



HAL
open science

AN ADAPTIVE FORWARD-BACKWARD-FORWARD SPLITTING ALGORITHM FOR SOLVING PSEUDO-MONOTONE INCLUSIONS

Flavia Chorobura, Ion I Necoara, Jean-Christophe Pesquet

► **To cite this version:**

Flavia Chorobura, Ion I Necoara, Jean-Christophe Pesquet. AN ADAPTIVE FORWARD-BACKWARD-FORWARD SPLITTING ALGORITHM FOR SOLVING PSEUDO-MONOTONE INCLUSIONS. 2024. hal-04571492

HAL Id: hal-04571492

<https://hal.science/hal-04571492v1>

Preprint submitted on 7 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AN ADAPTIVE FORWARD-BACKWARD-FORWARD SPLITTING ALGORITHM FOR SOLVING PSEUDO-MONOTONE INCLUSIONS*

FLAVIA CHOROBURA[†], ION NECOARA[‡], AND JEAN-CHRISTOPHE PESQUET[§]

Abstract. In this paper, we propose an adaptive forward-backward-forward splitting algorithm for finding a zero of a pseudo-monotone operator which is split as a sum of three operators: the first is continuous single-valued, the second is Lipschitzian, and the third is maximally monotone. This setting covers, in particular, constrained minimization scenarios, such as problems having smooth and convex functional constraints (e.g., quadratically constrained quadratic programs) or problems with a pseudo-convex objective function minimized over a simple closed convex set (e.g., quadratic over linear fractional programs). For the general problem, we design a forward-backward-forward splitting type method based on novel adaptive stepsize strategies. Under an additional generalized Lipschitz property of the first operator, sublinear convergence rate is derived for the sequence generated by our adaptive algorithm. Moreover, if the sum is uniformly pseudo-monotone, linear/sublinear rates are derived depending on the parameter of uniform pseudo-monotonicity. Preliminary numerical experiments demonstrate the good performance of our method when compared to some existing optimization methods and software.

Key words. Pseudo-monotone operators, forward-backward-forward splitting, adaptive stepsize, convergence analysis, nonconvex optimization. \LaTeX

MSC codes. 68Q25, 68R10, 68U05.

1. Introduction. Let \mathbb{H} be a finite-dimensional real vector space endowed with a scalar product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\| \cdot \|$. Our goal is to find a zero of a sum of three operators $A: \mathbb{H} \mapsto \mathbb{H}$, $B: \mathbb{H} \rightarrow \mathbb{H}$, and $C: \mathbb{H} \rightarrow 2^{\mathbb{H}}$, that is

$$(1.1) \quad \text{Find } \bar{z} \in \mathbb{H} \text{ such that } 0 \in A\bar{z} + B\bar{z} + C\bar{z}.$$

Finding a zero of a sum of operators is a very general problem and covers, in particular, constrained optimization, and minimax optimization problems frequently encountered in signal processing [23], triangulation in computer vision [2], semi-supervised learning [14], learning of kernel matrices [26], steering direction estimation for RADAR detection [18], generative adversarial networks [29] among others.

Previous work. The problem of finding a zero of a sum of operators is considered in many works. For example, [8, 15, 17, 24, 37] cover the monotone case, while [3, 10, 33] consider the nonmonotone case. In [17, 24] all three operators are assumed maximally monotone and, additionally, the first is Lipschitz continuous. Under these settings, algorithms based on resolvent and forward operators, activated one at a time successively, are proposed together with a detailed convergence analysis. Furthermore, finding a zero of a sum of two maximally monotone operators, $A + C$, such

*Submitted to the editors DATE.

Funding: The research leading to these results has received funding from: ITN-ETN project TraDE-OPT funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 861137; UEFISCDI, Romania, PN-III-P4-PCE-2021-0720, under project L2O-MOC, nr. 70/2022.

[†]Automatic Control and Systems Engineering Department, National University of Science and Technology Politehnica Bucharest, 060042 Bucharest, Romania, flavia.chorobura@stud.acs.upb.ro.

[‡]Automatic Control and Systems Engineering Department, National University of Science and Technology Politehnica Bucharest, 060042 Bucharest and Gheorghe Mihoc-Caius Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, 050711 Bucharest, Romania, ion.necoara@upb.ro.

[§]University Paris-Saclay, CentraleSupélec, CVN, Inria, Gif-sur-Yvette, France, jean-christophe.pesquet@centralesupelec.fr.

37 that A is a continuous single-valued operator, is investigated in [37] and a forward-
 38 backward-forward algorithm is proposed (also known as Tseng’s algorithm), where
 39 the stepsize is chosen constant when A is Lipschitz or based on an Armijo-Goldstein-
 40 type rule, otherwise. Linear rate was derived for this method when $A + C$ is strongly
 41 monotone. In [3, 10, 33], A is assumed Lipschitz, possibly nonmonotone, and C is
 42 maximally monotone, such that either $A + C$ satisfies the weak Minty condition or
 43 a cohypomonotonicity assumption. In particular, [33] considers an extragradient al-
 44 gorithm with adaptive and constant stepsizes, which reduces, for a specific choice of
 45 stepsize, to the forward-backward-forward algorithm in the monotone case. Moreover,
 46 [10] analyzes an optimistic gradient algorithm, while in [3] algorithms based on class-
 47 ical Halpern and Krasnosel’skii-Mann iterations are analyzed. For all these methods,
 48 under suitable assumptions, sublinear rates are derived. Finally, finding a zero of a
 49 sum of three operators $A + B + C$ is considered in [15], where A, C are maximally
 50 monotone and B is Lipschitz and monotone, and asymptotic convergence is proved
 51 for an error-tolerant forward-backward-forward algorithm.

52 The forward-backward-forward algorithm was also extended to solve variational ine-
 53 equalities. For example, [7, 36] consider a variational inequality, where the operator is
 54 Lipschitz, and a (modified) Tseng algorithm is employed with a constant stepsize or
 55 an adaptive stepsize, so that it is not necessary to know the Lipschitz constant. Con-
 56 vergence is derived when the operator is pseudo-monotone. Moreover, the Lipschitz
 57 assumption on the operator involved in the variational inequality is relaxed in [35],
 58 the operator being assumed continuous. Then, Tseng’s algorithm is considered with
 59 an Armijo-Goldstein rule for the stepsize. Under standard conditions, the weak and
 60 strong convergence of the method is obtained in the pseudo-monotone case. Our ap-
 61 proach differs from [35], as we consider that the operator A satisfies a relaxed Lipschitz
 62 condition and we employ Tseng’s algorithm with novel adaptive stepsize rules (e.g.,
 63 based on the positive root of a polynomial equation). Other methods for solving vari-
 64 ational inequalities with a Lipschitz operator in the monotone case were considered
 65 e.g., in [31] and in the nonmonotone case (under weak Minty condition) in [19].

66 Furthermore, specific algorithms were also developed for particular classes of varia-
 67 tional inequalities, such as convex-concave minimax optimization problems [13, 16].
 68 More specifically, these papers address problems of the form:

$$69 \quad (1.2) \quad \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \langle Lx, y \rangle + \varphi(x) - \psi(y),$$

70 where \mathcal{X} and \mathcal{Y} are Hilbert spaces and $\varphi : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ and $\psi : \mathcal{Y} \rightarrow \bar{\mathbb{R}}$ are proper, convex,
 71 lower semicontinuous functions. For such problems, a primal-dual proximal algorithm
 72 is proposed in [13] for which sublinear rate is derived in the optimality measure:

$$73 \quad G(\bar{x}, \bar{y}) = \max_{y \in \mathcal{Y}} \langle L\bar{x}, y \rangle - \psi(y) + \varphi(\bar{x}) - \min_{x \in \mathcal{X}} \langle Lx, \bar{y} \rangle + \varphi(x) - \psi(\bar{y}),$$

74 for a given $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$. An extension of the algorithm from [13] is given in [16],
 75 where φ is split as $\varphi_1 + \varphi_2$, with $\varphi_1 : \mathcal{X} \rightarrow \mathbb{R}$ convex, differentiable, and having
 76 Lipschitz continuous gradient, while φ_2 is a proper, convex, lower semicontinuous
 77 function. It is proved that this algorithm converges weakly to a solution to problem
 78 (1.2) and, if $\varphi_1 = 0$, then [16] recovers the primal-dual algorithm in [13].
 79

80 *Contributions.* In this paper, we propose a method for finding a zero of a sum of three
 81 operators, which are not necessarily monotone. For this general problem we design
 82 a forward-backward-forward splitting type method based on novel adaptive stepsize
 83 strategies and then perform a detailed convergence analysis. More specifically, our
 84 main contributions are the following.

