(\mathbf{x}^0, \mathbf{z}^0, \mathbf{u}^0)$. For all $k \geqslant 0$, consider the following update rule:*

Choose $\mathbf{y}^{k+1} \in \partial f(K\mathbf{x}^k)$.

Update $\mathbf{x}^{k+1} := \operatorname{Proj}_{\mathcal{S}}\left(\mathbf{u}^k + \frac{\theta_k}{\delta_k}K^*\mathbf{y}^{k+1} - \frac{1}{\delta_k}\nabla h(\mathbf{x}^k) - \frac{1}{\delta_k}A^*\mathbf{z}^k\right)$.

Update $\mathbf{u}^{k+1} := (1-\beta)\mathbf{u}^k + \beta\mathbf{x}^{k+1}$.
Set $\gamma_{k,0} := \gamma_k$.
Find the smallest $j_k \in \{0,1,2,\ldots\}$ *such that for* $\gamma_{k,j_k} := \gamma_{k,0}q^{j_k}$ *and*

$$\mathbf{z}^{k+1,j_k} := \operatorname{Prox}_{g*/\gamma_{k,j_k}}\left(\frac{A\mathbf{x}^{k+1}}{\gamma_{k,j_k}}\right)$$

it holds $\theta_{k+1} := \dfrac{\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1,j_k}, \mathbf{u}^{k+1}, \delta_k, \gamma_{k,j_k})}{f(K\mathbf{x}^{k+1})} > 0.$
Update $\gamma_{k+1} := \gamma_{k,j_k}$.
Update $\delta_{k+1} := \chi\left(L_{\nabla h} + \dfrac{2\sigma_A^2}{\gamma_{k+1}}\right)$.
Update $\mathbf{z}^{k+1} := \mathbf{z}^{k+1,j_k}$.
If $\|\mathbf{z}^{k+1}\| > \min\left(\dfrac{\varepsilon}{\gamma_{k+1}}, \sqrt{\dfrac{2\varepsilon}{\gamma_{k+1}}}\right)$, *then*
  Update $\gamma_{k+1} := \gamma_{k+1}q$.
  Update $\delta_{k+1} := \chi\left(L_{\nabla h} + \dfrac{2\sigma_A^2}{\gamma_{k+1}}\right)$.
End If

LEMMA 5.2 (Well-definedness of Algorithm 5.1). *Suppose Assumption 3.1 holds. Then the following statements are true:*

(i) *It holds* $\|\mathbf{z}^{k+1}\| \leqslant \ell + 1$ *for all* $k \geqslant 0$.
(ii) *The procedure of finding the smallest* $j_k \in \{0,1,2,\ldots\}$ *such that* $\theta_{k+1} > 0$ *is executed in every iteration of Algorithm 5.1 a finite number of times, and so the algorithm is well-defined. Moreover,* $\gamma_{k+1} \leqslant \gamma_k$ *for all* $k \geqslant 0$.
(iii) *There exists a constant* $\gamma > 0$, $\chi > 1$ *and an index* $K_0 \geqslant 0$ *such that* $\gamma_k = \gamma > 0$, $\delta_k = \delta := \chi(L_{\nabla h} + \frac{2\sigma_A^2}{\gamma})$, *and* $\|\mathbf{z}^{k+1}\| \leqslant \min\left(\frac{\varepsilon}{\gamma}, \sqrt{\frac{2\varepsilon}{\gamma}}\right)$ *for all* $k \geqslant K_0$.

Proof. (i) From the construction of the algorithm, $\mathbf{x}^k \in \mathcal{S}$ for all $k \geqslant 0$. So, by Assumption 3.1(f), for all $k \geqslant 0$ there exists $\tilde{\mathbf{z}}^k \in \partial g(A\mathbf{x}^{k+1})$ with $\|\tilde{\mathbf{z}}^k\| \leqslant \ell + 1$.

Taking into account the definitions of $\mathbf{z}^{k+1,j_k}$ and the proximal operator, for all $k \geqslant 0$, we have

$$\begin{aligned}
& g^*(\mathbf{z}^{k+1,j_k}) - \langle A\mathbf{x}^{k+1}, \mathbf{z}^{k+1,j_k}\rangle + \frac{\gamma_{k,j_k}}{2}\|\mathbf{z}^{k+1,j_k}\|^2 \\
\leqslant\ & g^*(\tilde{\mathbf{z}}^k) - \langle A\mathbf{x}^{k+1}, \tilde{\mathbf{z}}^k\rangle + \frac{\gamma_{k,j_k}}{2}\|\tilde{\mathbf{z}}^k\|^2 = -g(A\mathbf{x}^{k+1}) + \frac{\gamma_{k,j_k}}{2}\|\tilde{\mathbf{z}}^k\|^2 \\
\leqslant\ & g^*(\mathbf{z}^{k+1,j_k}) - \langle A\mathbf{x}^{k+1}, \mathbf{z}^{k+1,j_k}\rangle + \frac{\gamma_{k,j_k}}{2}\|\tilde{\mathbf{z}}^k\|^2.
\end{aligned} \tag{5.1}$$

Since $\gamma_{k,j_k} > 0$, it follows that $\|\mathbf{z}^{k+1,j_k}\|^2 \leqslant \|\tilde{\mathbf{z}}^k\|^2 \leqslant (\ell+1)^2$, consequently, $\|\mathbf{z}^{k+1}\| \leqslant \ell + 1$ for all $k \geqslant 0$.

(ii) Let $k \geqslant 0$ and $j_k \in \{0,1,2,...\}$. From (5.1) and Assumption 3.1(e), it holds

$$\begin{aligned}
& \Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1,j_k}, \mathbf{u}^{k+1}, \delta_k, \gamma_{k,j_k}) \\
&= \langle A\mathbf{x}^{k+1}, \mathbf{z}^{k+1,j_k}\rangle - g^*(\mathbf{z}^{k+1,j_k}) + h(\mathbf{x}^{k+1}) + \frac{\delta_k}{2}\|\mathbf{x}^{k+1} - \mathbf{u}^{k+1}\|^2 - \frac{\gamma_{k,j_k}}{2}\|\mathbf{z}^{k+1,j_k}\|^2 \\
&\geqslant g(A\mathbf{x}^{k+1}) + h(\mathbf{x}^{k+1}) + \frac{\delta_k}{2}\|\mathbf{x}^{k+1} - \mathbf{u}^{k+1}\|^2 - \frac{\gamma_{k,j_k}}{2}\|\tilde{\mathbf{z}}^k\|^2 \geqslant \alpha - \frac{\gamma_{k,j_k}}{2}\|\tilde{\mathbf{z}}^k\|^2.
\end{aligned}$$

Since $\|\tilde{\mathbf{z}}^k\| \leqslant \ell+1$, it is evident that after finitely many increases of j_k with 1 we obtain $\frac{\gamma_{k,j_k}}{2}\|\tilde{\mathbf{z}}^k\|^2 < \frac{\alpha}{2}$ and, therefore, $\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1,j_k}, \mathbf{u}^{k+1}, \delta_k, \gamma_{k,j_k}) > 0$. Consequently, Algorithm 5.1 is well-defined. Finally, from the formulation of the algorithm we easily see that $\gamma_{k+1} \leqslant \gamma_k$ for all $k \geqslant 0$.

(iii) In order to prove the statement, it is sufficient to show that there exist $\gamma > 0$ and $K_0 \geqslant 0$ such that $\gamma_k = \gamma > 0$ for all $k \geqslant K_0$. Assuming the contrary, there exists a strictly decreasing subsequence $\{\gamma_{k_s}\}$ such that $\gamma_{k_s} \to 0$ as $s \to +\infty$. As $\|\mathbf{z}^{k_s}\| \leqslant \ell + 1$ for all $s \geqslant 0$, there exists $s_0 \geqslant 0$ such that inequality in the "If-End If" statement is not verified for all $s \geqslant s_0$. Therefore, as $\{\gamma_{k_s}\}$ is strictly decreasing, for all $s \geqslant s_0$ there exists $\widehat{k}_s \in \mathbb{N}$ with $k_s \leqslant \widehat{k}_s \leqslant k_{s+1}$ such that $\theta_{\widehat{k}_s+1,0} \leqslant 0$. Using a similar argument as in (ii), this implies that $\gamma_{\widehat{k}_s} = \gamma_{\widehat{k}_s,0} \geqslant \frac{2\alpha}{(\ell+1)^2} > 0$ for all $s \geqslant s_0$. The monotonicity of the sequence $\{\gamma_k\}$ leads to $\gamma_{\widehat{k}_s} \to 0$ as $s \to +\infty$, and further to a contradiction. ∎

6. Convergence analysis of Adaptive FSPS.

We provide the convergence analysis for Algorithm 5.1.

6.1. Subsequential convergence.

To simplify the presentation, we denote $\boldsymbol{W}^k := (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)$ for all $k \geqslant 0$.

THEOREM 6.1. *Suppose Assumption 3.1 holds. Let* $0 < \beta < 2$, $\gamma > 0$, $\chi > 1$ *and* $K_0 \geqslant 0$ *satisfy* $\gamma_k = \gamma > 0$, $\delta_k = \delta := \chi(L_{\nabla h} + \frac{2\sigma_A^2}{\gamma})$, *and* $\|\mathbf{z}^{k+1}\| \leqslant \min\left(\frac{\varepsilon}{\gamma}, \sqrt{\frac{2\varepsilon}{\gamma}}\right)$ *for* $k \geqslant K_0$, *as indicated by Lemma 5.2 (iii). Let*

$$c_1 := \frac{(\chi - 1)\left(L_{\nabla h} + \frac{2\sigma_A^2}{\gamma}\right)}{2}, \quad c_2 := \delta(2-\beta)/2\beta, \quad c_3 := \gamma/2.$$

Then, for all $k \geqslant K_0 + 1$, *the following statements are true:*

(i) $\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma) + \theta_k f(K\mathbf{x}^k) - \theta_k\left(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})\right)$
$\leqslant \Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) - c_1\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - c_2\|\mathbf{u}^k - \mathbf{u}^{k+1}\|^2 - c_3\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2$;

(ii) $\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma) - \theta_k f(K\mathbf{x}^{k+1})$
$\leqslant -c_1\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - c_2\|\mathbf{u}^k - \mathbf{u}^{k+1}\|^2 - c_3\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2$.

Proof. Let $k \geqslant K_0 + 1$. (i) Similar to the proof of (i) in Theorem 4.3, we obtain:

$$\begin{aligned}&\Psi(\mathbf{x}^{k+1}, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) + \theta_k\left[f(K\mathbf{x}^k) - \left(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})\right)\right]\\ &\leqslant \Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) - \frac{\delta - L_{\nabla h}}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \end{aligned} \tag{6.1}$$

Similar to the proof for the inequality (4.7), we get

$$\begin{aligned}&\langle \mathbf{z}^{k+1}, A\mathbf{x}^{k+1}\rangle - g^*(\mathbf{z}^{k+1}) - \frac{\gamma}{2}\|\mathbf{z}^{k+1}\|^2\\ &\leqslant \langle \mathbf{z}^k, A\mathbf{x}^{k+1}\rangle - g^*(\mathbf{z}^k) - \frac{\gamma}{2}\|\mathbf{z}^k\|^2 + \frac{\sigma_A^2}{2\gamma}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2.\end{aligned} \tag{6.2}$$

Using $A\mathbf{x}^{k+1} - \gamma\mathbf{z}^{k+1} \in \partial g^*(\mathbf{z}^{k+1})$ and $A\mathbf{x}^k - \gamma\mathbf{z}^k \in \partial g^*(\mathbf{z}^k)$, and the monotonicity of the subdifferential operator, it yields

$$\gamma\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2 \leqslant -\langle \mathbf{z}^k - \mathbf{z}^{k+1}, A(\mathbf{x}^{k+1} - \mathbf{x}^k)\rangle \leqslant \frac{\gamma}{2}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2 + \frac{\sigma_A^2}{2\gamma}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2,$$

and further, in combination with (6.2),

$$\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^k, \delta, \gamma) \leqslant \Psi(\mathbf{x}^{k+1}, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) + \frac{\sigma_A^2}{\gamma}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - \frac{\gamma}{2}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2. \tag{6.3}$$

Using the extrapolation step, we get

$$\frac{\delta}{2}\|\mathbf{u}^{k+1}-\mathbf{x}^{k+1}\|^2=\frac{\delta}{2}\|\mathbf{u}^{k}-\mathbf{x}^{k+1}\|^2-\frac{\delta(1-(1-\beta)^2)}{2\beta^2}\|\mathbf{u}^{k}-\mathbf{u}^{k+1}\|^2,$$

which leads to

$$\Psi(\mathbf{x}^{k+1},\mathbf{z}^{k+1},\mathbf{u}^{k+1},\delta,\gamma)\leqslant\Psi(\mathbf{x}^{k+1},\mathbf{z}^{k+1},\mathbf{u}^{k},\delta,\gamma)-\frac{\delta(2-\beta)}{2\beta}\|\mathbf{u}^{k}-\mathbf{u}^{k+1}\|^2.$$

Finally, by adding (6.1), (6.3) with the above, the assertion follows by using the definition of δ.