85 (i) We consider the general problem (1.1) of finding a zero of a sum of three operators,
 86 $A+B+C$, such that A is continuous, B is Lipschitz, and C is maximally monotone. In
 87 contrast to other works that assume A to be Lipschitz and the sum to be monotone,
 88 we relax these conditions, i.e., we require the operator A to satisfy a generalized
 89 Lipschitz condition and the sum to be pseudo-monotone. Our assumptions cover
 90 important classes of optimization problems such as problems minimizing smooth and
 91 convex functional constraints (e.g., quadratically constrained quadratic programs) or
 92 problems minimizing pseudo-convex objective functions over a simple closed convex
 93 set (e.g., quadratic over linear fractional programs).

94 (ii) For solving this general problem we propose a variant of the forward-backward-
 95 forward algorithm [37], based on *two novel adaptive stepsize strategies*. In contrast
 96 to previous works where computationally expensive Armijo-Goldstein stepsize rules
 97 are used when the operator is continuous, we propose two *adaptive* stepsize strategies
 98 that require finding the root of a certain nonlinear equation whose coefficients depend
 99 on the current iterate and on the parameters characterizing the operator properties.
 100 In particular, for quadratically constrained quadratic (resp. quadratic over linear
 101 fractional) programs the stepsize is computed solving a second-order (resp. third-
 102 order) polynomial equation.

103 (iii) Within the considered settings, we provide a detailed convergence analysis for
 104 the forward-backward-forward algorithm based on our adaptive stepsize rules. In
 105 particular, when the sum of the operators is pseudo-monotone, we prove the global
 106 asymptotic convergence for the whole sequence generated by the algorithm and, ad-
 107 ditionally, establish sublinear convergence rate. An improved linear rate is obtained
 108 when the sum is uniformly pseudo-monotone of order $q \in [1, 2]$. Finally, detailed nu-
 109 merical experiments using synthetic and real data demonstrate the effectiveness of our
 110 method and allows us to evaluate its performance when compared to some existing
 111 state-of-the-art optimization methods from [37, 35], and existing software [20].

112 **2. Background.** We denote by $\text{zer}(A)$ the set of zeros of an operator A and
 113 by $\Gamma_0(\mathbb{H})$ the set of proper lower semicontinuous convex functions on \mathbb{H} with values
 114 in $(-\infty, +\infty]$. Further, let us recall the definition of the subdifferential of a convex
 115 function.

116 **DEFINITION 2.1.** *The subdifferential of a proper convex function $f : \mathbb{H} \rightarrow \bar{\mathbb{R}}$ is the*
 117 *set-valued operator $\partial f : \mathbb{H} \rightarrow 2^{\mathbb{H}}$ which maps every point $x \in \mathbb{H}$ to the set*

$$\text{H}\S \quad \partial f(x) = \{u \in \mathbb{H} \mid (\forall y \in \mathbb{H}) \langle y - x, u \rangle + f(x) \leq f(y)\}.$$

120 Note that $\partial f(x) = \emptyset$ for $x \notin \text{dom} f$. For example, let D be a nonempty closed and
 121 convex subset of \mathbb{H} and let its indicator function ι_D be defined as

$$122 \quad (2.1) \quad \iota_D : \mathbb{H} \rightarrow \bar{\mathbb{R}} : x \mapsto \begin{cases} 0 & \text{if } x \in D \\ +\infty, & \text{otherwise.} \end{cases}$$

123 Then, $\partial \iota_D = \mathcal{N}_D$, where \mathcal{N}_D is the normal cone to D , i.e.

$$124 \quad (2.2) \quad \mathcal{N}_D(x) = \begin{cases} \{u \in \mathbb{H} \mid (\forall y \in D) \langle y - x, u \rangle \leq 0\} & \text{if } x \in D \\ \emptyset & \text{otherwise.} \end{cases}$$

125 Moreover, if f is differentiable at a point $x \in \text{dom} f$, its gradient is denoted by $\nabla f(x)$.
 126 Let us also recall the definition of functions with Hölder continuous gradient.

127 DEFINITION 2.2. Let $\nu \in (0, 1]$. Then, the differentiable function $g: \mathbb{H} \rightarrow \mathbb{R}$ has
 128 a ν -Hölder continuous gradient, if there exists $L_g > 0$ such that

$$129 \quad (2.3) \quad (\forall(x, w) \in \mathbb{H}^2) \quad \|\nabla g(x) - \nabla g(w)\| \leq L_g \|x - w\|^\nu.$$

130 If g has ν -Hölder continuous gradient, then the following inequality holds:

$$131 \quad (2.4) \quad (\forall(x, w) \in \mathbb{H}^2) \quad |g(w) - g(x) - \langle \nabla g(x), w - x \rangle| \leq \frac{L_g}{1 + \nu} \|w - x\|^{1+\nu}.$$

132 Next, we present the definitions of pseudo-convex functions and operators.

133 DEFINITION 2.3. Let $\mathcal{X} \subseteq \mathbb{H}$ be an open set, $f: \mathcal{X} \rightarrow \mathbb{R}$ be a differentiable function
 134 and let \mathcal{Z} be a subset of \mathcal{X} . Then, f is said to be pseudo-convex on \mathcal{Z} if, for every
 135 $(x, w) \in \mathcal{Z}^2$, one has:

$$137 \quad \langle \nabla f(x), w - x \rangle \geq 0 \implies f(x) \leq f(w).$$

138 Clearly, any convex function is pseudo-convex and any stationary point of a pseudo-
 139 convex function is a global minimum. However, there are also pseudo-convex functions
 140 that are not convex. For example, consider an open convex set $\mathcal{X} \subset \mathbb{R}^n$ and differ-
 141 entiable functions $g: \mathcal{X} \rightarrow [0, +\infty[$ and $h: \mathcal{X} \rightarrow (0, +\infty)$ such that g is convex and
 142 h is concave. Then, the function $f: \mathcal{X} \rightarrow (0, +\infty): x \mapsto g(x)/h(x)$, is pseudo-convex
 143 on any subset of \mathcal{X} [6]. Other examples of pseudo-convex functions are given in Ex-
 144 ample 3.3 below, see also [27]. The notion of pseudo-convexity was also extended to
 145 nondifferentiable functions, see for example [4].

146 DEFINITION 2.4. An operator $T: \mathbb{H} \rightarrow 2^{\mathbb{H}}$ is said to be pseudo-monotone if

$$147 \quad (\forall(x, y) \in \mathbb{H}^2) (\exists \hat{x} \in Tx) \langle \hat{x}, y - x \rangle \geq 0 \implies (\forall \hat{y} \in Ty) \langle \hat{y}, y - x \rangle \geq 0.$$

149 For example, [25] shows that any differentiable pseudo-convex function has a pseudo-
 150 monotone gradient. In addition, [4] proves that a lower semicontinuous radially con-
 151 tinuous function is pseudo-convex if and only if its subdifferential is pseudo-monotone.
 152 Moreover, note that every monotone operator is pseudo-monotone

153 Finally, let us recall the definition of the resolvent of an operator $C: \mathbb{H} \rightarrow 2^{\mathbb{H}}$. The
 154 resolvent of C is the operator $J_C = (\text{Id} + C)^{-1}$, that is

$$155 \quad (\forall(x, p) \in \mathbb{H}^2) \quad p \in J_C x \iff x - p \in Cp.$$

156 If $C: \mathbb{H} \rightarrow 2^{\mathbb{H}}$ is maximally monotone, then J_C is single-valued, defined everywhere
 157 on \mathbb{H} , and firmly nonexpansive [5]. Moreover, if $C = \partial f$ (the subdifferential operator
 158 of a convex function f), then its resolvent is the proximal mapping $\text{prox}_{\gamma f}$. If $f = \iota_D$,
 159 where ι_D is defined in (2.1) and D is a nonempty closed convex subset of \mathbb{H} , then
 160 $\text{prox}_{\gamma \iota_D} = \text{proj}_D$, where proj_D is the projection operator onto the set D .

161 **3. Assumptions and examples.** In this section we provide several examples
 162 of problems that fit into our framework and also our main assumptions. First, let us
 163 present some important examples of optimization problems that can be recast as (1.1).