(ii) Follows from (i) by using that $f(K\mathbf{x}^{k+1})\geqslant\langle K\mathbf{x}^{k+1},\mathbf{y}^{k+1}\rangle-f^*(\mathbf{y}^{k+1})$, $\theta_k>0$, and $\theta_k f(K\mathbf{x}^k)=\Psi(\mathbf{x}^k,\mathbf{z}^k,\mathbf{u}^k,\delta,\gamma)$. ∎

Let γ and δ be the constants indicated in Lemma 5.2 (iii), and the merit function $\Pi:\mathbb{R}^n\times\mathrm{dom}g^*\times\mathbb{R}^n\to\overline{\mathbb{R}}$ defined by

$$\Pi(\mathbf{x},\mathbf{z},\mathbf{u})=\frac{\Psi(\mathbf{x},\mathbf{z},\mathbf{u},\delta,\gamma)}{f(K\mathbf{x})}=\frac{\langle\mathbf{z},A\mathbf{x}\rangle-g^*(\mathbf{z})+h(\mathbf{x})+\iota_{\mathcal{S}}(\mathbf{x})+\frac{\delta}{2}\|\mathbf{x}-\mathbf{u}\|^2-\frac{\gamma}{2}\|\mathbf{z}\|^2}{f(K\mathbf{x})}.$$

Theorem 6.2 (Subsequential convergence). *Suppose Assumption 3.1 holds. Let $0<\beta<2$, $\chi>1$, $\gamma>0$ and $K_0\geqslant 0$ satisfy $\gamma_k=\gamma>0$, $\delta_k=\delta:=(\chi-1)\left(L_{\nabla h}+\frac{2\sigma_A^2}{\gamma}\right)$, and $\|\mathbf{z}^{k+1}\|\leqslant\min\left(\frac{\varepsilon}{\gamma},\sqrt{\frac{2\varepsilon}{\gamma}}\right)$ for $k\geqslant K_0$, as indicated by Lemma 5.2 (iii). Let Ω be the set of the accumulation points of the sequence $\{\boldsymbol{W}^k\}$. Then, the following statements are true:*

(i) The sequence $\left\{\theta_k=\frac{\Psi(\mathbf{x}^k,\mathbf{z}^k,\mathbf{u}^k,\delta,\gamma)}{f(K\mathbf{x}^k)}=\Pi(\mathbf{x}^k,\mathbf{z}^k,\mathbf{u}^k)\right\}$ is nonincreasing and there exists a scalar $\overline{\theta}\geqslant 0$ such that $\lim_{k\to+\infty}\theta_k=\overline{\theta}$.

(ii) The sequence $\{\boldsymbol{W}^k\}$ is bounded.

(iii) For every $(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{\mathbf{z}},\overline{\mathbf{u}})\in\Omega$ it holds $\Pi(\overline{\mathbf{x}},\overline{\mathbf{z}},\overline{\mathbf{u}})=\overline{\theta}$.

(iv) If $K_\varepsilon:=\{\mathbf{x}\,|\,\mathrm{dist}(\mathbf{x},A(\mathcal{S}))\leqslant\varepsilon\}\subseteq\mathrm{int}(\mathrm{dom}g)$, then g is Lipschitz continuous on the compact set K_ε with some Lipschitz constant $\kappa>0$. In this case, any accumulation point of the sequence $\{\mathbf{x}^k\}$ is a limiting $(2\kappa\varepsilon,(2\kappa+1)\varepsilon)$-lifted approximate stationary point of (1.1).

(v) It holds that $\lim\limits_{k\to+\infty}\dfrac{\langle K\mathbf{x}^{k+1},\mathbf{y}^{k+1}\rangle-f^*(\mathbf{y}^{k+1})}{f(K\mathbf{x}^k)}=1$. Furthermore, there exists an index $K_1\geqslant K_0+1$ such that

$$0<m\leqslant\langle K\mathbf{x}^k,\mathbf{y}^k\rangle-f^*(\mathbf{y}^k)\leqslant f(K\mathbf{x}^k)\leqslant M\quad\forall k\geqslant K_1,\tag{6.4}$$

where m and M are the bounds from Lemma 3.5.

Proof. (i) It follows from Theorem 6.1 (ii) that for all $k\geqslant K_0+1$

$$\theta_{k+1}\leqslant\theta_k-\frac{1}{M}\left(c_1\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2+c_2\|\mathbf{u}^k-\mathbf{u}^{k+1}\|^2+c_3\|\mathbf{z}^k-\mathbf{z}^{k+1}\|^2\right),\tag{6.5}$$

where $M>0$ is the constant provided by Lemma 3.5. Thus,

$$\|\mathbf{x}^k-\mathbf{x}^{k+1}\|\to 0,\ \|\mathbf{u}^k-\mathbf{u}^{k+1}\|\to 0,\ \|\mathbf{z}^{k+1}-\mathbf{z}^{k}\|\to 0,\ \|\mathbf{x}^{k+1}-\mathbf{u}^{k}\|\to 0,\tag{6.6}$$

as $k\to+\infty$ and the sequence $\{\theta_k\}$ is nonincreasing. Thus, $\overline{\theta}:=\lim_{k\to\infty}\theta_k\geqslant 0$ exists.

(ii) Since $\mathcal{S}$ is a compact set, the sequence $\{\mathbf{x}^k\}$ is bounded by construction, which, according to (6.6), guarantees that $\{\mathbf{u}^k\}$ is bounded. The sequence $\{\mathbf{y}^k\}$ is bounded due to Assumption 3.1(d), and the sequence $\{\mathbf{z}^k\}$ is bounded due to Assumption 3.1(f).

(iii) Let $\overline{\boldsymbol{W}} = (\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}, \overline{\mathbf{u}})$ be an accumulation point of the sequence and $\{\boldsymbol{W}^k\}$ and $\{\boldsymbol{W}^{k_j}\}$ be a subsequence such that $\lim_{j\to+\infty} \boldsymbol{W}^{k_j} = \overline{\boldsymbol{W}}$.

From $\lim_{j\to+\infty} \frac{\Psi(\mathbf{x}^{k_j},\mathbf{z}^{k_j},\mathbf{u}^{k_j},\delta,\gamma)}{f(K\mathbf{x}^{k_j})} = \lim_{j\to+\infty} \theta_{k_j} = \overline{\theta}$ and $\lim_{j\to+\infty} f(K\mathbf{x}^{k_j}) = f(K\overline{\mathbf{x}}) > 0$, which holds due to Assumption 3.1(d), by noting that $\{K\mathbf{x}^{k_j}\} \subseteq K(\mathcal{S}) \subseteq \mathrm{int}(\mathrm{dom} f)$ and $K(\mathcal{S})$ is closed, we have that the following limit exists:

$$\overline{\Psi} := \lim_{j\to\infty} \Psi(\mathbf{x}^{k_j}, \mathbf{z}^{k_j}, \mathbf{u}^{k_j}, \delta, \gamma) \in \mathbb{R}. \tag{6.7}$$

Next, we show that $\overline{\Psi} = \Psi(\overline{\mathbf{x}}, \overline{\mathbf{z}}, \overline{\mathbf{u}}, \delta, \gamma)$. From (6.7), $\mathbf{x}^{k_j} \in \mathcal{S}$, g^* is lower semicontinuous and the definition of $\Psi(\cdot,\cdot,\cdot,\delta,\gamma)$, we have that $\Psi(\overline{\mathbf{x}}, \overline{\mathbf{z}}, \overline{\mathbf{u}}, \delta, \gamma) \geqslant \overline{\Psi}$. Invoking the update scheme, for every $j \geqslant 0$ such that $k_j \geqslant K_0 + 1$ it holds

$$g^*(\overline{\mathbf{z}}) - \langle \overline{\mathbf{z}}, A\mathbf{x}^{k_j}\rangle + \frac{\gamma}{2}\|\overline{\mathbf{z}}\|^2 \geqslant g^*(\mathbf{z}^{k_j}) - \langle \mathbf{z}^{k_j}, A\mathbf{x}^{k_j}\rangle + \frac{\gamma}{2}\|\mathbf{z}^{k_j}\|^2$$

and, further,

$$-g^*(\overline{\mathbf{z}}) + \langle \overline{\mathbf{z}}, A\mathbf{x}^{k_j}\rangle - \frac{\gamma}{2}\|\overline{\mathbf{z}}\|^2 + h(\mathbf{x}^{k_j}) \leqslant -g^*(\mathbf{z}^{k_j}) + \langle \mathbf{z}^{k_j}, A\mathbf{x}^{k_j}\rangle - \frac{\gamma}{2}\|\mathbf{z}^{k_j}\|^2 + h(\mathbf{x}^{k_j}).$$

We let $j \to +\infty$ and get

$$-g^*(\overline{\mathbf{z}}) + \langle \overline{\mathbf{z}}, A\overline{\mathbf{x}}\rangle - \frac{\gamma}{2}\|\overline{\mathbf{z}}\|^2 + h(\overline{\mathbf{x}}) \leqslant \lim_{j\to+\infty}(-g^*(\mathbf{z}^{k_j}) + \langle \mathbf{z}^{k_j}, A\mathbf{x}^{k_j}\rangle - \frac{\gamma}{2}\|\mathbf{z}^{k_j}\|^2 + h(\mathbf{x}^{k_j})),$$

so, $\Psi(\overline{\mathbf{x}}, \overline{\mathbf{z}}, \overline{\mathbf{u}}, \delta, \gamma) \leqslant \overline{\Psi}$. In conclusion, $\Psi(\overline{\mathbf{x}}, \overline{\mathbf{z}}, \overline{\mathbf{u}}, \delta, \gamma) = \overline{\Psi}$ and $\Pi(\overline{\mathbf{x}}, \overline{\mathbf{z}}, \overline{\mathbf{u}}) = \overline{\theta}$.

(iv) Invoking the update rules for $\mathbf{x}^{k+1}$, $\mathbf{y}^{k+1}$, $\mathbf{z}^{k+1}$ and $\mathbf{u}^{k+1}$, for all $k \geqslant K_0 + 1$ it yields

$$\begin{cases} \mathbf{y}^{k+1} \in \partial f(K\mathbf{x}^k), \\ 0 \in \partial\iota_{\mathcal{S}}(\mathbf{x}^{k+1}) + A^*\mathbf{z}^k + \nabla h(\mathbf{x}^k) - \theta_k K^*\mathbf{y}^{k+1} + \delta(\mathbf{x}^{k+1} - \mathbf{u}^k), \\ A\mathbf{x}^{k+1} - \gamma\mathbf{z}^{k+1} \in \partial g^*(\mathbf{z}^{k+1}), \\ \mathbf{u}^{k+1} = (1-\beta)\mathbf{u}^k + \beta\mathbf{x}^{k+1}. \end{cases} \tag{6.8}$$

Let $\overline{\boldsymbol{W}} = (\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}, \overline{\mathbf{u}})$ be an accumulation point of the sequence of $\{\boldsymbol{W}^k\}$, and let $\{\boldsymbol{W}^{k_j} = (\mathbf{x}^{k_j}, \mathbf{y}^{k_j}, \mathbf{z}^{k_j}, \mathbf{u}^{k_j})\}$ be a subsequence converging to $\overline{\boldsymbol{W}}$ as $j \to +\infty$. From (6.6), we see that $\mathbf{x}^{k_j-1} \to \overline{\mathbf{x}}$ and $\mathbf{u}^{k_j-1} \to \overline{\mathbf{u}}$ as $j \to +\infty$. Then, letting $k = k_j - 1$ and $j \to +\infty$ in the above system and taking into account the fact that the graph of the convex subdifferential is closed, we obtain

$$\begin{cases} \overline{\mathbf{y}} \in \partial f(K\overline{\mathbf{x}}), \\ \mathbf{0} \in \partial\iota_{\mathcal{S}}(\overline{\mathbf{x}}) + A^*\overline{\mathbf{z}} + \nabla h(\overline{\mathbf{x}}) - \overline{\theta} K^*\overline{\mathbf{y}}, \\ A\overline{\mathbf{x}} - \gamma\overline{\mathbf{z}} \in \partial g^*(\overline{\mathbf{z}}), \\ \overline{\mathbf{u}} = \overline{\mathbf{x}}. \end{cases} \tag{6.9}$$

Since $\|\mathbf{z}^{k+1}\| \leqslant \min\left(\frac{\varepsilon}{\gamma}, \sqrt{\frac{2\varepsilon}{\gamma}}\right)$ for all $k \geqslant K_0$, it yields $\|\overline{\mathbf{z}}\| \leqslant \min\left(\frac{\varepsilon}{\gamma}, \sqrt{\frac{2\varepsilon}{\gamma}}\right)$. The third inclusion relation in (6.9) guarantees that $\mathrm{dist}(\partial g^*(\overline{\mathbf{z}}), A\overline{\mathbf{x}}) \leqslant \varepsilon$. Therefore, according to Lemma 2.4, $\overline{\mathbf{z}} \in \partial_{2\kappa\varepsilon} g(A\overline{\mathbf{x}})$, which, combined with the first two inclusion relations in (6.9), leads to $0 \in (A^*\partial_{2\kappa\varepsilon} g(A\overline{\mathbf{x}}) + \nabla h(\overline{\mathbf{x}}) + \partial\iota_{\mathcal{S}}(\overline{\mathbf{x}}))\, f(K\overline{\mathbf{x}}) - \overline{\Psi} K^*\partial f(K\overline{\mathbf{x}})$.