164 *Example 3.1.* (Minimizing the sum of three functions). The most straightforward
 165 example of inclusion (1.1) arises from the optimization problem:

$$166 \quad (3.1) \quad \min_{x \in \mathbb{R}^n} F(x) := f(x) + g(x) + h(Lx),$$

167 where $L \in \mathbb{R}^{m \times n}$, f has Hölder gradient, g has Lipschitz gradient, and $h \in \Gamma_0(\mathbb{R}^m)$
 168 is finite at a point in the relative interior of the range of L . This formulation covers
 169

170 smooth ($f = 0$), Hölder smooth ($g = 0$) or hybrid smooth composite problems,
 171 respectively. The first-order optimality condition for (3.1) at \bar{x} reduces to $0 \in A\bar{x} +$
 172 $B\bar{x} + C\bar{x}$, where $A = \nabla f$, $B = \nabla g$, and $C = L^\top \circ \partial h \circ L$.

173 *Example 3.2.* (Minimax problems). Let m and n be positive integers and con-
 174 sider the following minimax problem:

$$175 \quad (3.2) \quad \min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} F(x, y) + \varphi(x) - \psi(y),$$

176 where F is a differentiable function, $\psi = \psi_1 + \psi_2$ and $\varphi = \varphi_1 + \varphi_2$, with ψ_1 and φ_1
 177 having Lipschitz gradients, $\varphi_2 \in \Gamma_0(\mathbb{R}^n)$, and $\psi_2 \in \Gamma_0(\mathbb{R}^m)$. Note that the minimax
 178 problem (3.2) is more general than problem (1.2) considered in previous works, as
 179 we allow more general expressions for $F(x, y)$ beyond bilinear terms. The first-order
 180 optimality conditions for this problem are equivalent to solving the inclusion:

$$181 \quad (3.3) \quad 0 \in A(\bar{x}, \bar{y}) + B(\bar{x}, \bar{y}) + C(\bar{x}, \bar{y}),$$

183 where $\mathbb{H} = \mathbb{R}^n \times \mathbb{R}^m$ and the three operators are

$$184 \quad (3.4) \quad A : \mathbb{H} \rightarrow \mathbb{H} : (x, y) \mapsto (\nabla_x F(x, y), -\nabla_y F(x, y))$$

$$185 \quad B : \mathbb{H} \rightarrow \mathbb{H} : (x, y) \mapsto (\nabla \varphi_1(x), \nabla \psi_1(y)),$$

$$186 \quad C : \mathbb{H} \rightarrow 2^{\mathbb{H}} : (x, y) \mapsto \partial \varphi_2(x) \times \partial \psi_2(y).$$

188 One concrete application of the above minimax formulation is the quadratically con-
 189 strained quadratic program (QCQP) problem:

$$190 \quad \min_{x \in \mathbb{R}^n, x \geq 0} \frac{1}{2} x^\top Q_0 x + b^\top x + c$$

$$191 \quad (3.5) \quad \text{s.t.} \quad \frac{1}{2} x^\top Q_i x + l_i^\top x \leq r_i \quad \forall i \in \{1, \dots, \bar{m}\}, \quad l_i^\top x = r_i \quad \forall i \in \{\bar{m} + 1, \dots, m\},$$

$$192$$

193 where $(Q_i)_{0 \leq i \leq \bar{m}}$ are positive semidefinite matrices of dimension $n \times n$, and $(l_i)_{1 \leq i \leq m}$
 194 and b are vectors in \mathbb{R}^n . Rewriting the QCQP into the Lagrange primal-dual form
 195 using the dual variables $y = (y_i)_{1 \leq i \leq m}$, we get:

$$196 \quad A : \mathbb{H} \rightarrow \mathbb{H} : (x, y) \mapsto \left(\sum_{i=1}^{\bar{m}} (Q_i x + l_i) y_i, \left(-\frac{1}{2} x^\top Q_i x - l_i^\top x + r_i \right)_{1 \leq i \leq m} \right)$$

$$197 \quad (3.6) \quad B : \mathbb{H} \rightarrow \mathbb{H} : (x, y) \mapsto (Q_0 x + b, \mathbf{0}_m)$$

$$198 \quad C : \mathbb{H} \rightarrow 2^{\mathbb{H}} : (x, y) \mapsto (\mathcal{N}_{[0, +\infty)^n}(x) \times (\mathcal{N}_{[0, +\infty)^{\bar{m}}}((y_i)_{1 \leq i \leq \bar{m}}) \times \{\mathbf{0}_{m-\bar{m}}\}),$$

200 where we have set $Q_i = 0$ for every $i \in \{\bar{m} + 1, \dots, m\}$. QCQP's have many ap-
 201 plications, e.g., in signal processing [23], triangulation in computer vision [2], semi-
 202 supervised learning [14], learning of kernel matrices [26], steering direction estimation
 203 for RADAR detection [18].

204 *Example 3.3.* (Fractional programming). Consider the following quadratic frac-
 205 tional programming problem:

$$206 \quad (3.7) \quad \min_{x \in D} f(x) := \frac{\frac{1}{2} x^\top Q x - h^\top x + h_0}{d^\top x + d_0} \quad \text{with} \quad D = \{x \in \mathbb{R}^n \mid d^\top x \geq 0\},$$

$$207$$

208 $d_0 \in (0, +\infty)$, $h_0 \in \mathbb{R}$, $(h, d) \in (\mathbb{R}^n)^2$, and $Q \in \mathbb{R}^{n \times n}$ a symmetric matrix. If \bar{x} satisfies
 209 the first-order optimality condition for (3.7), then we have the following inclusion:

$$(3.8) \quad 0 \in A\bar{x} + C\bar{x},$$

where the operators are

$$(3.9) \quad Ax = \begin{cases} \nabla f(x) & \text{if } x \in D \\ \emptyset & \text{otherwise,} \end{cases} \quad \text{and } C = \mathcal{N}_D.$$

When the matrix Q is positive semidefinite, the function f is pseudo-convex since it is the ratio of convex over concave functions, see [6]. Otherwise, [11, 12] present necessary and sufficient conditions for the function f to be pseudo-convex. Particular cases of (3.7) are problems whose objective is a sum of a linear and a linear fractional function, i.e., when $Q = (rd^\top + dr^\top)/2$, which yields the following formulation:

$$(3.10) \quad \min_{x \in D} f(x) := r^\top x + \frac{h^\top x + h_0}{d^\top x + d_0} \quad \text{with } D = \{x \in \mathbb{R}^n \mid d^\top x \geq 0\}.$$

Reference [28] presents several cases when f is pseudo-convex over the polyhedral set $\{x \in \mathbb{R}^n \mid d^\top x + d_0 > 0\}$, namely, if $r = \eta d$, with $\eta \geq 0$, or $h = \zeta d$, with $h_0 - \zeta d_0 \geq 0$ (see [28, Theorem 1] for more details). Fractional programming arises e.g., in portfolio and transportation problems (see [28] for more details).

Now, we are ready to present our standing assumptions for operators A , B , and C .

ASSUMPTION 1. *The following assumptions hold for problem (1.1):*

- i) C has nonempty closed convex domain, $\text{dom } C$, and is maximally monotone.
- ii) A is a continuous single-valued operator on $\text{dom } C \subseteq \mathbb{H}$.
- iii) B is a single-valued operator and Lipschitz on $\text{dom } C$ with a Lipschitz constant $L_B > 0$ (when $B = 0$, we can take an arbitrarily small positive value for L_B).
- iv) $A + B + C$ is a pseudo-monotone operator.
- v) There exist $(\zeta, \tau) \in (0, +\infty)^2$ such that for every $(u, w) \in \mathbb{H}^2$, $\gamma \in (0, +\infty)$, $q = \text{proj}_{\text{dom } C} w$, and $z = q - \gamma u$, the following holds:

$$(3.11) \quad \|q - J_{\gamma C} z\| \leq \gamma(\zeta \|u\| + \tau).$$

- vi) A satisfies a generalized Lipschitz condition, that is, there exist $\mu \in (0, +\infty)$, $(\beta, \theta) \in [2, +\infty]^2$ and continuous functions a , b , and c from \mathbb{H} to $[0, +\infty[$ such that, for every $(z_1, z_2) \in (\text{dom } C)^2$,

$$(3.12) \quad \|Az_1 - Az_2\|^2 \leq a(z_1)\|z_1 - z_2\|^\mu + b(z_1)\|z_1 - z_2\|^\theta + c(z_1)\|z_1 - z_2\|^\beta.$$

Note that our assumptions are quite general. Clearly, Example 3.1 satisfies Assumption 1, if the objective function F is pseudo-convex. Below we provide other representative examples of important classes of problems that fit into our settings.