As seen in the proof of statement (iii), we have $\overline{\Psi} = \langle A\overline{\mathbf{x}}, \overline{\mathbf{z}}\rangle - g^*(\overline{\mathbf{z}}) + h(\overline{\mathbf{x}}) - \frac{\gamma}{2}\|\overline{\mathbf{z}}\|^2$, therefore

$$\begin{aligned} |\overline{\Psi} - (g(A\overline{\mathbf{x}}) + h(\overline{\mathbf{x}}))| &= (g(A\overline{\mathbf{x}}) + h(\overline{\mathbf{x}})) - \left(\langle A\overline{\mathbf{x}}, \overline{\mathbf{z}}\rangle - g^*(\overline{\mathbf{z}}) + h(\overline{\mathbf{x}}) - \frac{\gamma}{2}\|\overline{\mathbf{z}}\|^2\right) \\ &= g(A\overline{\mathbf{x}}) + g^*(\overline{\mathbf{z}}) - \langle A\overline{\mathbf{x}}, \overline{\mathbf{z}}\rangle + \frac{\gamma}{2}\|\overline{\mathbf{z}}\|^2 \leqslant 2\kappa\varepsilon + \varepsilon = (2\kappa+1)\varepsilon. \end{aligned}$$

Thus, $\overline{\mathbf{x}}$ is a limiting $(2\kappa\varepsilon, (2\kappa+1)\varepsilon)$-lifted approximate stationary point of (1.1).

(v) Invoking the first inclusion relation in (6.8) and Lemma 3.5, we obtain for all $k \geqslant K_0 + 1$

$$\begin{aligned} \frac{|f(K\mathbf{x}^k) - \left(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})\right)|}{f(K\mathbf{x}^k)} &= \frac{|\langle \mathbf{y}^{k+1}, K(\mathbf{x}^k - \mathbf{x}^{k+1})\rangle|}{f(K\mathbf{x}^k)} \\ &\leqslant \frac{|\langle \mathbf{y}^{k+1}, K(\mathbf{x}^k - \mathbf{x}^{k+1})\rangle|}{m}. \end{aligned}$$

Using that $\mathbf{x}^k - \mathbf{x}^{k+1} \to 0$ as $k \to +\infty$ and the boundedness of $\{\mathbf{y}^k\}$, we obtain

$$\lim_{k\to+\infty} \frac{|f(K\mathbf{x}^k) - \left(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})\right)|}{f(K\mathbf{x}^k)} = 0. \tag{6.10}$$

The second statement is a direct consequence of (6.10). □

Let $m > 0$ be the scalar introduced in Lemma 3.5, γ and δ the constants indicated in Lemma 5.2 (iii), and the following modified merit function $\Gamma : \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathrm{dom} f^* : \langle K\mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y}) > m/2\} \times \mathrm{dom} g^* \times \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by

$$\Gamma(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}) = \frac{\Psi(\mathbf{x}, \mathbf{z}, \mathbf{u}, \delta, \gamma)}{\langle K\mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y})} = \frac{\langle \mathbf{z}, A\mathbf{x}\rangle - g^*(\mathbf{z}) + h(\mathbf{x}) + \iota_{\mathcal{S}}(\mathbf{x}) + \frac{\delta}{2}\|\mathbf{x} - \mathbf{u}\|^2 - \frac{\gamma}{2}\|\mathbf{z}\|^2}{\langle K\mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y})}.$$

In the following we show that values of Γ along the sequence $\{(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)\}$ converge to $\overline{\theta}$ as $k \to +\infty$ and that it takes this value at every point of Ω.

THEOREM 6.3. *Suppose Assumption 3.1 holds. Let $0 < \beta < 2$, $\gamma > 0$, $\chi > 1$ and $K_0 \geqslant 0$ satisfy $\gamma_k = \gamma > 0$, $\delta_k = \delta := \chi(L_{\nabla h} + \frac{2\sigma_A^2}{\gamma})$, and $\|\mathbf{z}^{k+1}\| \leqslant \min\left(\frac{\varepsilon}{\gamma}, \sqrt{\frac{2\varepsilon}{\gamma}}\right)$ for $k \geqslant K_0$, as indicated by Lemma 5.2 (iii), and $K_1 \geqslant K_0 + 1$ such that (6.4) holds, as indicated by Theorem 6.2 (v). Then, the following statements are true:*

(i) There exists $c > 0$ such that for all $k \geqslant K_1$

$$\begin{aligned} &\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}) \\ &\leqslant \Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) - c\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - c\|\mathbf{u}^k - \mathbf{u}^{k+1}\|^2 - c\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2; \end{aligned} \tag{6.11}$$

(ii) $\lim_{k\to+\infty} \Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)$ exists and it is equal to $\overline{\theta} = \lim_{k\to+\infty} \theta_k$;

(iii) For every $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}, \overline{\mathbf{u}}) \in \Omega$ it holds $\Gamma(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}, \overline{\mathbf{u}}) = \overline{\theta}$.

Proof. (i) By using the fact of $\theta_k f(K\mathbf{x}^k) = \Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma)$, from Theorem 6.1 (i) we obtain for all $k \geqslant K_1$

$$\begin{aligned} \Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma) &\leqslant \theta_k \left(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})\right) \\ &\quad - c_1\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - c_2\|\mathbf{u}^k - \mathbf{u}^{k+1}\|^2 - c_3\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2. \end{aligned}$$

Since $0 < m \leqslant \langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1}) \leqslant M$ for all $k \geqslant K_1$, it yields

$$\eta_{k+1} \leqslant \theta_k - \frac{c_1}{M}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - \frac{c_2}{M}\|\mathbf{u}^k - \mathbf{u}^{k+1}\|^2 - \frac{c_3}{M}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2, \tag{6.12}$$

where $\eta_{k+1} := \Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}) = \frac{\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma)}{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})}$. Then one can choose $c := \frac{1}{M}\min(c_1, c_2, c_3)$ and the conclusion follow as $\theta_k \leqslant \eta_k$ for all $k \geqslant K_1$. The proofs of (ii) and (iii) follow similarly to items (i) and (iii) of Theorem 6.2 and are therefore omitted. ∎

6.2. Global convergence. To this end, we will provide two different settings in which we can bound the distance between the origin and the limiting subdifferential of Γ and Π, respectively. The two settings are considered below by supposing that Assumption 3.1 holds, $0 < \beta < 2$, $\gamma > 0$, $\chi > 1$ and $K_0 \geqslant 0$ satisfy $\gamma_k = \gamma > 0$, $\delta_k = \delta := \chi(L_{\nabla h} + \frac{2\sigma_A^2}{\gamma})$ for all $k \geqslant K_0$, as indicated by Lemma 5.2 (iii), and $K_1 \geqslant K_0 + 1$ is such that (6.4) holds, as indicated by Theorem 6.2 (v).

Case I: **f^* satisfies the calm condition over its effective domain and g is essentially strictly convex.** The following characterization of the Fréchet subdifferential of the merit function Γ follows from Lemma 2.3.

LEMMA 6.4. *Suppose Assumption 3.1 holds. Let f^* satisfy the calm condition at $\widehat{\mathbf{y}} \in \mathrm{dom} f^*$, $\widehat{\mathbf{x}} \in \mathcal{S}$ be such that $\langle K\widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle - f^*(\widehat{\mathbf{y}}) > m/2$, and g^* be differentiable at $\widehat{\mathbf{z}} \in \mathrm{int}(\mathrm{dom} g^*)$. Denote $\alpha_1 := \Psi(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \widehat{\mathbf{u}}, \delta, \gamma)$ and $\alpha_2 := \langle K\widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle - f^*(\widehat{\mathbf{y}})$, and suppose that $\alpha_1 > 0$. Then, there exist open sets $\mathcal{O}_i$, $i = 1, 2$, such that $\langle K\widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle - f^*(\widehat{\mathbf{y}}) > m/2$ for all $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in \mathcal{O}_1 \times \mathcal{O}_2$, and*

$$\hat{\partial}\Gamma(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}, \widehat{\mathbf{z}}, \widehat{\mathbf{u}}) = \left\{ (\boldsymbol{\xi}_{\mathbf{x}}, \boldsymbol{\xi}_{\mathbf{y}}, \boldsymbol{\xi}_{\mathbf{z}}, \boldsymbol{\xi}_{\mathbf{u}}) \;\middle|\; \begin{array}{l} \boldsymbol{\xi}_{\mathbf{x}} \in \dfrac{\alpha_2(A^*\widehat{\mathbf{z}} + \nabla h(\widehat{\mathbf{x}}) + \partial\iota_{\mathcal{S}}(\widehat{\mathbf{x}}) + \delta(\widehat{\mathbf{x}} - \widehat{\mathbf{u}})) - \alpha_1 K^*\widehat{\mathbf{y}}}{(\langle K\widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle - f^*(\widehat{\mathbf{y}}))^2} \\ \boldsymbol{\xi}_{\mathbf{y}} \in \dfrac{\alpha_1(\partial f^*(\widehat{\mathbf{y}}) - K\widehat{\mathbf{x}})}{(\langle K\widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle - f^*(\widehat{\mathbf{y}}))^2} \\ \boldsymbol{\xi}_{\mathbf{z}} = \dfrac{A\widehat{\mathbf{x}} - \nabla g^*(\widehat{\mathbf{z}}) - \gamma\widehat{\mathbf{z}}}{\langle K\widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle - f^*(\widehat{\mathbf{y}})} \\ \boldsymbol{\xi}_{\mathbf{u}} = \dfrac{\delta(\widehat{\mathbf{u}} - \widehat{\mathbf{x}})}{\langle K\widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle - f^*(\widehat{\mathbf{y}})} \end{array} \right\}.$$

THEOREM 6.5. *Suppose that f^* satisfies the calm condition over its effective domain and g is essentially strictly convex. Then there exists $\zeta > 0$ such that for all $k \geqslant K_1$*

$$\mathrm{dist}(\mathbf{0}, \partial\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})) \leqslant \zeta(\|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \|\mathbf{u}^k - \mathbf{u}^{k+1}\| + \|\mathbf{z}^k - \mathbf{z}^{k+1}\|).$$

Proof. Let $k \geqslant K_1$ be fixed. It holds

$$\mathrm{dist}(\mathbf{0}, \partial\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})) \leqslant \mathrm{dist}(\mathbf{0}, \hat{\partial}\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})).$$

Since g is essentially strictly convex, g^* is essentially smooth [26, Theorem 26.3]. According to the third inclusion relation in (6.8), we have $A\mathbf{x}^{k+1} - \gamma\mathbf{z}^{k+1} \in \partial g^*(\mathbf{z}^{k+1})$, which means $\mathbf{z}^{k+1} \in \mathrm{int}(\mathrm{dom} g^*)$. In addition, $m \leqslant \langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1}) \leqslant M$ and $\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma) > 0$. Thus, one can make use of the formula provided in Lemma 6.4 to characterize the subdifferential of Γ at $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})$. Invoking again (6.8), we have $-(A^*\mathbf{z}^k + \nabla h(\mathbf{x}^k) - \theta_k K^*\mathbf{y}^{k+1} + \delta(\mathbf{x}^{k+1} - \mathbf{u}^k)) \in \hat{\partial}\iota_{\mathcal{S}}(\mathbf{x}^{k+1})$ and $\mathbf{y}^{k+1} \in \partial f(K\mathbf{x}^k)$ or, equivalently, $K\mathbf{x}^k \in \partial f^*(\mathbf{y}^{k+1})$, and $A\mathbf{x}^{k+1} - (\nabla g^*(\mathbf{z}^{k+1}) + \gamma\mathbf{z}^{k+1}) = 0$.

Thus, for

$$\boldsymbol{\xi}_{\mathbf{x}}^{k+1} := \frac{A^*\mathbf{z}^{k+1} + \nabla h(\mathbf{x}^{k+1}) + \delta(\mathbf{x}^{k+1} - \mathbf{u}^{k+1})}{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})} - \frac{A^*\mathbf{z}^k + \nabla h(\mathbf{x}^k) - \theta_k K^*\mathbf{y}^{k+1} + \delta(\mathbf{x}^{k+1} - \mathbf{u}^k)}{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})}$$

$$- \frac{\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma) K^* \mathbf{y}^{k+1}}{(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1}))^2}$$

$$\boldsymbol{\xi}_{\mathbf{y}}^{k+1} := \frac{\Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma)(-K\mathbf{x}^{k+1} + K\mathbf{x}^k)}{(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1}))^2},$$

$$\boldsymbol{\xi}_{\mathbf{z}}^{k+1} := \frac{A\mathbf{x}^{k+1} - (\nabla g^*(\mathbf{z}^{k+1}) + \gamma \mathbf{z}^{k+1})}{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})} = 0, \ \ \boldsymbol{\xi}_{\mathbf{u}}^{k+1} := \frac{\delta(\mathbf{u}^{k+1} - \mathbf{x}^{k+1})}{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})},$$

we have that $(\boldsymbol{\xi}_{\mathbf{x}}^{k+1}, \boldsymbol{\xi}_{\mathbf{y}}^{k+1}, \boldsymbol{\xi}_{\mathbf{z}}^{k+1}, \boldsymbol{\xi}_{\mathbf{u}}^{k+1}) \in \hat{\partial}\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})$. Consequently,

$$\text{dist}(\mathbf{0}, \hat{\partial}\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})) \leqslant \|\boldsymbol{\xi}_{\mathbf{x}}^{k+1}\| + \|\boldsymbol{\xi}_{\mathbf{y}}^{k+1}\| + \|\boldsymbol{\xi}_{\mathbf{u}}^{k+1}\|. \tag{6.13}$$

Due to the boundedness of the four sequences, the values

$$B_{\mathbf{x}} := \sup_k \|\mathbf{x}^k\|, \ B_{\mathbf{y}} := \sup_k \|\mathbf{y}^k\|, \ B_{\mathbf{z}} := \sup_k \|\mathbf{z}^k\|, \ B_{\mathbf{u}} := \sup_k \|\mathbf{u}^k\|$$

are finite. Since $\{\theta_k\}$ and $\{f(K\mathbf{x}^k)\}$ are bounded, the sequence $\{\Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma)\}$ is also bounded. Let $B_\Psi := \sup_k |\Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma)| < +\infty$. Further, as $\{\mathbf{z}^k\} \subseteq \text{int}(\text{dom}\, g^*)$, g^* is Lipschitz continuous on the closure of $\{\mathbf{z}^k\}$. We denote by L_{g^*} the corresponding Lipschitz constant. This being given, it is evident that

$$\begin{aligned} &|\Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) - \Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma)| \\ &\leqslant \varrho_1 \|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \varrho_2 \|\mathbf{z}^k - \mathbf{z}^{k+1}\| + \varrho_3 \|\mathbf{u}^{k+1} - \mathbf{u}^k\|, \end{aligned} \tag{6.14}$$

where $\varrho_1 := B_{\mathbf{z}}\sigma_A + \delta(B_{\mathbf{x}} + B_{\mathbf{u}}) + L_h$, $\varrho_2 := \sigma_A B_{\mathbf{x}} + L_{g^*} + \gamma B_{\mathbf{z}}$, $\varrho_3 := \delta(B_{\mathbf{x}} + B_{\mathbf{u}})$. Since

$$\begin{aligned} \boldsymbol{\xi}_{\mathbf{x}}^{k+1} = {} & \frac{A^*(\mathbf{z}^{k+1} - \mathbf{z}^k) + \nabla h(\mathbf{x}^{k+1}) - \nabla h(\mathbf{x}^k) + \delta(\mathbf{u}^k - \mathbf{u}^{k+1})}{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})} \\ & + \frac{\Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) \frac{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})}{f(K\mathbf{x}^k)} - \Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma)}{(\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1}))^2} K^* \mathbf{y}^{k+1}, \end{aligned}$$

we obtain

$$\begin{aligned} \|\boldsymbol{\xi}_{\mathbf{x}}^{k+1}\| \leqslant {} & \frac{1}{m}(L_{\nabla h}\|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \sigma_A \|\mathbf{z}^k - \mathbf{z}^{k+1}\| + \delta \|\mathbf{u}^k - \mathbf{u}^{k+1}\|) \\ & + \frac{B_{\mathbf{y}}\sigma_K}{m^2} \left| \Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) \left(\frac{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})}{f(K\mathbf{x}^k)} - 1 \right) \right| \\ & + \frac{B_{\mathbf{y}}\sigma_K}{m^2} \left| \Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) - \Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}, \delta, \gamma) \right|. \end{aligned}$$

From

$$\begin{aligned} & \left| \Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) \left(\frac{\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1})}{f(K\mathbf{x}^k)} - 1 \right) \right| \\ = & \left| \Psi(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta, \gamma) \frac{\langle K(\mathbf{x}^{k+1} - \mathbf{x}^k), \mathbf{y}^{k+1}\rangle}{f(K\mathbf{x}^k)} \right| \leqslant \frac{B_\Psi B_{\mathbf{y}} \sigma_K}{m} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|, \end{aligned}$$

and (6.14), it yields $\|\boldsymbol{\xi}_{\mathbf{x}}^{k+1}\| \leqslant \eta_1 \|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \eta_2 \|\mathbf{u}^k - \mathbf{u}^{k+1}\| + \eta_3 \|\mathbf{z}^k - \mathbf{z}^{k+1}\|$, with $\eta_1 := \frac{L_{\nabla h}}{m} + \frac{B_\Psi B_{\mathbf{y}}^2 \sigma_K^2}{m^3} + \varrho_1 \frac{B_{\mathbf{y}}\sigma_K}{m^2}$, $\eta_2 := \frac{\delta}{m} + \varrho_3 \frac{B_{\mathbf{y}}\sigma_K}{m^2}$ and $\eta_3 := \frac{\sigma_A}{m} + \varrho_2 \frac{B_{\mathbf{y}}\sigma_K}{m^2}$. In addition, we have that $\|\boldsymbol{\xi}_{\mathbf{y}}^{k+1}\| \leqslant B_\Psi \frac{\sigma_K}{m^2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|$ and $\|\boldsymbol{\xi}_{\mathbf{u}}^{k+1}\| \leqslant \frac{\delta|1-\beta|}{m\beta} \|\mathbf{u}^k - \mathbf{u}^{k+1}\|$, which, in the light of (6.13), leads to the conclusion. ∎

Case II: f **is differentiable with Lipschitz continuous gradient over an open set containing** $K(\mathcal{S})$ **and** g **is essentially strictly convex.** The workhorse of our analysis will be the merit function Π. The following statement is a direct consequence of [8, Lemma 2.1 (ii)].

LEMMA 6.6. *Suppose Assumption 3.1 holds. Let f be differentiable at $K\widehat{\mathbf{x}} \in \operatorname{int}(\operatorname{dom} f)$ for $\widehat{\mathbf{x}} \in \mathcal{S}$, and g^* be differentiable at $\widehat{\mathbf{z}} \in \operatorname{int}(\operatorname{dom} g^*)$. Denote $\alpha_1 := \Psi(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \widehat{\mathbf{u}}, \delta, \gamma)$ and $\alpha_2 := f(K\widehat{\mathbf{x}})$, and suppose that $\alpha_1 > 0$. Then,*

$$\hat{\partial}\Pi(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \widehat{\mathbf{u}}) = \left\{ (\boldsymbol{\xi}_{\mathbf{x}}, \boldsymbol{\xi}_{\mathbf{z}}, \boldsymbol{\xi}_{\mathbf{u}}) \;\middle|\; \begin{array}{l} \boldsymbol{\xi}_{\mathbf{x}} \in \dfrac{\alpha_2(A^*\widehat{\mathbf{z}} + \nabla h(\widehat{\mathbf{x}}) + \partial\iota_{\mathcal{S}}(\widehat{\mathbf{x}}) + \delta(\widehat{\mathbf{x}} - \widehat{\mathbf{u}})) - \alpha_1 K^*\nabla f(K\widehat{\mathbf{x}})}{(f(K\widehat{\mathbf{x}}))^2} \\ \boldsymbol{\xi}_{\mathbf{z}} = \dfrac{A\widehat{\mathbf{x}} - \nabla g^*(\widehat{\mathbf{z}}) - \gamma\widehat{\mathbf{z}}}{f(K\widehat{\mathbf{x}})} \\ \boldsymbol{\xi}_{\mathbf{u}} = \dfrac{\delta(\widehat{\mathbf{u}} - \widehat{\mathbf{x}})}{f(K\widehat{\mathbf{x}})} \end{array} \right\}.$$

THEOREM 6.7. *Suppose that f is differentiable with Lipschitz continuous gradient on an open set containing $K(\mathcal{S})$, and g is essentially strictly convex. Then there exists $\zeta > 0$ such that for all $k \geqslant K_1$*

$$\operatorname{dist}(\mathbf{0}, \partial\Pi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})) \leqslant \zeta(\|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \|\mathbf{u}^k - \mathbf{u}^{k+1}\| + \|\mathbf{z}^k - \mathbf{z}^{k+1}\|).$$

Proof. The proof is similar to Theorem 6.5, thus omitted here. ◻

Remark 6.8 (Comments on the assumption of essential strict convexity). The assumption of g being essentially strictly convex can be enforced by redefining the functions g and h as $\widetilde{g}(\mathbf{x}) := g(\mathbf{x}) + \frac{s}{2}\|\mathbf{x}\|^2$ and $\widetilde{h}(\mathbf{x}) := h(\mathbf{x}) - \frac{s}{2}\|A\mathbf{x}\|^2$ with $s > 0$. We noticed that, for small $s > 0$, the algorithm exhibits comparable (or simply the same) numerical performance as for $s = 0$.

Remark 6.9. We require that either f^* satisfies the calm condition over its effective domain or f is differentiable with Lipschitz continuous gradient over an open set containing $K(\mathcal{S})$. These conditions can be satisfied in many applications. For example, if f is supercocercive, that is, $\lim_{\|\mathbf{x}\|\to+\infty} \frac{f(\mathbf{x})}{\|\mathbf{x}\|} = +\infty$, then f^* is a real-valued convex function with full domain [4, Proposition 14.15], and so, it is locally Lipschitz (and, in particular, calm). This applies, for instance, to example (b) in the introduction. Regarding example (a), if $p \in (1, +\infty)$, noting that $K(\mathcal{S})$ is a compact set which does not contain the origin, then $f = \|\cdot\|_p$ is differentiable with Lipschitz continuous gradient over an open set containing $K(\mathcal{S})$.

Remark 6.10. According to Definition 2.1, the Kurdyka-Łojasiewicz (KL) property requires that the underlying function is proper and lower semicontinuous. Suppose that g is strictly convex; then its conjugate g^* is differentiable on $\operatorname{int}(\operatorname{dom} g^*)$. Consequently, Γ is lower semicontinuous on

$$\left\{(\mathbf{x}, \mathbf{y}) \in \operatorname{int}(\operatorname{dom} g) \times \operatorname{dom} f^* : \langle K\mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y}) > \frac{m}{2}\right\} \times \operatorname{int}(\operatorname{dom} g^*) \times \mathbb{R}^n.$$

Assume that Γ satisfies the KL property at a point

$$\boldsymbol{W} := (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}) \in \left\{K_\varepsilon \times \operatorname{dom} f^* : \langle K\mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y}) > \frac{m}{2}\right\} \times \operatorname{int}(\operatorname{dom} g^*) \times \mathbb{R}^n \subseteq \partial\Gamma.$$

Then we can restrict the neighborhood U of $\boldsymbol{W}$ such that $\operatorname{Proj}_{\mathbf{z}}(U) \cap \operatorname{int}(\operatorname{dom} g^*)$ is open, where $\operatorname{Proj}_{\mathbf{z}}(U)$ denotes the projection on the space where the block variable $\mathbf{z}$ belongs to. Then, by shrinking U if necessary, we have

$$U = U \cap \left(\{(\mathbf{x}, \mathbf{y}) \in K_\varepsilon \times \operatorname{dom} f^* : \langle K\mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y}) > \frac{m}{2}\} \times \operatorname{int}(\operatorname{dom} g^*) \times \mathbb{R}^n\right),$$

on which Γ remains lower semicontinuous. A similar argument applies to the merit function Π.

Finally, we provide the global convergence result which is in line with [20, Theorem 4] and [11, Theorem 3.4].

THEOREM 6.11. *Let $\varepsilon > 0$. Suppose Assumption 3.1 holds, $K_\varepsilon \subseteq \text{int}(\text{dom}\, g)$, g is nonsmooth and essentially strictly convex and one of the following conditions are fulfilled:*

(i) *f^* satisfies the calm condition over its effective domain and Γ satisfies KL property at every point of $\{(\mathbf{x},\mathbf{y}) \in K_\varepsilon \times \text{dom} f^* : \langle K\mathbf{x},\mathbf{y}\rangle - f^*(\mathbf{y}) > m/2\} \times \text{int}(\text{dom} g^*) \times \mathbb{R}^n$.*

(ii) *f is differentiable with Lipschitz continuous gradient over an open set containing $K(\mathcal{S})$ and Π satisfies KL property at every point of $K_\varepsilon \times \text{int}(\text{dom} g^*) \times \mathbb{R}^n$.*

Let $\{\boldsymbol{W}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)\}$ be the sequence generated by Algorithm 5.1. Then, $\sum_k \left(\|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \|\mathbf{u}^k - \mathbf{u}^{k+1}\| + \|\mathbf{z}^k - \mathbf{z}^{k+1}\|\right) < +\infty$, and $\{\mathbf{x}^k\}$ converges to a limiting $(2\kappa\varepsilon, (2\kappa+1)\varepsilon)$-lifted stationary point of (1.1), where κ is the Lipschitz constant of g on K_ε.