Examples satisfying Assumption 1.iv.

- a) Consider Example 3.2, where the operators A , B , C are defined by (3.4) with

$$(3.13) \quad (\forall (x, y) \in \mathbb{H}) \quad F(x, y) = \sum_{i=1}^m y_i g_i(x),$$

where $(g_i)_{1 \leq i \leq m}$ are twice differentiable convex functions, φ and ψ_1 are convex functions, and ψ_2 is the indicator function of the set $[0, +\infty)^m$. Then, $A + \{\mathbf{0}_n\} \times \mathcal{N}_{[0, +\infty)^m}$ is a maximally monotone operator. Indeed, we have

$$(3.14) \quad A : \mathbb{H} \rightarrow \mathbb{H} : (x, y) \mapsto \left(\sum_{i=1}^m y_i \nabla g_i(x), -g(x) \right),$$

251 where $g(x) = [g_1(x), \dots, g_m(x)]^\top$. The Jacobian \mathcal{J}_A of A at $(x, y) \in \mathbb{H}^2$ is

$$252 \quad \mathcal{J}_A(x, y) = \begin{bmatrix} \sum_{i=1}^m y_i \nabla^2 g_i(x) & \nabla g(x) \\ -\nabla g(x)^\top & \mathbf{0}_{m \times m} \end{bmatrix} \quad \text{with } \nabla g(x) = [\nabla g_1(x), \dots, \nabla g_m(x)].$$

254 Note that, for every $(x, y) \in \mathbb{R}^n \times [0, +\infty)^m$, $\mathcal{J}_A(x, y)$ is positive semidefinite ma-
 255 trix. Since A is continuous and monotone on $\mathbb{R}^n \times [0, +\infty[^m$, then A is maximally
 256 monotone on $\mathbb{R}^n \times [0, +\infty[^m$, see [5]. Moreover, since φ and ψ_1 are proper lower
 257 semicontinuous convex functions, then $A + B + C$ is a monotone operator, which
 258 is an instance of a pseudo-monotone operator.

259 b) Consider the following problem

$$261 \quad (3.15) \quad \min_{x \in D} f(x),$$

263 where D is a nonempty closed convex subset of $\mathbb{H} = \mathbb{R}^n$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a
 264 differentiable pseudo-convex function on D . Then, the operators A , B , and C can
 265 be defined as $A = \nabla f$, $B = 0$, and $C = \mathcal{N}_D$. Moreover, $A + C$ is pseudo-monotone.
 266 Indeed, consider $(x, w) \in D^2$. Assume that

$$267 \quad (3.16) \quad (\forall \hat{x} \in Cx) \quad \langle \nabla f(x) + \hat{x}, w - x \rangle \geq 0.$$

269 We need to show that

$$270 \quad (3.17) \quad (\forall \hat{w} \in Cw) \quad \langle \nabla f(w) + \hat{w}, w - x \rangle \geq 0.$$

272 It follows from the definition of the normal cone in (2.2) that

$$273 \quad (3.18) \quad (\forall \hat{x} \in Cx) \quad \langle \hat{x}, w - x \rangle \leq 0 \quad \text{and} \quad (\forall \hat{w} \in Cw) \quad \langle \hat{w}, x - w \rangle \leq 0.$$

275 Combining (3.16) and the first inequality in (3.18) yields

$$276 \quad \langle \nabla f(x), w - x \rangle \geq 0.$$

278 Since f is pseudo-convex, then the above inequality implies that

$$279 \quad \langle \nabla f(w), w - x \rangle \geq 0.$$

281 Hence, from the previous inequality and the second one in (3.18), we derive (3.17).
 282 Therefore, Assumption 1.iv holds. Some examples of pseudo-convex functions are
 283 encountered in fractional programs, see Example 3.3 and also [34, 21].

284 **Example satisfying Assumption 1.v.**

285 Let g be a convex function defined as $g = g_1 + g_2$, where g_1 is the indicator function
 286 of a nonempty closed convex set D and g_2 is a proper lower-semicontinuous convex
 287 function which is Lipschitz on its domain with modulus $L_{g_2} > 0$. We assume that
 288 $D \subseteq \text{dom } g_2$ and there is a point in the intersection of the relative interiors of D
 289 and $\text{dom } g_2$. The latter condition ensures that $\partial(g_1 + g_2) = \partial g_1 + \partial g_2$. Consider the
 290 operator $C = \partial g$. Let $(u, w) \in \mathbb{H}^2$, $q = \text{proj}_{\text{dom } C} w$, and $z = q - \gamma u$. Then, we have

$$291 \quad \text{dom } C = D \quad \text{and} \quad J_{\gamma C} = \text{prox}_{\gamma(g_1 + g_2)}.$$

293 In [1] it was proved that

$$J_{\gamma C} = \text{proj}_{\text{dom}C} \circ \text{prox}_{\gamma g_2}^{g_1}, \quad \text{where} \quad \text{prox}_{\gamma g_2}^{g_1} = (I + \gamma \partial g_2 \circ \text{proj}_{\text{dom}C})^{-1}.$$

Define $p = \text{prox}_{\gamma g_2}^{g_1} z$. Then, $p + \gamma \hat{p} = z$ for some $\hat{p} \in \partial g_2(\text{proj}_{\text{dom}C} p)$. Moreover, $\|\hat{p}\| \leq L_{g_2}$ and

$$\begin{aligned} \|\text{proj}_{\text{dom}C} w - J_{\gamma C} z\| &\leq \|\text{proj}_{\text{dom}C} w - \text{prox}_{\gamma g_2}^{g_1}(z)\| = \|\text{proj}_{\text{dom}C} w - z + \gamma \hat{p}\| \\ &\leq \|\text{proj}_{\text{dom}C} w - z\| + \gamma \|\hat{p}\| \leq \|\text{proj}_{\text{dom}C} w - z\| + \gamma L_{g_2} = \gamma \|u\| + \gamma L_{g_2}, \end{aligned}$$

where, in the first inequality we have used the nonexpansiveness of the projection operator and, in the last one, we have used the linear relation between z , $\text{proj}_{\text{dom}C} w$, and u . Therefore, in this case $\zeta = 1$ and $\tau = L_{g_2}$.

Examples satisfying Assumption 1.vi.

Next we present some examples where Assumption 1.vi is satisfied. In the first two examples we consider the operator A defined in (3.3)-(3.4), where, in the first case, $F(x, y) = yg(x)$ and g has a ν -Hölder continuous gradient, and, in the second case, F is given by (3.13) where the functions $(g_i)_{1 \leq i \leq m}$ have Lipschitz continuous gradients. In the third example, we consider the operator A defined with a pseudo-convex function f as in Example 3.3.

a) Consider the nonlinear operator

$$A : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n \times \mathbb{R} : (x, y) \rightarrow (\nabla g(x)y, -g(x)).$$

where g has a ν -Hölder continuous gradient with constant L_g . Then, for every $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}$ and $\bar{z} = (\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}$, we have

$$\|Az - A\bar{z}\|^2 \leq 2\|\nabla g(x)\|^2 \|z - \bar{z}\|^2 + 4L_g^2 \|z - \bar{z}\|^{2+2\nu} + 4L_g^2 |y|^2 \|z - \bar{z}\|^{2\nu}.$$

Indeed, from the definition of A ,

$$(3.19) \quad \|A(x, y) - A(\bar{x}, \bar{y})\|^2 = \|\nabla g(x)y - \nabla g(\bar{x})\bar{y}\|^2 + |g(\bar{x}) - g(x)|^2.$$