Proof. We prove the statement only in the setting of assumption (i). The proof of the other case can be done analogously. The sequence $\{\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)\}_{k \geqslant K_1}$ is nonincreasing and it converges to $\overline{\theta}$ as $k \to +\infty$. Thus, $\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) \geqslant \overline{\theta}$ for all $k \geqslant K_1$, which allows us to divide the proof into two cases. Case I. There exists $K_2 \geqslant K_1$ such that $\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) = \overline{\theta}$ for $k \geqslant K_2$. Then, $(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}) = (\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k)$ for all $k \geqslant K_2$ due to (6.11), and the conclusion follows. Case II. $\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) > \overline{\theta}$ for all $k \geqslant K_1$. Let Ω denote the set of accumulation points of $\{(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)\}$. Then, Ω is compact. Invoking Theorem 6.2 (iii), according to the uniformized KL property [7], there exist $\varrho > 0$ and $\mu > 0$ and a desingularization function ϕ with the property that for all $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u})$ with $\text{dist}((\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}), \Omega) < \varrho$ and $\overline{\theta} < \Gamma(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}) < \overline{\theta} + \mu$, it holds $\phi'(\Gamma(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}) - \overline{\theta})\text{dist}(\mathbf{0}, \partial\Gamma(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u})) \geqslant 1$. Then, there exists $K_2 \geqslant K_1$ such that $\text{dist}((\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k), \Omega) < \varrho$ and $\overline{\theta} < \Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) < \overline{\theta} + \mu$ for all $k \geqslant K_2$.

Thus, by using Theorem 6.3 and Theorem 6.5, for all $k \geqslant K_2$ it holds

$$
\begin{aligned}
&\phi(\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) - \overline{\theta}) - \phi(\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}) - \overline{\theta}) \\
\geqslant\ & \phi'(\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) - \overline{\theta}) \left(\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) - \Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1})\right) \\
\geqslant\ & \frac{c}{\text{dist}(\mathbf{0}, \partial\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k))} \left(\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \|\mathbf{u}^k - \mathbf{u}^{k+1}\|^2 + \|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2\right) \\
\geqslant\ & \frac{c}{3\zeta} \frac{(\|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \|\mathbf{u}^k - \mathbf{u}^{k+1}\| + \|\mathbf{z}^k - \mathbf{z}^{k+1}\|)^2}{(\|\mathbf{x}^k - \mathbf{x}^{k-1}\| + \|\mathbf{u}^k - \mathbf{u}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\|)},
\end{aligned}
$$

where c and ζ are given as in Theorems 6.3 and 6.5. By denoting $\delta^k_{\mathbf{x},\mathbf{z},\mathbf{u}} := \|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \|\mathbf{u}^k - \mathbf{u}^{k+1}\| + \|\mathbf{z}^k - \mathbf{z}^{k+1}\|$, it follows that for all $k \geqslant K_2$

$$
\begin{aligned}
2\delta^k_{\mathbf{x},\mathbf{z},\mathbf{u}} &\leqslant 2\sqrt{\frac{3\zeta}{c}\left(\phi(\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) - \overline{\theta}) - \phi(\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}) - \overline{\theta})\right)\delta^{k-1}_{\mathbf{x},\mathbf{z},\mathbf{u}}} \\
&\leqslant \delta^{k-1}_{\mathbf{x},\mathbf{z},\mathbf{u}} + \frac{3\zeta}{c}\left(\phi(\Gamma(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k) - \overline{\theta}) - \phi(\Gamma(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}) - \overline{\theta})\right).
\end{aligned}
$$

So, the conclusion follows. ∎

Remark 6.12. If $\mathcal{S}$ is a semialgebraic set, and f, g, and h are semialgebraic functions (that is, their graphs can be written as a finite union or intersection of sets described by polynomial inequalities), then Γ and Π are also semialgebraic functions. So, they satisfy the KL property at every point of the domain of their subdifferential [3]. We also remark that, in this case, the desingularization function ϕ of the KL property takes the form of $\phi(s) = c\,s^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$. Here, θ is often called the corresponding KL exponent, see [33] for recent developments in estimating the KL exponents. Then, (local) convergence rate analysis of the algorithm can be deduced following the techniques used in [2] with the information of the KL exponents. For brevity, we omit the details here.

Remark 6.13. (Tightness of the convergence results of the conceptual FSPS algorithm) As seen in the proof of Theorem 6.2, every accumulation point $(\overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}, \overline{\mathbf{u}})$ of the sequence $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)$ need to fulfill the system of optimality conditions (6.9). Due to the existence of $\gamma > 0$ in the third inclusion of (6.9), we cannot anticipate $\overline{\mathbf{x}}$ as an exact limiting lifted stationary point of (1.1). To ensure that the accumulation point is an exact lifted stationary point, as the sequence $\{\gamma_k\}$ is non-increasing, without loss of generality, we can assume that one of the following two must hold:

(1.) $\gamma_k \equiv 0$ for all $k \geqslant K$, for some finite index K, or

(2.) $\gamma_k \downarrow 0$ as $k \to \infty$.

The following example illustrates that, in general, our convergence results are sharp.

EXAMPLE 6.14. *Consider problem (1.1) for* $\mathcal{S} = [0, 1]^2 \subseteq \mathbb{R}^2$*,* $A = K = I$ *where* I *is the identity mapping, and* $g, h, f : \mathbb{R}^2 \to \mathbb{R}$ *are given by* $g(\mathbf{x}) = \|\mathbf{x}\|_1$*,* $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 + \mathbf{e}^\top\mathbf{x} + \frac{1}{2}$ *and* $f(\mathbf{x}) = \mathbf{e}^\top\mathbf{x} + \frac{1}{2}$*, where* $\mathbf{e} = (1, 1)^\top$*. We consider two cases: (1.)* $\gamma_k \equiv 0$ *for all* k*. Let* $\beta = 1$*,* $\delta_k \equiv 1$*,* $\theta_0 := 2$ *and* $\mathbf{z}^0 = \mathbf{u}^0 = \mathbf{x}^0 := (1, 0)^\top$*. For the fourth update block in FSPS, we will choose* $\mathbf{z}^{k+1}$ *as the minimum norm solution. Then, FSPS generates a sequence* $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)$ *such that*

$$\mathbf{z}^k = \mathbf{u}^k = \mathbf{x}^k = \begin{cases} (0,1)^\top, & \text{if } k \text{ is odd,} \\ (1,0)^\top, & \text{if } k \text{ is even,} \end{cases} \qquad \mathbf{y}^k = \mathbf{e} \quad \text{and} \quad \theta_k = 2 \quad \forall k \geqslant 1.$$

One can verify that neither $(1, 0)^\top$ *nor* $(0, 1)^\top$ *is a limiting lifted stationary point of Example 6.14. Thus, the subsequential convergence to an exact limiting lifted stationary point cannot be guaranteed in this case.*

(2.) $\gamma_k = \frac{1}{k+1}$ *for all* k*. We can show that any accumulation point of the sequence generated by the FSPS may not be an exact limiting lifted stationary point, see Appendix A for details. Note that, in this case,* $\delta_k \equiv 1$*, which violates the choice in Theorem 4.3, where* $\delta_k = \chi\left(L_{\nabla h} + \frac{\sigma_A^2}{\gamma_k}\right) \to +\infty$*.*

7. Discussion on its variants with counterexamples guided. It is interesting to see when the basic algorithm FSPS can converge to an exact limiting lifted stationary point. Consider the conceptual algorithm FSPS with $\gamma_k \equiv 0$ reads for all $k \geqslant 0$:

$$(7.1) \qquad \begin{cases} \mathbf{y}^{k+1} & \in & \partial f(K\mathbf{x}^k) \\ \mathbf{x}^{k+1} & = & \mathrm{Proj}_{\mathcal{S}}\left(\mathbf{u}^k + \frac{\tilde{\theta}_k}{\delta_k}K^*\mathbf{y}^{k+1} - \frac{1}{\delta_k}\nabla h(\mathbf{x}^k) - \frac{1}{\delta_k}A^*\mathbf{z}^k\right), \\ \mathbf{u}^{k+1} & = & (1-\beta)\mathbf{u}^k + \beta\mathbf{x}^{k+1}, \\ \mathbf{z}^{k+1} & = & \arg\min_{\mathbf{z}}\left[g^*(\mathbf{z}) - \langle A\mathbf{x}^{k+1}, \mathbf{z}\rangle\right], \\ \tilde{\theta}_{k+1} & = & \dfrac{\tilde{\Psi}(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^{k+1}; \delta_k)}{f(K\mathbf{x}^{k+1})}, \end{cases}$$

where $\tilde{\Psi}(\mathbf{x}, \mathbf{z}, \mathbf{u}, \delta) := \langle \mathbf{z}, A\mathbf{x}\rangle - g^*(\mathbf{z}) + h(\mathbf{x}) + \iota_{\mathcal{S}}(\mathbf{x}) + \frac{\delta}{2}\|\mathbf{x} - \mathbf{u}\|^2$.[4]

[4]Note that $\gamma_k \equiv 0$. Then, the function Ψ in (4.1) reduces to $\tilde{\Psi}$.

Next, we show that the sequence $\{\mathbf{x}^k\}$ generated by (7.1) converges to an exact limiting lifted stationary point of (1.1) if g is ℓ-smooth.

THEOREM 7.1. *Suppose Assumption 3.1 holds, g is ℓ-smooth ($\ell > 0$) and essentially strictly convex, and one of the following conditions is fulfilled:*

(i) *f^* satisfies the calm condition over its effective domain and Γ satisfies KL property at every point of of $\{(\mathbf{x},\mathbf{y}) \in K_\varepsilon \times \mathrm{dom} f^* : \langle K\mathbf{x},\mathbf{y}\rangle - f^*(\mathbf{y}) > m/2\} \times \mathrm{int}(\mathrm{dom} g^*) \times \mathbb{R}^n$.*

(ii) *f is differentiable with Lipschitz continuous gradient over an open set containing $K(\mathcal{S})$ and Π satisfies KL property at every point of $K_\varepsilon \times \mathrm{int}(\mathrm{dom} g^*) \times \mathbb{R}^n$.*

Let $0 < \beta < 2$, $\chi > 1$, $\delta_k \equiv \delta := \chi(L_{\nabla h} + 2\ell\sigma_A^2)$ for $k \geqslant 0$, and $\{\boldsymbol{W}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)\}$ be the sequence generated by (7.1). Then, $\sum_k \left(\|\mathbf{x}^k - \mathbf{x}^{k+1}\| + \|\mathbf{u}^k - \mathbf{u}^{k+1}\| + \|\mathbf{z}^k - \mathbf{z}^{k+1}\|\right) < +\infty$, and $\{\mathbf{x}^k\}$ converges to a limiting lifted stationary point of (1.1).

Proof. First, analogous to the proof to (6.1), one can show that for all $k \geqslant 0$

$$\begin{aligned}
&\tilde{\Psi}(\mathbf{x}^{k+1}, \mathbf{z}^k, \mathbf{u}^k, \delta) + \tilde{\theta}_k \left[f(K\mathbf{x}^k) - (\langle K\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\rangle - f^*(\mathbf{y}^{k+1}))\right] \\
&\leqslant \tilde{\Psi}(\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k, \delta) - \frac{\delta - L_{\nabla h}}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2.
\end{aligned}$$

Second, using the optimality condition of $\mathbf{z}^{k+1}$ in (7.1), it yields

$$\begin{aligned}
\langle A\mathbf{x}^{k+1}, \mathbf{z}^{k+1}\rangle - g^*(\mathbf{z}^{k+1}) \leqslant{}& \langle A\mathbf{x}^{k+1}, \mathbf{z}^k\rangle - g^*(\mathbf{z}^k) + \langle \mathbf{z}^{k+1} - \mathbf{z}^k, A\mathbf{x}^{k+1} - A\mathbf{x}^k\rangle \\
&- \frac{1}{2\ell}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2.
\end{aligned}$$

So, $\tilde{\Psi}(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{u}^k, \delta) \leqslant \tilde{\Psi}(\mathbf{x}^{k+1}, \mathbf{z}^k, \mathbf{u}^k, \delta) + \ell\sigma_A^2\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - \frac{1}{4\ell}\|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2$. The remaining proofs are similar to Theorems 6.1, 6.3 to establish the descent property of the merit functions, and show the subsequential convergence by following the proof routines in Theorem 6.2, and the global convergence routines in Theorem 6.11, thus omitted here. ∎

Another interesting question is to see *what happens if we replace the updating step of $\mathbf{z}^{k+1}$ in the conceptual FSPS (7.1) with the following:*

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \left\{ g^*(\mathbf{z}) - \langle A\mathbf{x}^{k+1}, \mathbf{z}\rangle + \frac{\alpha_k}{2}\|\mathbf{z} - \mathbf{z}^k\|^2 \right\}, \tag{7.2}$$

where $\alpha_k > 0$. We call this variant as P-FSPS. If g is nonsmooth, then P-FSPS can also exhibit a cycling phenomenon, as illustrated by the following example.

EXAMPLE 7.2. *Consider the problem (1.1) for $\mathcal{S} = [0,1]^2 \subseteq \mathbb{R}^2$, $A = \frac{1}{2}I$, $K = 2I$ where I is the identity mapping, and $g, h, f : \mathbb{R}^2 \to \mathbb{R}$ given by $g(\mathbf{x}) = \|\mathbf{x}\|_1$, $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 + \frac{1}{2}\mathbf{e}^\top\mathbf{x} + \frac{3}{2}$ and $f(\mathbf{x}) = \frac{1}{2}\mathbf{e}^\top\mathbf{x} + 1$. Let $\beta = 1$, $\delta_k \equiv \delta = \frac{1}{2}$ for $k \geqslant 0$, $\tilde{\theta}_0 := \frac{3}{2}$ and $\mathbf{z}^0 = (1,1)^\top$, $\mathbf{u}^0 = \mathbf{x}^0 := (1,0)^\top$. For any $\alpha_k > 0$, P-FSPS generates a sequence $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)$ such that for all $k \geqslant 0$*

$$\mathbf{u}^k = \mathbf{x}^k = \begin{cases} (0,1)^\top, & \text{if } k \text{ is odd,} \\ (1,0)^\top, & \text{if } k \text{ is even,} \end{cases} \qquad \mathbf{y}^k = \frac{1}{2}\mathbf{e}, \ \ \mathbf{z}^k = \mathbf{e} \quad \text{and} \quad \tilde{\theta}_k = \frac{3}{2}.$$

Direct verification shows that neither $(1,0)^\top$ nor $(0,1)^\top$ is a limiting lifted stationary point of Example 7.2.

8. Numerical results. We present numerical results to demonstrate the efficacy of the proposed algorithmic framework. All algorithms are implemented using MATLAB R2016a and executed on a desktop running Windows 10 equipped with an Intel Core i7-7600U CPU processor (2.80GHz) and 16GB of memory.

8.1. Implementation details. To allow for larger step sizes δ_k and mitigate the dependence on the unknown parameter $L_{\nabla h}$, we propose practical variants of Algorithms 4.1 and 5.1 by incorporating a nonmonotone line search strategy [32]. These variants are referred to as S-FSPS-nls and Adaptive FSPS-nls, respectively. Due to space limitations, we only present the details of Adaptive FSPS-nls in Algorithm 8.1.

ALGORITHM 8.1 (Adaptive FSPS algorithm with *nonmonotone line search*). *Let* $0 < \beta < 2$, $\chi > 1$, $0 < q < 1$, $\delta_0 > 0$, $\gamma_0 = 1$, *and* $\varepsilon > 0$, $\eta > 1$, $0 < \mu < 1$, $c > 0$, $T, \ell, t \in \mathbb{N}$. *Let* $(\mathbf{x}^0, \mathbf{u}^0)$ *be a given starting point. We use* MaxIt *to indicate the maximal number of iterations.*

For $k = 0 : \text{MaxIt}$ *do*
  Set $\gamma_{k,0} := \gamma_k$.
  For $j = 0 : \ell - 1$ *do*
    Set $\gamma_{k,j} := \gamma_{k,0} q^j$.
    Set $\mathbf{z}^{k+1,j} := \text{Prox}_{g^*/\gamma_{k,j}} \left(\dfrac{A\mathbf{x}^k}{\gamma_{k,j}} \right)$.
    If $\theta_{k+1} := \dfrac{\Psi(\mathbf{x}^k, \mathbf{z}^{k+1,j}, \mathbf{u}^k, \delta_k, \gamma_{k,j})}{f(K\mathbf{x}^k)} > 0$, *then*
      Update $\gamma_k := \gamma_{k,j}$, $\mathbf{z}^{k+1} := \mathbf{z}^{k+1,j}$.
      Break
    End If
  End For
  Set $\delta_{k,0} := \chi(L_{\nabla h} + \dfrac{2\sigma_A^2}{\gamma_k})$.
  Choose $\mathbf{y}^{k+1} \in \partial f(K\mathbf{x}^k)$.
  Set $\mathbf{d}^{k+1} := \theta_{k+1} K^* \mathbf{y}^{k+1} - \nabla h(\mathbf{x}^k) - A^* \mathbf{z}^{k+1}$.
  For $s = 0 : t - 1$
    Set $\delta_{k,s} := \mu \eta^s \delta_{k,0}$.
    Set $\tilde{\mathbf{x}}^{k+1} := \text{Proj}_{\mathcal{S}} \left(\mathbf{u}^k + \dfrac{\mathbf{d}^{k+1}}{\delta_{k,s}} \right)$.
    If $F(\tilde{\mathbf{x}}^{k+1}) \leqslant \max\limits_{[k-T]_+ \leqslant j \leqslant k} F(\mathbf{x}^j) - \dfrac{c}{2} ||\mathbf{x}^k - \tilde{\mathbf{x}}^{k+1}||^2$, *then*
      Update $\mathbf{x}^{k+1} := \tilde{\mathbf{x}}^{k+1}$.
      Break
    End If.
  End For.
  Update $\mathbf{u}^{k+1} := \mathbf{u}^k - \beta(\mathbf{u}^k - \mathbf{x}^{k+1})$.
  Update $\gamma_{k+1} := \gamma_k$, $\delta_{k+1} := \delta_k$.
  If $\|\mathbf{z}^{k+1}\| > \min \left(\dfrac{\varepsilon}{\gamma_k}, \sqrt{\dfrac{2\varepsilon}{\gamma_k}} \right)$, *then*
    Update $\gamma_{k+1} := \gamma_k q$, $\delta_{k+1} := \chi(L_{\nabla h} + \dfrac{2\sigma_A^2}{\gamma_{k+1}})$.
  End If
End For

8.2. Limited-angle CT reconstruction. We solved the problem (1.2) by comparing S-FSPS-nls, Adaptive FSPS-nls with the Extrapolated Proximal Subgradient algorithm (e-PSA) from [8], and the Proximity Gradient Subgradient algorithm with Backtracked Extrapolation (PGSA_BE) from [21]. We set $\tau = 0.1$ and $p = 2$ in (1.2) throughout the numerical tests. Each algorithm was initialized with the zero vector (with a safeguard mechanism of computing the denominator of (1.1) via $\max(\|\nabla \mathbf{x}\|_2, \text{eps})$) and used the same stopping criterion defined by:

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|}{\max\{\text{eps}, \|\mathbf{x}^k\|\}} < 10^{-6} \ \text{ or } \ k > \text{MaxIt}, \tag{8.1}$$

where eps represents the machine precision. We also adopted a *two-stage approach with a warm start strategy*, where the last iterate of the first stage served as the initial point for the second stage. The warm start will be beneficial for solving the imaging processing problem, but it also requires careful parameter tuning for two phases.

When implementing Adaptive FSPS-nls, we set $f(\mathbf{x}) := \|\mathbf{x}\|_2$, $A = K = \nabla$, and followed Remark 6.8 by setting $g(\mathbf{x}) := \tau\|\mathbf{x}\|_1 + \frac{s}{2}\|\mathbf{x}\|^2$ and $h(\mathbf{x}) := \frac{1}{2}\|P\mathbf{x} - \mathbf{f}\|^2 - \frac{s}{2}\|A\mathbf{x}\|^2$ where $s = 0.1$. The superscripts $^{(1)}$ and $^{(2)}$ represent the stage one and stage two, respectively. The parameter settings were $\beta^{(1)} = 1.1$, $\beta^{(2)} = 1.45$, $\chi^{(1)} = 1.1$, $\chi^{(2)} = 1.001$, $\mu^{(1)} = \mu^{(2)} = 0.4$, $\eta^{(1)} = \eta^{(2)} = 1.5$, $q^{(1)} = q^{(2)} = 0.999$, $T^{(1)} = T^{(2)} = 5$, $c^{(1)} = c^{(2)} = $ 1e-4, $t^{(1)} = t^{(2)} = 250$, $\ell^{(1)} = \ell^{(2)} = 1000$, $\text{MaxIt}^{(1)} = 50$, $\text{MaxIt}^{(2)} = 5000$, and $\varepsilon^{(1)} = \varepsilon^{(2)} = 1e-6$. To implement S-FSPS-nls, we use the same parameters as those in Adaptive FSPS-nls, except that we set $\gamma_k^{(1)} = \gamma_k^{(2)} = \frac{1}{k^{0.05}}$ and $\chi^{(1)} = \chi^{(2)} = 2$.

When applying PGSA_BE (Algorithm 1 in [21]), we set $f(\mathbf{x}) := \tau\|\nabla\mathbf{x}\|_1$, $h(\mathbf{x}) := \frac{1}{2}\|P\mathbf{x} - \mathbf{f}\|^2$, $g(\mathbf{x}) := \|\nabla\mathbf{x}\|_2$. The inner loop amounts to solving in each iteration

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\mathcal{B}} \left[\tau\|\nabla\mathbf{x}\|_1 + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{q}^k\|^2\right], \tag{8.2}$$

with $\mathbf{q}^k = \mathbf{u}^{k+1} - \alpha P^*(P\mathbf{u}^{k+1} - \mathbf{f}) + \alpha c_k \frac{\nabla^*(\nabla\mathbf{x}^k)}{\|\nabla\mathbf{x}^k\|_2}$, $\mathbf{u}^{k+1} = \mathbf{x}^k + \beta_k(\mathbf{x}^k - \mathbf{x}^{k-1})$ and $c_k = \frac{f(\mathbf{x}^k) + h(\mathbf{x}^k)}{g(\mathbf{x}^k)}$. We applied ADMM to (8.2) by introducing $\nabla\mathbf{x} = \mathbf{y}$ and $\mathbf{x} = \mathbf{z}$, and with ρ_1 and ρ_2 being the penalty parameters. For the outer loop parameters we set $\ell^{(1)} = \ell^{(2)} = 0, \beta_k^{(1)} = \beta_k^{(2)} \equiv 0.1, \alpha^{(1)} = 0.0015, \alpha^{(2)} = 0.001$, $\varepsilon^{(1)} = \varepsilon^{(2)} =$ 1e-3 (in the backtracking condition), $\text{MaxIt}^{(1)} = 50$, $\text{MaxIt}^{(2)} = 5000$. For the inner loop parameters we set $\text{Inner\_tol}^{(1)} = \text{Inner\_tol}^{(2)} =$ 1e-6, and $\text{Inner\_MaxIt}^{(1)} = 1000$, $\text{Inner\_MaxIt}^{(2)} = 200$, $\rho_1^{(1)} = \rho_1^{(2)} =$ 1e-4, $\rho_2^{(1)} = \rho_2^{(2)} =$ 1e-2.

When applying e-PSA (Algorithm 4.1 in [8]), we set $f^n(\mathbf{x}) := \tau\|\nabla\mathbf{x}\|_1$, $f^s(\mathbf{x}) := \frac{1}{2}\|P\mathbf{x} - \mathbf{f}\|^2$, and $g(\mathbf{x}) := \|\nabla\mathbf{x}\|_2$. Due to the absence of boundedness condition (BC), $\overline{\mu} = \overline{\kappa} = 0$, and so, $\kappa_k = \mu_k = 0$ and $\mathbf{u}^k = \mathbf{v}^k = \mathbf{x}^k$ for all $k \geqslant 0$. The inner loop amounts to solving in each iteration

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\mathcal{B}} \left[\tau\|\nabla\mathbf{x}\|_1 + \frac{1}{2\tau_k}\|\mathbf{x} - \mathbf{p}^k\|^2 + \frac{\ell}{2}\|\mathbf{x} - \mathbf{x}^k\|^2\right], \tag{8.3}$$

with $\mathbf{p}^k = \mathbf{x}^k + \tau_k\theta_k \frac{\nabla^*\nabla\mathbf{x}^k}{\|\nabla\mathbf{x}^k\|} - \tau_k P^*(P\mathbf{x}^k - \mathbf{f})$ and $\theta_k = \frac{f^n(\mathbf{x}^k) + f^s(\mathbf{x}^k)}{g(\mathbf{x}^k)}$. We solved (8.3) also via ADMM, with ρ_1 and ρ_2 being the penalty parameters. For the outer loop parameters we set $\beta^{(1)} = \beta^{(2)} = 0$, $\ell^{(1)} = \ell^{(2)} = \|P\mathbf{x}^{\text{true}}\|/\|\mathbf{x}^{\text{true}}\|$, $\tau_k^{(1)} = \tau_k^{(2)} \equiv 760$, and $\text{MaxIt}^{(1)} = 50$, $\text{MaxIt}^{(2)} = 5000$. For the inner loop parameters we set $\text{Inner\_tol}^{(1)} = \text{Inner\_tol}^{(2)} =$ 1e-6, and $\text{Inner\_MaxIt}^{(1)} = 1000$, $\text{Inner\_MaxIt}^{(2)} = 200$, $\rho_1^{(1)} = \rho_1^{(2)} = \rho_2^{(1)} = \rho_2^{(2)} =$ 1e-2.

We assessed performance based on two metrics: the root mean squared error (RMSE) [29] and the overall structural similarity index (SSIM) [31]. We conducted tests on parallel beam CT reconstruction of the Shepp-Logan phantom using projection ranges of 90°, 120°, and 150°. We evaluated both noiseless and noisy scenarios, where the Gaussian noise had a zero mean and standard deviations (σ) of 0.001 and 0.005. The performance of the three algorithms is summarized in Table 1. The results indicate that S-FSPS-nls and Adaptive FSPS-nls outperform the recently introduced double-loop algorithms, PGSA_BE and e-PSA, in terms of SSIM, RMSE, and CPU time (in seconds).