Moreover,

$$\begin{aligned} \|\nabla g(x)y - \nabla g(\bar{x})\bar{y}\|^2 &= \|\nabla g(x)y - \nabla g(x)\bar{y} + \nabla g(x)\bar{y} - \nabla g(\bar{x})\bar{y}\|^2 \\ &\leq 2\|\nabla g(x)\|^2 |y - \bar{y}|^2 + 2|\bar{y}|^2 \|\nabla g(x) - \nabla g(\bar{x})\|^2 \\ &\stackrel{(2.3)}{\leq} 2\|\nabla g(x)\|^2 |y - \bar{y}|^2 + 2L_g^2 |\bar{y}|^2 \|x - \bar{x}\|^{2\nu} \\ (3.20) \quad &\leq 2\|\nabla g(x)\|^2 |y - \bar{y}|^2 + 4L_g^2 |y|^2 \|x - \bar{x}\|^{2\nu} + 4L_g^2 |y - \bar{y}|^2 \|x - \bar{x}\|^{2\nu}, \end{aligned}$$

where, in the first and last inequalities, we used the fact that $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$. On other hand, from (2.4) we deduce that

$$\begin{aligned} |g(\bar{x}) - g(x)|^2 &\leq 2|\langle \nabla g(x), \bar{x} - x \rangle|^2 + \frac{2L_g^2}{(1 + \nu)^2} \|\bar{x} - x\|^{2+2\nu} \\ &\leq 2\|\nabla g(x)\|^2 \|\bar{x} - x\|^2 + \frac{2L_g^2}{(1 + \nu)^2} \|\bar{x} - x\|^{2+2\nu} \\ (3.21) \quad &\leq 2\|\nabla g(x)\|^2 \|\bar{x} - x\|^2 + 4L_g^2 \|\bar{x} - x\|^{2+2\nu}. \end{aligned}$$

Altogether, (3.19), (3.20), and (3.21) lead to

$$\begin{aligned} \|A(x, y) - A(\bar{x}, \bar{y})\|^2 &\leq 2\|\nabla g(x)\|^2 \|(x, y) - (\bar{x}, \bar{y})\|^2 + 4L_g^2 \|(x, y) - (\bar{x}, \bar{y})\|^2 \|\bar{x} - x\|^{2\nu} + 4L_g^2 |y|^2 \|x - \bar{x}\|^{2\nu} \\ &\leq 2\|\nabla g(x)\|^2 \|(x, y) - (\bar{x}, \bar{y})\|^2 + 4L_g^2 \|(x, y) - (\bar{x}, \bar{y})\|^{2+2\nu} + 4L_g^2 |y|^2 \|x - \bar{x}\|^{2\nu}. \end{aligned}$$

Finally, the inequality $\|x - \bar{x}\| \leq \|(x, y) - (\bar{x}, \bar{y})\|$ allows us to prove the statement.

339 b) Consider the nonlinear operator A defined in (3.13)-(3.14) where, for every $i \in$
 340 $\{1, \dots, m\}$, g_i has a Lipschitz continuous gradient with constant $L_{g_i} > 0$ (e.g., g_i
 341 is a quadratic function). Then, for every $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $\bar{z} = (\bar{x}, \bar{y}) \in$
 342 $\mathbb{R}^n \times \mathbb{R}^m$, we have:

$$343 \quad (3.22) \quad \|A(x, y) - A(\bar{x}, \bar{y})\|^2 \leq b(x, y) \|(x, y) - (\bar{x}, \bar{y})\|^2 + c \|(x, y) - (\bar{x}, \bar{y})\|^4,$$

344 with $c = \frac{5}{2} \sum_{i=1}^m L_{g_i}^2$, $b(x, y) = 2(\rho(x, y) + \sum_{i=1}^m \|\nabla g_i(x)\|^2)$, and

$$346 \quad \rho(x, y) = 2 \max \left(m \max_{1 \leq i \leq m} \|\nabla g_i(x)\|^2, \left(\sum_{i=1}^m L_{g_i} |y_i| \right)^2 \right).$$

347 Indeed, similarly to the previous example,
 348

$$349 \quad (3.23) \quad \|A(x, y) - A(\bar{x}, \bar{y})\|^2 = \left\| \sum_{i=1}^m \nabla g_i(x) y_i - \nabla g_i(\bar{x}) \bar{y}_i \right\|^2 + \sum_{i=1}^m |g_i(\bar{x}) - g_i(x)|^2.$$

350 Moreover,
 351

$$352 \quad \left\| \sum_{i=1}^m \nabla g_i(x) y_i - \nabla g_i(\bar{x}) \bar{y}_i \right\|^2 = \left\| \sum_{i=1}^m \nabla g_i(x) y_i - \nabla g_i(x) \bar{y}_i + \nabla g_i(x) \bar{y}_i - \nabla g_i(\bar{x}) \bar{y}_i \right\|^2$$

$$353 \quad \leq \left(\sum_{i=1}^m \|\nabla g_i(x)\| \|y_i - \bar{y}_i\| + \|\bar{y}_i\| \|\nabla g_i(x) - \nabla g_i(\bar{x})\| \right)^2$$

$$354 \quad \leq \left(\max_{1 \leq i \leq m} \|\nabla g_i(x)\| \sum_{i=1}^m |y_i - \bar{y}_i| + \sum_{i=1}^m L_{g_i} |\bar{y}_i| \|x - \bar{x}\| \right)^2$$

$$355 \quad \leq 2 \left(\sqrt{m} \max_{1 \leq i \leq m} \|\nabla g_i(x)\| \|y - \bar{y}\| + \sum_{i=1}^m L_{g_i} |y_i| \|x - \bar{x}\| \right)^2 + 2 \left(\sum_{i=1}^m L_{g_i} |\bar{y}_i - y_i| \|x - \bar{x}\| \right)^2$$

356 where in the fourth inequality we used the Cauchy-Schwarz inequality. Hence,
 357

$$358 \quad \left\| \sum_{i=1}^m \nabla g_i(x) y_i - \nabla g_i(\bar{x}) \bar{y}_i \right\|^2$$

$$359 \quad \leq \rho(x, y) (\|y - \bar{y}\| + \|x - \bar{x}\|)^2 + 2 \left(\sum_{i=1}^m L_{g_i}^2 \right) \|\bar{y}_i - y_i\|^2 \|x - \bar{x}\|^2$$

$$360 \quad (3.24) \quad \leq 2\rho(x, y) \|(x, y) - (\bar{x}, \bar{y})\|^2 + 2 \left(\sum_{i=1}^m L_{g_i}^2 \right) \|(x, y) - (\bar{x}, \bar{y})\|^4.$$

361 On other hand, using (2.4) with $\nu = 1$,

$$363 \quad \sum_{i=1}^m |g_i(\bar{x}) - g_i(x)|^2 \leq \sum_{i=1}^m 2|\langle \nabla g_i(x), \bar{x} - x \rangle|^2 + \frac{L_{g_i}^2}{2} \|x - \bar{x}\|^4$$

$$364 \quad (3.25) \quad \leq 2 \sum_{i=1}^m \|\nabla g_i(x)\|^2 \|x - \bar{x}\|^2 + \sum_{i=1}^m \frac{L_{g_i}^2}{2} \|x - \bar{x}\|^4.$$

365 Hence, (3.23), (3.24), (3.25), and the fact $\|x - \bar{x}\| \leq \|(x, y) - (\bar{x}, \bar{y})\|$ yield (3.22).
 366

367 c) Consider problem (3.10) and operator A defined in (3.9). The Hessian of f is:

$$368 \quad (\forall x \in D) \quad \nabla^2 f(x) = \frac{2(h^\top x + h_0)}{(d^\top x + d_0)^3} dd^\top - \frac{1}{(d^\top x + d_0)^2} (dh^\top + hd^\top).$$

370 Consider $(x, \bar{x}) \in D^2$. By the mean value inequality, there exists $w \in (x, \bar{x})$ s.t.

$$371 \quad \|\nabla f(x) - \nabla f(\bar{x})\|^2 \leq \|\nabla^2 f(w)\|^2 \|x - \bar{x}\|^2$$

$$372 \quad \leq \left(\frac{2}{(d^\top w + d_0)^3} \|d\|^2 |h^\top w + h_0| + \frac{2}{(d^\top w + d_0)^2} |d^\top h| \right)^2 \|x - \bar{x}\|^2$$

$$373 \quad \leq \left(\frac{2}{d_0^3} \|d\|^2 |h^\top(x - w)| + \frac{2}{d_0^3} \|d\|^2 |h^\top x + h_0| + \frac{2}{d_0^2} |d^\top h| \right)^2 \|x - \bar{x}\|^2$$

$$374 \quad \leq 2 \left[\left(\frac{2}{d_0^3} \|d\|^2 |h^\top(x - w)| \right)^2 + \left(\frac{2}{d_0^3} \|d\|^2 |h^\top x + h_0| + \frac{2}{d_0^2} |d^\top h| \right)^2 \right] \|x - \bar{x}\|^2$$

$$375 \quad \leq \frac{8}{d_0^6} \|d\|^4 \|h\|^2 \|x - \bar{x}\|^4 + \frac{8}{d_0^4} \left(\frac{\|d\|^2}{d_0} |h^\top x + h_0| + |d^\top h| \right)^2 \|x - \bar{x}\|^2,$$