When comparing S-FSPS-nls and Adaptive FSPS-nls, we observe that S-FSPS-nls requires less CPU time, primarily because Adaptive FSPS-nls needs to perform backtracking to ensure the non-negativity of θ_k, whereas S-FSPS-nls does not necessarily require this. However, S-FSPS-nls achieves slightly lower SSIM values in some cases compared to Adaptive FSPS-nls.

8.3. Robust Sharpe ratio type minimization problem. We tested Adaptive FSPS-nls also on the robust sharp-ratio minimization problem (1.3), and compared it with PGSA_BE, e-PSA and the Dinkelbach's method with Surrogation (DLS) [14, Algorithm 7.2.7]. The data $((r_i)_{i=1}^{m_1}, (\mathbf{a}_i)_{i=1}^{m_1}, (C_i)_{i=1}^{m_2})$ were generated as follows: (1) each vector $\mathbf{a}_i$ was generated such that each entry is drawn from a uniform distribution over the interval $[0, 1]$; (2) r_i was set to be greater than $\|\mathbf{a}_i\|_\infty$; (3) each matrix C_i was generated such that each eigenvalue conforms to a uniform distribution over the interval $[10^{-3}, 1 + 10^{-3}]$.

We measured the performance in terms of the objective value `obj`, the infeasibility `infea` $:= \|\max(-\mathbf{x}, 0)\|_1 + |\|\mathbf{x}\|_1 - 1|$, and the lifted stationarity residual

$$\texttt{stat} := \operatorname{dist}\big(\mathbf{0},\ (A^*\partial g(A\mathbf{x}) + \nabla h(\mathbf{x}) + \partial\iota_{\mathcal{S}}(\mathbf{x}))\, f(K\mathbf{x}) - (g(A\mathbf{x}) + h(\mathbf{x}))K^*\partial f(K\mathbf{x})\big).$$

All metrics are evaluated at the last iterate. We also used (8.1) as a stopping criterion.

When implementing Adaptive FSPS-nls, we set f, $\mathbf{r}$, A, and K as in Section 1, and set $g(\mathbf{x}) := \|\mathbf{r} - \mathbf{x}\|_\infty + \frac{s}{2}\|\mathbf{x}\|^2$ and $h(\mathbf{x}) := -\frac{s}{2}\|A\mathbf{x}\|^2$ with $s = 0.01$, by following Remark 6.8. We set the algorithm parameters as $\ell := 100$, $L_{\nabla h} := s\|A^*A\|$, $\chi := 1.1$, $\eta := 1.15$, $q := 0.999$, $\mu := 0.005$, $c := 10^{-4}$, $T := 5$, $\delta_0 := \chi(L_{\nabla h} + 2\sigma_A^2)$, $t := 250$, MaxIt $:= 500$, $\varepsilon := 1e-8$, and $\beta := 1.6$. To implement S-FSPS-nls, we use the same parameters as those in Adaptive FSPS-nls, except that we set $\gamma_k = \frac{1}{k^{1/3}}$ and $\chi = 1.5$.

When implementing PGSA_BE, we defined $f(\mathbf{x}) := \max_{1\leqslant i\leqslant m_1}\{r_i - \mathbf{a}_i^\top\mathbf{x}\}$, $h(\mathbf{x}) := 0$, and $g(\mathbf{x}) := \max_{1\leqslant i\leqslant m_2}\mathbf{x}^\top C_i\mathbf{x}$. The inner loop amounts to solving in each iteration

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\Delta}\left[\max_{1\leqslant i\leqslant m_1}\{r_i - \mathbf{a}_i^\top\mathbf{x}\} + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{u}^{k+1} - \alpha c_k\mathbf{y}^k\|^2\right], \tag{8.4}$$

with $\mathbf{y}^k \in \partial g(\cdot)(\mathbf{x}^k)$ and $\mathbf{u}^{k+1} = \mathbf{x}^k + \beta_k(\mathbf{x}^k - \mathbf{x}^{k-1})$. The inner loops of both e-PSA and DLS amounts to solving in each iteration a similar problem as (8.4).

For fair comparisons, we solved the inner loop subproblems for all these double-loop algorithms via ADMM. We used MaxIt $= 500$ for all these test algorithms. In addition, we used for e-PSA as outer loop parameters $\beta = 0$, $\tau_k \equiv 0.5$ and as inner loop parameters $\rho_1 = 0.1$, $\rho_2 = 0.1$; we used for PGSA_BE as outer loop parameters $\beta_k \equiv 0.5$, $\alpha = 0.5$, $\varepsilon = 1 \times 10^{-3}$ (in the backtracking) and as inner loop parameters $\rho_1 = 0.5$, $\rho_2 = 0.5$, and we used for DLS as inner loop parameters $\rho_1 = \rho_2 = 0.5$. We conducted numerical tests by setting (n, m_1, m_2) to $(100, 5, 20)$, $(100, 20, 5)$, $(100, 20, 20)$, $(400, 20, 10)$, $(400, 10, 20)$, and $(400, 20, 20)$. We performed

TABLE 1
Parallel beam CT reconstruction of the Shepp-Logan phantom for different projection ranges.

SD of noise	Range	e-PSA [8]			PGSA_BE [21]			Time Ratios	
		SSIM	RMSE	T_1	SSIM	RMSE	T_2	T_4 / T_1	T_4 / T_2
0	90°	0.9946	3.38e-04	6.32e+02	0.9958	3.65e-04	6.68e+02	1.40%	1.32%
	120°	0.9970	1.98e-04	5.98e+02	0.9999	3.37e-05	6.30e+02	1.67%	1.58%
	150°	0.9991	1.06e-04	4.39e+02	1.0000	1.95e-05	3.28e+02	2.32%	3.12%
0.001	90°	0.9946	3.38e-04	6.33e+02	0.9955	3.76e-04	6.68e+02	1.40%	1.32%
	120°	0.9970	1.97e-04	4.65e+02	0.9999	3.55e-05	5.62e+02	1.82%	1.51%
	150°	0.9991	1.08e-04	4.35e+02	1.0000	2.27e-05	2.27e+02	1.92%	3.68%
0.005	90°	0.9944	3.21e-04	5.74e+02	0.9865	6.22e-04	4.34e+02	1.58%	2.09%
	120°	0.9967	2.15e-04	6.58e+02	0.9994	9.83e-05	4.42e+02	1.45%	2.05%
	150°	0.9984	1.54e-04	5.87e+02	0.9996	8.35e-05	4.05e+02	1.72%	2.49%

SD of noise	Range	S-FSPS-nls			Adaptive FSPS-nls			Time Ratios	
		SSIM	RMSE	T_3	SSIM	RMSE	T_4	T_3 / T_4	T_4 / T_3
0	90°	0.9999	3.83e-05	6.06e+00	1.0000	2.39e-05	8.85e+00	68.47%	146.04%
	120°	1.0000	2.60e-05	5.48e+00	1.0000	1.21e-05	9.96e+00	55.02%	181.75%
	150°	1.0000	1.76e-05	3.21e+00	1.0000	8.47e-06	1.22e+01	26.31%	327.56%
0.001	90°	0.9999	3.83e-05	6.06e+00	0.9999	2.94e-05	8.85e+00	68.47%	146.03%
	120°	0.9999	2.85e-05	4.32e+00	1.0000	1.49e-05	8.48e+00	50.94%	196.30%
	150°	1.0000	2.09e-05	2.56e+00	1.0000	1.10e-05	8.36e+00	30.62%	327.56%
0.005	90°	0.9991	1.16e-04	6.12e+00	0.9992	1.12e-04	9.05e+00	67.62%	147.88%
	120°	0.9995	9.09e-05	8.72e+00	0.9996	8.65e-05	9.33e+00	93.46%	159.12%
	150°	0.9996	8.36e-05	3.81e+00	0.9996	8.13e-05	1.01e+01	37.72%	264.30%

Table 2
S-FSPS-nls and Adpative FSPS-nls versus double-loop algorithms for robust sharp-ratio Problem

(n, m_1, m_2)		FSPS-nls	S-FSPS	e-PSG	PGSA_BE	DLS
(100, 5, 20)	`obj`	1.52e+00	1.52e+00	1.56e+00	1.59e+00	1.54e+00
	`infea`	4.22e-09	3.56e-09	4.11e-07	3.98e-07	3.98e-05
	`stat`	2.53e-07	2.53e-07	2.21e-07	2.97e-07	1.50e-07
	`CPU`	2.61e-02	2.74e-02	5.58e-02	5.17e-02	3.27e-01
(100, 20, 5)	`obj`	1.76e+00	1.76e+00	1.79e+00	1.79e+00	1.75e+00
	`infea`	4.35e-09	4.70e-09	3.98e-07	4.95e-07	3.20e-05
	`stat`	3.20e-07	3.20e-07	5.04e-07	5.45e-07	3.88e-07
	`CPU`	2.10e-02	2.60e-02	6.29e-02	5.34e-02	9.55e-01
(100, 20, 20)	`obj`	1.68e+00	1.68e+00	1.67e+00	1.69e+00	1.75e+00
	`infea`	3.12e-09	4.47e-09	3.69e-07	5.82e-07	3.18e-05
	`stat`	4.07e-07	4.07e-07	4.16e-07	5.04e-07	2.62e-03
	`CPU`	2.66e-02	3.28e-02	7.02e-02	6.32e-02	2.59e+00
(400, 20, 10)	`obj`	1.88e+00	1.88e+00	1.89e+00	1.88e+00	2.09e+00
	`infea`	2.84e-09	2.57e-09	3.90e-07	4.29e-07	5.10e-05
	`stat`	6.30e-05	6.30e-05	6.28e-05	6.02e-05	5.52e-03
	`CPU`	5.41e-01	5.52e-01	8.26e-01	7.60e-01	1.03e+02
(400, 10, 20)	`obj`	1.70e+00	1.70e+00	1.78e+00	1.78e+00	1.89e+00
	`infea`	4.13e-09	4.29e-09	4.03e-07	4.32e-07	5.01e-05
	`stat`	3.02e-05	3.02e-05	2.75e-05	2.88e-05	1.36e-02
	`CPU`	1.33e+00	1.25e+00	1.36e+00	1.43e+00	1.50e+02
(400, 20, 20)	`obj`	1.84e+00	1.84e+00	1.85e+00	1.82e+00	1.93e+00
	`infea`	4.04e-09	3.42e-09	4.87e-07	3.90e-07	4.41e-05
	`stat`	4.32e-05	4.32e-05	4.25e-05	3.86e-05	4.04e-04
	`CPU`	6.24e-01	6.17e-01	7.34e-01	7.71e-01	1.01e+02

50 trials for each configuration. The average values of the considered performance metrics, along with the CPU time (in seconds), are reported in Table 2.

As observed, S-FSPS-nls and Adaptive FSPS-nls outperform e-PSA, PGSA_BE, and DLS by achieving smaller `infeas`, comparable `stat`, and `obj` values, while requiring less computation time. Their performance is nearly identical, mainly due to the choice $\varepsilon = 1 \times 10^{-8}$.

9. Conclusions. The paper focuses on a class of structural fractional programs characterized by linear compositions with nonsmooth functions in both the numerator and denominator. We develop a proximal subgradient algorithm framework with two versions (S-FSPS and Adaptive FSPS) to overcome the challenges in computing the proximal point of the linear composition with the nonsmooth component in the numerator. Our contributions include establishing the subsequential convergence to an exact lifted stationary point for the S-FSPS while establishing the global convergence of Adaptive FSPS toward an approximate lifted stationary point under the KL property, without imposing full-row rank assumptions. We explain the rationale behind the convergence to an approximate lifted stationary point of the Adaptive FSPS and construct counterexamples to show that pursuing an exact solution in the adaptive version might lead to divergence. Finally, we demonstrate the superiority of these practical versions of the newly proposed algorithms over the existing state-of-the-art methods for two concrete applications.

Appendix A. Accumulation points of the sequence generated by the FSPS may fail to be a lifted stationary point when $\gamma_k \downarrow 0$.

Consider the counter-example of Example 6.14. For $\beta = 1$, $\gamma_k := \frac{1}{k+1}$, $\delta_k \equiv 1$, $\theta_0 := 2$, and $\mathbf{z}^0 = \mathbf{u}^0 = \mathbf{x}^0 := (1,0)^\top$, FSPS generates a sequence $\{(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)\}$. The sequence $\{\mathbf{x}^k\}$ has two accumulation points: $(1,0)^\top$ and $(0,1)^\top$. Indeed, neither $(1,0)^\top$ nor $(0,1)^\top$ is a limiting lifted stationary point of Example 6.14. We provide the details in the following lemma.