377 where in the third inequality we used the fact that, since D is convex, $w \in D$,
378 hence $d^\top w \geq 0$. In the fourth inequality, we used the convexity of $(\cdot)^2$, and in the
379 last one we used that $w \in (x, \bar{x})$. Hence, Assumption 1.vi holds with $a(z_1) = 0$,

$$380 \quad b(z_1) = \frac{8}{d_0^4} \left(\frac{\|d\|^2}{d_0} |h^\top z_1 + h_0| + |d^\top h| \right)^2, \quad c(z_1) = \frac{8}{d_0^6} \|d\|^4 \|h\|^2, \quad \theta = 2, \quad \text{and } \beta = 4.$$

381 d) Consider problem (3.7) and operator A defined in (3.9). The Hessian of f is

$$382 \quad (\forall x \in D) \quad \nabla^2 f(x) = \frac{Q}{d^\top x + d_0} + \frac{(2f(x)dd^\top - (Qx - h)d^\top - d(Qx - h)^\top)}{(d^\top x + d_0)^2}.$$

384 Consider $(x, \bar{x}) \in D^2$. By the mean value inequality, there exists $w \in (x, \bar{x})$ s.t.

$$385 \quad (3.26) \quad \|\nabla f(x) - \nabla f(\bar{x})\| \leq \|\nabla^2 f(w)\| \|x - \bar{x}\|.$$

387 By proceeding similarly to the previous example, we get

$$388 \quad \|\nabla^2 f(w)\| \leq \frac{\|Q\|}{d_0} + \frac{2\|d\|^2}{d_0^3} (\|Q\| \|x\|^2 + |h^\top x - h_0|) + \frac{2\|d\|}{d_0^2} \|Qx - h\|$$

$$389 \quad (3.27) \quad + \left(\frac{2\|d\|^2}{d_0^3} \|h\| + \frac{2\|d\|}{d_0^2} \|Q\| \right) \|\bar{x} - x\| + \frac{2\|d\|^2}{d_0^3} \|Q\| \|\bar{x} - x\|^2.$$

391 We deduce from (3.26) and (3.27) that

$$392 \quad \|\nabla f(x) - \nabla f(\bar{x})\|^2$$

$$393 \quad \leq 3 \left(\frac{2\|d\|^2}{d_0^3} \|h\| + \frac{2\|d\|}{d_0^2} \|Q\| \right)^2 \|\bar{x} - x\|^4 + \frac{12\|d\|^4}{d_0^6} \|Q\|^2 \|\bar{x} - x\|^6$$

$$394 \quad + 3 \left(\frac{\|Q\|}{d_0} + \frac{2\|d\|^2}{d_0^3} (\|Q\| \|x\|^2 + |h^\top x - h_0|) + \frac{2\|d\|}{d_0^2} \|Qx - h\| \right)^2 \|\bar{x} - x\|^2.$$

396 Therefore, Assumption 1.vi is satisfied with $\mu = 6$, $\theta = 4$, $\beta = 2$,

$$397 \quad c(z_1) = 3 \left(\frac{\|Q\|}{d_0} + \frac{2\|d\|^2}{d_0^3} (\|Q\| \|z_1\|^2 + |h^\top z_1 - h_0|) + \frac{2\|d\|}{d_0^2} \|Qz_1 - h\| \right)^2,$$

$$398 \quad (3.28) \quad b(z_1) = 12 \left(\frac{\|d\|^2}{d_0^3} \|h\| + \frac{\|d\|}{d_0^2} \|Q\| \right)^2, \quad \text{and } a(z_1) = \frac{12\|d\|^4}{d_0^6} \|Q\|^2.$$

399

400 From the previous discussion, one can see that our assumptions cover a broad range
401 of optimization problems arising in applications.

402 **4. An adaptive forward-backward-forward algorithm.** Adaptive methods
403 are very popular in optimization as they make stepsize tuning cheap [7, 36, 33]. In
404 contrast to previous works on the forward-backward-forward algorithm in the line
405 of [37], where computationally expensive Armijo-Goldstein stepsize rules are used
406 when the operator is assumed continuous [37, 35], we will propose *two novel adaptive*
407 *strategies*. In these strategies, the stepsize is computed by leveraging the current
408 iterate and the parameters characterizing the operator properties.

409 **4.1. Investigated algorithm.** In this section, we introduce a new algorithm
410 for solving problem (1.1). Our algorithm is similar to the forward-backward-forward
411 splitting algorithm in [37] as it also involves two explicit (forward) steps using A and
412 B , and one implicit (backward) step using C . However, the novelty of our iterative
413 process lies in the adaptive way we choose the stepsize γ_k , which is adapted to the
414 assumptions considered on the operators A, B , and C (see Assumption 1).

Adaptive Forward-Backward-Forward Algorithm (AFBF):

1. Choose the initial estimate $x_0 \in \text{dom } C$.
2. For $k \geq 0$ do:
 - (a) Compute the stepsize $\gamma_k > 0$ and update:
 - (b) $z_k = x_k - \gamma_k(Ax_k + Bx_k)$
 - (c) $p_k = J_{\gamma_k C} z_k$
 - (d) $q_k = p_k - \gamma_k(Ap_k + Bp_k)$
 - (e) $\hat{x}_k = q_k - z_k + x_k$
 - (f) $x_{k+1} = \text{proj}_{\text{dom } C}(\hat{x}_k)$

416 Typically, to prove the convergence of a forward-backward-forward splitting algo-
417 rithm, one needs the operators A and B to satisfy a Lipschitz type inequality [37]:

$$418 \quad (4.1) \quad \gamma_k^2 \|Ax_k + Bx_k - Ap_k - Bp_k\|^2 \leq \alpha_k \|x_k - p_k\|^2,$$

421 where $\alpha_k \in (\alpha_{\min}, \alpha_{\max}) \subset (0, 1)$, $k \in \mathbb{N}$. In our case it is difficult to find a positive
422 stepsize γ_k satisfying (4.1) as the operator A is not assumed to be Lipschitz. However,
423 imposing appropriate assumptions on the operator A (e.g., some generalized Lipschitz
424 type inequality as in Assumption 1.vi), we can ensure (4.1). In the next sections we
425 provide two adaptive choices for γ_k that enable us to prove AFBF convergence.

426 **4.2. First adaptive choice for the stepsize.** In this section, we design a novel
427 strategy to choose γ_k when the operator A satisfies Assumption 1.vi with $\mu \geq 2$.

Stepsize Choice 1:

1. Choose $0 < \alpha_{\min} \leq \alpha_{\max} < 1$ and $\sigma > 0$.
2. For $k \geq 0$ do:
 - (a) Compute $d(x_k) = \zeta \|Ax_k + Bx_k\| + \tau$ and choose $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$.
 - (b) Choose γ_k such that

$$428 \quad (4.2) \quad \gamma_k \in \begin{cases} [\sigma, \bar{\gamma}_k] & \text{if } \sigma \leq \bar{\gamma}_k \\ \bar{\gamma}_k & \text{otherwise,} \end{cases}$$

where $\bar{\gamma}_k > 0$ is the root of the following equation in γ :

$$429 \quad (4.3) \quad a(x_k)d(x_k)^{\mu-2}\gamma^\mu + b(x_k)d(x_k)^{\theta-2}\gamma^\theta + c(x_k)d(x_k)^{\beta-2}\gamma^\beta + L_B^2\gamma^2 = \frac{\alpha_k}{2}.$$

429 Note that equation (4.3), which is a polynomial equation when μ, θ and β are integers,
430 is well defined, i.e., there exists $\bar{\gamma}_k > 0$ satisfying equation (4.3). Indeed, define

431

432

$$(\forall \gamma \in [0, +\infty)) \quad h(\gamma) = 2a(x_k)d(x_k)^{\mu-2} + 2b(x_k)d(x_k)^{\theta-2}\gamma^\theta \\ + 2c(x_k)d(x_k)^{\beta-2}\gamma^\beta + 2L_B^2\gamma^2 - \alpha_k,$$

433

435

436

437

438

439

and $w_k = \sqrt{\alpha_k}/L_B$. Note that, we have $h(w_k) \geq \alpha_k > 0$ and $h(0) < 0$. Since h is continuous on $[0, w_k]$, there exists $\bar{\gamma}_k \in (0, w_k)$ such that $h(\bar{\gamma}_k) = 0$. Moreover, since $h'(\gamma) \geq 4L_B^2\gamma > 0$ for every $\gamma \in (0, +\infty)$, then h is strictly increasing on $(0, +\infty)$. Hence, there exists exactly one $\bar{\gamma}_k > 0$ such that the equality in (4.3) is satisfied and $h(\gamma_k) \leq 0 = h(\bar{\gamma}_k)$ for γ_k defined in (4.2).