LEMMA A.1. *Let the sequence $\{(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \mathbf{u}^k)\}$ be generated by FSPS (4.1) for solving Example 6.14 with $\beta = 1$, $\gamma_k := \frac{1}{k+1}$, $\delta_k \equiv 1$, $\theta_0 := 2$, and $\mathbf{z}^0 = \mathbf{u}^0 = \mathbf{x}^0 := (1,0)^\top$. Then, we have $\mathbf{y}^k = \mathbf{e}$ for all $k \geqslant 1$ and the following statements hold:*

(i)

$$\textcircled{1}_k: \ \mathbf{z}^k = \begin{cases} (0,1)^\top, & \text{if } k \text{ is odd}, \\ (1,0)^\top, & \text{if } k \text{ is even}, \end{cases}$$

(ii)

$$\textcircled{2}_k: \ \mathbf{x}^k = \begin{cases} (0, \ \theta_{k-1} - 1)^\top, & \text{if } k \text{ is odd}, \\ (\theta_{k-1} - 1, \ 0)^\top, & \text{if } k \text{ is even}. \end{cases}$$

where θ_k is given by

$$\text{(A.1)} \quad \textcircled{3}_k: \ \theta_k = \frac{\theta_{k-1} - 1 + (0.5 * (\theta_{k-1} - 1)^2 + (\theta_{k-1} - 1) + 0.5) - \frac{1}{2k}}{\theta_{k-1} - 1/2}.$$

(iii) $\lim_{k\to\infty} \theta_k = 2$*, and hence, the sequence $\{\mathbf{x}^k\}$ has two accumulation points: $(1,0)^\top$ and $(0,1)^\top$.*

Proof. First, we define a sequence $\{b_k\}$ via the following recurrence formula: $b_0 = 1$ and, for all $k \geqslant 0$, $b_{k+1} = \frac{\frac{1}{2}b_k^2 + b_k - \frac{1}{2(k+1)}}{b_k + 1/2}$. For this sequence, we first use *mathematical induction* to see that

$$\text{(A.2)} \qquad 1/2 < b_k < 1, \ \forall k \geqslant 1.$$

By direct calculations, we see that $b_1 = 2/3$, $b_2 = 23/42$, $b_3 = 936.5/1848$ and $b_4 = \frac{936.5^2/(1848*2)+705.5}{1860.5}$. Thus, (A.2) holds with $k = 1,2,3,4$. Suppose that (A.2) holds with $k = k_0$ for some $k_0 \geqslant 4$, that is, $1/2 < b_{k_0} < 1$. We now show that (A.2) holds with $k = k_0 + 1$. To see this, we first note that $b_{k_0+1} = \frac{1}{2}b_{k_0} + \frac{3}{4} - \frac{\frac{3}{8} + \frac{1}{2(k_0+1)}}{b_{k_0}+1/2}$. Define a one-variable function $f(x) := \frac{1}{2}x + \frac{3}{4} - \frac{\frac{3}{8} + \frac{1}{2(k_0+1)}}{x+1/2}$. Direct verification shows that f is an increasing function. So, $b_{k_0+1} = f(b_{k_0}) \geqslant f(1/2) = 5/8 - \frac{1}{2(k_0+1)} > 1/2$, where the last strict inequality holds as $k_0 \geqslant 4$. Moreover, as $b_{k_0} < 1$, $b_{k_0+1} = f(b_{k_0}) \leqslant f(1) < 1$. Thus, (A.2) holds.

Next, we show the main results of this lemma. Clearly, from the definition of f and the construction, $\mathbf{y}^k = \mathbf{e}$ for all $k \geqslant 1$.

[Proof of (i) & (ii)] We use mathematical induction to verify $\textcircled{1}_k$, $\textcircled{2}_k$ and $\textcircled{3}_k$ hold for all $k \geqslant 1$. A direct verification shows that the statements of $\textcircled{1}_k$, $\textcircled{2}_k$ and $\textcircled{3}_k$ hold for $k = 1,2$; Suppose that $\textcircled{1}_k$, $\textcircled{2}_k$ and $\textcircled{3}_k$ hold for $k \leqslant k_0$ with $k_0 \geqslant 2$. Using (A.2) with $b_{k_0} = \theta_{k_0} - 1$, we see that $3/2 < \theta_{k_0} < 2$. Using the update formula of $\mathbf{x}^{k+1}$ in (4.1), a direct verification shows that $\textcircled{2}_k$ holds with $k = k_0 + 1$. Note from the update formula of $\mathbf{z}^{k+1}$ in (4.1) that $\mathbf{z}^{k_0+1} := \mathrm{Proj}_{\mathcal{B}_1^\infty}(\frac{(\theta_{k_0}-1, \ 0)^\top}{1/(k_0+1)})$ or

$\mathbf{z}^{k_0+1} := \mathrm{Proj}_{\mathcal{B}_1^\infty}(\frac{(0,\ \theta_{k_0}-1)^\top}{1/(k_0+1)})$ where $\mathcal{B}_1^\infty$ is the unit ball defined by the ℓ_∞-norm. Since $3/2 < \theta_{k_0} < 2$, we have $\textcircled{1}_k$ holds with $k = k_0 + 1$. Finally, using the update formula of θ_{k+1} in (4.1), $\textcircled{3}_k$ with $k = k_0 + 1$ also follows.

[Proof of (iii)] To see (iii), we first establish that $b_{k+1} \geqslant b_k$ when $k \geqslant 4$. From the definition of the sequence $\{b_k\}$, this is equivalent to

$$\frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{k+1}}\right) \leqslant b_k \leqslant \frac{1}{2}\left(1 + \sqrt{1 - \frac{4}{k+1}}\right). \tag{A.3}$$

Clearly, (A.3) is true with $k = 4$ by direct computation. Suppose now (A.3) holds with $k = k_0$ with $k_0 \geqslant 4$. We now show that (A.3) holds with $k = k_0 + 1$, that is,

$$\frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{k_0+2}}\right) \overset{(\clubsuit)}{\leqslant} b_{k_0+1} \overset{(\spadesuit)}{\leqslant} \frac{1}{2}\left(1 + \sqrt{1 - \frac{4}{k_0+2}}\right).$$

For $(\clubsuit)$, it holds obviously due to $b_k > 1/2$ for all $k \geqslant 1$. To prove $(\spadesuit)$, recall the one-variable f defined as above. We have $b_{k_0+1} = f(b_{k_0}) \leqslant f(\frac{1+\sqrt{1-\frac{4}{k_0+1}}}{2})$, where the last inequality follows by the induction hypothesis and the fact that f is increasing. Thus, it remains to show that $f(\frac{1+\sqrt{1-\frac{4}{k_0+1}}}{2}) \leqslant \frac{1+\sqrt{1-\frac{4}{k_0+2}}}{2}$. By letting $\delta := \frac{1}{k_0+1}$, $\kappa := \frac{1}{k_0+2}$ and noting that $\kappa \leqslant \delta$. Let $c := \sqrt{1-4\delta}$ and $d := \sqrt{1-4\kappa}$. Thus we have $\frac{c^2}{4} \leqslant \frac{1}{4}cd + \frac{d}{2} - \frac{c}{2}$. Consequently, $1 + \frac{\sqrt{1-4\delta}}{4} - \frac{\frac{3}{8}+\frac{\delta}{2}}{1+\frac{1}{2}\sqrt{1-4\delta}} \leqslant \frac{1}{2} + \frac{1}{2}\sqrt{1-4\kappa}$. With some elementary calculations, it leads to $f(\frac{1+\sqrt{1-\frac{4}{k_0+1}}}{2}) \leqslant \frac{1+\sqrt{1-\frac{4}{k_0+2}}}{2}$.

Therefore, the sequence of $\{b_k\}$ is monotone and bounded, thus $\lim_{k\to+\infty} b_k$ exists. Consequently, $\lim_{k\to+\infty} \theta_k$ exists, and $\lim_{k\to+\infty} \theta_k = 2$. ∎

Acknowledgments: The authors would also like to thank the anonymous reviewers and the associate editor for their constructive comments and suggestions.

REFERENCES

[1] W. van Ackooij, P. Pérez-Aros, C. Soto, E. Vilches, *Inner Moreau Envelope of Nonsmooth Conic Chance-Constrained Optimization Problems,* Mathematics of Operations Research 49 (2023), pp. 1419-1451.

[2] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features,* Mathematical Programming, 116(2009), pp. 5–16.

[3] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming 137 (2013), pp. 91–129.

[4] H. H. Bauschke and P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, second edition, Springer, 2017.

[5] W. Bian and X. Chen, *A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty, SIAM Journal on Numerical Analysis*, vol. 58, no. 1, pp. 858–883, 2020.

[6] J. Bolte, A. Daniilidis, and A. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17 (2007), pp. 1205–1223.

[7] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming 146 (2014), pp. 459-494.

[8] R. I. Boţ, M. Dao, and G. Li, *Extrapolated proximal subgradient algorithms for nonconvex and nonsmooth fractional programs*, Mathematics of Operations Research 47 (2022), pp. 2415–2443.

[9] R. I. Boţ, M. Dao, G. Li, *Inertial proximal block coordinate method for a class of nonsmooth sum-of-ratios optimization problems*, SIAM Journal on Optimization 33 (2023), pp. 361–393.

[10] R. I. Boţ, E.-R. Csetnek, *Proximal-gradient algorithms for fractional programming*, Optimization 66 (2017), pp. 1383–1396.
[11] R. I. Boţ, E. R. Csetnek, D.-K. Nguyen, *A proximal minimization algorithm for structured nonconvex and nonsmooth problems*, SIAM Journal on Optimization 29 (2019), pp. 1300–1328.
[12] S. Banert, R. I. Boţ, *A general double-proximal gradient algorithm for d.c. programming*, Mathematical Programming 178 (2019), pp. 301–326.
[13] L. Chen, S. He and S. Z. Zhang, *When all risk-adjusted performance measures are the same: in praise of the Sharpe ratio*, Quantitative Finance, 11 (2011), pp. 1439–1447.
[14] Y. Cui and J. S. Pang, *Modern Nonconvex Nondifferentiable Optimization*, MOS-SIAM Series on Optimization, 2021.
[15] W. Dinkelbach, *On nonlinear fractional programming*, Management Science 13 (1967), pp. 492–498.
[16] T. Liu, T. K. Pong and A. Takeda, *A successive difference-of-convex approximation method for a class of nonconvex nonsmooth optimization problems*, Mathematical Programming, 176 (2019), pp. 339-367.
[17] M. Y. Hong, Z. Q. Luo, and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM Journal on Optimization, 26 (2016), pp. 337–364.
[18] T. Ibaraki, *Parametric approaches to fractional programs*, Mathematical Programming 26 (1983), pp. 345–362.
[19] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, SIAM, Philadelphia, 2001.
[20] G. Li and T. K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM Journal on Optimization 25 (2015), pp. 2434–2460.
[21] Q. Li, L. X. Shen, N. Zhang and J. P. Zhou, *A proximal algorithm with backtracked extrapolation for a class of structured fractional programming*, Applied and Computational Harmonic Analysis 56 (2022), pp. 98–122.
[22] B. S. Mordukhovich, N. M. Nam and N. D. Yen, *Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming*, Optimization 55 (2006), pp. 685–708.
[23] J.-S. Pang, *A parametric linear complementarity technique for optimal portfolio selection with a risk-free asset*, Operations Research 28 (1980), pp. 927–941.
[24] J.-S. Pang and M. Tao, *Decomposition methods for computing directional stationary solutions of a class of nonsmooth nonconvex optimization problems,* SIAM Journal on Optimization 28 (2018), pp. 1640–1669.
[25] P. Pérez-Aros, E. Vilches, Moreau envelope of supremum functions with applications to infinite and stochastic programming. SIAM J. Optim. 31 (2021), pp. 1635–1657.
[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
[27] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
[28] M. Tao, *Minimization of L_1 over L_2 for sparse signal recovery with convergence guarantee*, SIAM Journal on Scientific Computing 44 (2022), pp. A770–A797.
[29] C. Wang, M. Tao, J. G. Nagy, and Y. Lou, *Limited-angle CT reconstruction via the L_1/L_2 minimization,* SIAM Journal on Imaging Sciences 14 (2021), pp. 749–777.
[30] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., 78 (2019), pp. 1–35.
[31] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, *Image quality assessment: from error visibility to structural similarity,* IEEE Transactions on Image Processing 13 (2004), pp. 600–612.
[32] S. J. Wright, R. D. Nowak and M. A. T. Figueiredo, *Sparse reconstruction by separable approximation,* IEEE Transactions on Signal Processing, 57 (2009), pp. 2479–2493.
[33] P. Yu, G. Li and T.K. Pong, *Kurdyka–Lojasiewicz Exponent via Inf-projection.* Foundations of Computational Mathematics, 22 (2022), pp. 1171–1217.
[34] L. Zeng, P. Yu, and T. K. Pong, *Analysis and algorithms for some compressed sensing models based on L1/L2 minimization,* SIAM Journal on Optimization 31 (2021), pp. 1576–1603.
[35] N. Zhang and Q. Li, *First-order algorithms for a class of fractional optimization problems,* SIAM Journal on Optimization 32 (2022), pp. 100–129.
[36] J. Zhang, P. Xiao, R. Sun, and Z. Luo, *A single-loop smoothed gradient descent-ascent algorithm for nonconvex-concave min-max problems,* arXiv:2010.15768v2, 2022.
[37] J. Zeng, W. Yin and D. Zhou, *Moreau Envelope Augmented Lagrangian Method for Nonconvex Optimization with Linear Constraints,* Journal of Scientific Computing, 91, 61 (2022).