440

441

LEMMA 4.1. *Let Assumption 1 hold with $\mu \geq 2$. Let $k \geq 0$ and let γ_k be given by (4.2). Then, inequality (4.1) is satisfied and*

442

443

$$\gamma_k < \eta := \sqrt{\frac{\alpha_{\max}}{2L_B^2}}.$$

444

Proof. From basic properties of the norm,

445

$$\gamma_k^2 \|Ax_k + Bx_k - Ap_k - Bp_k\|^2 \leq 2\gamma_k^2 \|Bx_k - Bp_k\|^2 + 2\gamma_k^2 \|Ax_k - Ap_k\|^2.$$

447

Using Lipschitzianity of operator B on $\text{dom } C$ and Assumption 1.vi, we get

448

449

450

$$\gamma_k^2 \|Ax_k + Bx_k - Ap_k - Bp_k\|^2 \leq 2\gamma_k^2 L_B^2 \|x_k - p_k\|^2 \\ + 2\gamma_k^2 a(x_k) \|x_k - p_k\|^\mu + 2\gamma_k^2 b(x_k) \|x_k - p_k\|^\theta + 2\gamma_k^2 c(x_k) \|x_k - p_k\|^\beta \\ = 2\gamma_k^2 \left(L_B^2 + a(x_k) \|x_k - p_k\|^{\mu-2} + b(x_k) \|x_k - p_k\|^{\theta-2} \right. \\ \left. + c(x_k) \|x_k - p_k\|^{\beta-2} \right) \|x_k - p_k\|^2.$$

451

452

453

454

Using (3.11) with $q = x_k$, $w = \hat{x}_{k-1}$, $u = Ax_k + Bx_k$, and $\gamma = \gamma_k$,

$$\|x_k - p_k\| = \|\text{proj}_{\text{dom } C}(\hat{x}_{k-1}) - J_{\gamma_k C} z_k\|$$

455

456

$$\stackrel{(3.11)}{\leq} \gamma_k (\zeta \|Ax_k + Bx_k\| + \tau) = \gamma_k d(x_k).$$

457

From (4.3), (4.5), (4.6) and the fact that $h(\gamma_k) \leq h(\bar{\gamma}_k)$, we deduce that

458

459

460

461

462

463

$$\gamma_k^2 \|Ax_k + Bx_k - Ap_k - Bp_k\|^2 \\ \stackrel{(4.5),(4.6)}{\leq} 2(L_B^2 \gamma_k^2 + a(x_k) d(x_k)^{\mu-2} \gamma_k^\mu) \|x_k - p_k\|^2 \\ + (b(x_k) d(x_k)^{\theta-2} \gamma_k^\theta + c(x_k) d(x_k)^{\beta-2} \gamma_k^\beta) \|x_k - p_k\|^2 \\ \leq 2(L_B^2 \bar{\gamma}_k^2 + a(x_k) d(x_k)^{\mu-2} \bar{\gamma}_k^\mu) \|x_k - p_k\|^2 \\ + (b(x_k) d(x_k)^{\theta-2} \bar{\gamma}_k^\theta + c(x_k) d(x_k)^{\beta-2} \bar{\gamma}_k^\beta) \|x_k - p_k\|^2 \\ \stackrel{(4.3)}{=} \alpha_k \|x_k - p_k\|^2.$$

465

466

467

From the above inequality, the first statement holds. Moreover, from (4.3), since $b(x_k)$, $c(x_k)$, and $d(x_k)$ are nonnegative for every $k \geq 0$, we have $2L_B^2 \bar{\gamma}_k^2 - \alpha_k \leq 0$. Since $\alpha_k < \alpha_{\max}$, for every $k \geq 0$, inequality (4.4) holds. \square

468

469

470

471

From previous examples, one can see that Stepsize Choice 1 requires the computation of a positive root of a second-order polynomial equation for quadratically constrained quadratic programs (3.5), while for quadratic over linear fractional programs (3.7), one needs to compute the positive root of a third-order equation. More explicitly:

472 (i) If we consider the quadratically constrained quadratic program (3.5), then the
 473 operator A defined in (3.6) for problem (3.5) satisfies (3.22) where, for every $i \in$
 474 $\{1, \dots, m\}$ $g_i: x \mapsto \frac{1}{2}x^\top Q_i x + l_i^\top x - r_i$. Hence, equation (4.3) becomes:

$$475 \quad c d(x_k)^2 \gamma^4 + (\|Q_0\|^2 + b(x_k)) \gamma^2 - \frac{\alpha_k}{2} = 0,$$

477 with function b and c given in (3.22). Solving the corresponding quadratic equation
 478 in γ^2 yields a second-order equation whose positive root is

$$479 \quad \bar{\gamma}_k = \left(\frac{\sqrt{(\|Q_0\|^2 + b(x_k))^2 + 2c d(x_k)^2 \alpha_k} - (\|Q_0\|^2 + b(x_k))}{2c d(x_k)^2} \right)^{1/2}.$$

481 (ii) For the quadratic fractional program (3.7), equation (4.3) becomes

$$482 \quad a(x_k) d(x_k)^4 \gamma^6 + b(x_k) d(x_k)^2 \gamma^4 + (c(x_k) + L_B^2) \gamma^2 - \frac{\alpha_k}{2} = 0,$$

484 where functions a , b , and c are given in (3.28). Setting $\eta = \gamma^2$, we obtain a cubic
 485 equation with a positive root η_k , and then $\bar{\gamma}_k = \sqrt{\eta_k}$.

486 **4.2.1. Convergence results under pseudo-monotonicity.** Next, we show
 487 the asymptotic convergence of the sequences $(x_k)_{k \in \mathbb{N}}$ and $(p_k)_{k \in \mathbb{N}}$ generated by AFBF
 488 Algorithm when $\mu \geq 2$ and the stepsize is computed according to (4.2). The following
 489 sequence will play a key role in our convergence analysis:

$$490 \quad (4.8) \quad (\forall k \in \mathbb{N}) \quad u_k = \gamma_k^{-1} (z_k - p_k) + Ap_k + Bp_k \in Ap_k + Bp_k + Cp_k.$$

491 **THEOREM 4.2.** *Suppose that $\text{zer}(A + B + C) \neq \emptyset$ and Assumption 1 holds with*
 492 *$\mu \geq 2$. Let $(x_k)_{k \in \mathbb{N}}$, $(z_k)_{k \in \mathbb{N}}$, $(p_k)_{k \in \mathbb{N}}$, and $(q_k)_{k \in \mathbb{N}}$ be sequences generated by AFBF*
 493 *algorithm with stepsizes $(\gamma_k)_{k \in \mathbb{N}}$ given by (4.2). Then, the following hold:*

- 494 *i) $(x_k)_{k \in \mathbb{N}}$ is a Fejèr monotone sequence with respect to $\text{zer}(A + B + C)$;*
 495 *ii) $\sum_{k=0}^{+\infty} \|x_k - p_k\|^2 < +\infty$ and $\sum_{k=0}^{+\infty} \|z_k - q_k\|^2 < +\infty$;*
 496 *iii) there exists $\bar{z} \in \text{zer}(A + B + C)$ such that $x_k \rightarrow \bar{z}$, $p_k \rightarrow \bar{z}$, and $u_k \rightarrow 0$.*

497 *Proof.* i) Let $k \in \mathbb{N}$ and let $\bar{z} \in \text{zer}(A + B + C)$. Then,

$$498 \quad \|x_k - \bar{z}\|^2 = \|x_k - p_k + p_k - \bar{z}\|^2$$

$$499 \quad = \|x_k - p_k\|^2 + 2\langle x_k - p_k, p_k - \bar{z} \rangle + \|p_k - \hat{x}_k + \hat{x}_k - \bar{z}\|^2$$

$$500 \quad = \|x_k - p_k\|^2 + \|p_k - \hat{x}_k\|^2 + \|\hat{x}_k - \bar{z}\|^2 + 2\langle x_k - p_k, p_k - \bar{z} \rangle + 2\langle p_k - \hat{x}_k, \hat{x}_k - \bar{z} \rangle$$

$$501 \quad = \|x_k - p_k\|^2 - \|p_k - \hat{x}_k\|^2 + \|\hat{x}_k - \bar{z}\|^2 + 2\langle x_k - p_k, p_k - \bar{z} \rangle + 2\langle p_k - \hat{x}_k, p_k - \bar{z} \rangle$$

$$502 \quad = \|x_k - p_k\|^2 - \|p_k - \hat{x}_k\|^2 + \|\hat{x}_k - \bar{z}\|^2 + 2\langle x_k - \hat{x}_k, p_k - \bar{z} \rangle.$$

504 Moreover, using $p_k - \hat{x}_k = \gamma_k (Ap_k + Bp_k - Ax_k - Bx_k)$, we deduce that

$$505 \quad \|x_k - \bar{z}\|^2 = \|x_k - p_k\|^2 - \gamma_k^2 \|Ap_k + Bp_k - Ax_k - Bx_k\|^2 + \|\hat{x}_k - \bar{z}\|^2$$

$$506 \quad + 2\langle x_k - \hat{x}_k, p_k - \bar{z} \rangle.$$

508 Note that $\bar{z} \in \text{dom}C$. Using the nonexpansiveness of the projection, (4.9) yields

$$509 \quad \|x_{k+1} - \bar{z}\|^2 \leq \|\hat{x}_k - \bar{z}\|^2$$

$$510 \quad \stackrel{(4.9)}{=} \|x_k - \bar{z}\|^2 - \|x_k - p_k\|^2 + \gamma_k^2 \|Ap_k + Bp_k - Ax_k - Bx_k\|^2$$

$$511 \quad - 2\langle x_k - \hat{x}_k, p_k - \bar{z} \rangle.$$

513 We deduce from Lemma 4.1 that

$$514 \quad (4.11) \quad \|x_{k+1} - \bar{z}\|^2 \leq \|x_k - \bar{z}\|^2 - (1 - \alpha_k)\|x_k - p_k\|^2 - 2\langle x_k - \hat{x}_k, p_k - \bar{z} \rangle.$$

516 Since $z_k \in (\text{Id} + \gamma_k C)p_k$, the inclusion relation in (4.8) holds and

$$517 \quad (4.12) \quad x_k - \hat{x}_k = \gamma_k u_k.$$

518 Since $A + B + C$ is pseudo-monotone and \bar{z} is a zero of $A + B + C$, we obtain:

$$519 \quad (4.13) \quad \langle u_k, p_k - \bar{z} \rangle \geq 0.$$

520 Using the last inequality with (4.12) and $\alpha_k \leq \alpha_{\max} < 1$, it follows from (4.11) that

$$521 \quad (4.14) \quad \|x_{k+1} - \bar{z}\|^2 \leq \|x_k - \bar{z}\|^2 - (1 - \alpha_{\max})\|x_k - p_k\|^2 \leq \|x_k - \bar{z}\|^2.$$

523 This shows that $(x_k)_{k \in \mathbb{N}}$ is a Fejèr monotone sequence w.r.t. $\text{zer}(A + B + C)$.

524

525 ii) Since $(x_k)_{k \in \mathbb{N}}$ is a Fejèr monotone sequence, then it is bounded and (4.14) yields

$$526 \quad (4.15) \quad (1 - \alpha_{\max}) \sum_{j=0}^k \|x_j - p_j\|^2 \leq \|x_0 - \bar{z}\|^2 < +\infty.$$

527

528 It follows that $(p_k)_{k \in \mathbb{N}}$ is also bounded. In addition, by using Steps 2.(b) and 2.(e)
529 of AFBF algorithm, Lemma 4.1, Cauchy-Schwarz inequality, and the fact that $\alpha_k \leq$
530 α_{\max} , we further get

$$\begin{aligned} 531 \quad \|z_k - q_k\|^2 &\stackrel{2.(e),(4.12)}{=} \gamma_k^2 \|u_k\|^2 \stackrel{2.(b),(4.8)}{=} \|x_k - p_k + \gamma_k(Ap_k + Bp_k - Ax_k - Bx_k)\|^2 \\ 532 \quad &= \|x_k - p_k\|^2 + 2\gamma_k \langle x_k - p_k, Ap_k + Bp_k - Ax_k - Bx_k \rangle \\ 533 \quad &\quad + \gamma_k^2 \|Ap_k + Bp_k - Ax_k - Bx_k\|^2 \\ 534 \quad &\leq \|x_k - p_k\|^2 + 2\gamma_k \|x_k - p_k\| \|Ap_k + Bp_k - Ax_k - Bx_k\| \\ 535 \quad &\quad + \gamma_k^2 \|Ap_k + Bp_k - Ax_k - Bx_k\|^2 \\ 536 \quad (4.16) \quad &\leq (1 + \sqrt{\alpha_k})^2 \|x_k - p_k\|^2. \end{aligned}$$

538 As a consequence of (4.15) and the boundedness of $(\alpha_k)_{k \in \mathbb{N}}$, $\sum_{k=0}^{+\infty} \|z_k - q_k\|^2 < +\infty$.

539 iii) Let u_k be defined by (4.8). According to (4.16), since $\alpha_k \in (0, 1)$,

$$540 \quad (4.17) \quad \|u_k\| \leq (\gamma_k^{-1}) (1 + \sqrt{\alpha_k}) \|x_k - p_k\| \leq 2\gamma_k^{-1} \|x_k - p_k\|.$$

542 On other hand, from the definition of γ_k and Lemma 4.1, it follows that

$$(4.18)$$

$$\begin{aligned} 543 \quad \alpha_{\min} &\leq \alpha_k \\ 544 \quad &= 2\bar{\gamma}_k^2 \left(L_B^2 + a(x_k)d(x_k)^{\mu-2}\bar{\gamma}_k^{\mu-2} + b(x_k)d(x_k)^{\theta-2}\bar{\gamma}_k^{\theta-2} + c(x_k)d(x_k)^{\beta-2}\bar{\gamma}_k^{\beta-2} \right) \\ 545 \quad &\stackrel{(4.4)}{\leq} 2\bar{\gamma}_k^2 \left(L_B^2 + a(x_k)d(x_k)^{\mu-2}\eta^{\mu-2} + b(x_k)d(x_k)^{\theta-2}\eta^{\theta-2} + c(x_k)d(x_k)^{\beta-2}\eta^{\beta-2} \right). \end{aligned}$$

547 Since A , B , a , b , and c are continuous on $\text{dom } C$ and, $(x_k)_{k \in \mathbb{N}}$ and $(\gamma_k)_{k \in \mathbb{N}}$ are
548 bounded, then $(d_k)_{k \in \mathbb{N}}$ is bounded and there exist $(R_1, R_2, R_3) \in (0, +\infty)^3$ such that

549

$$550 \quad a(x_k)d(x_k)^{\mu-2} \leq R_1, \quad b(x_k)d(x_k)^{\theta-2} \leq R_2, \quad \text{and} \quad c(x_k)d(x_k)^{\beta-2} \leq R_3.$$

551

552 This allows us to lower-bound γ_k as follows:

$$553 \quad (4.19) \quad \gamma_k \geq \gamma_{\min} := \min \left\{ \sigma, \sqrt{\frac{\alpha_{\min}}{2(L_B^2 + R_1\eta^{\mu-2} + R_2\eta^{\theta-2} + R_3\eta^{\beta-2})}} \right\}.$$

555 Hence, from (4.17), we deduce that

$$556 \quad (4.20) \quad \|u_k\| \leq 2\gamma_{\min}^{-1} \|x_k - p_k\|.$$

558 As (4.15) implies that $x_k - p_k \rightarrow 0$, we have

$$559 \quad (4.21) \quad u_k \rightarrow 0.$$

561 To prove the convergence of $(x_k)_{k \in \mathbb{N}}$, according to the Fejèr-monotone convergence
562 theorem [9, Lemma 6], it is sufficient to show that every sequential cluster point of
563 $(x_k)_{k \in \mathbb{N}}$ is a zero of $A + B + C$. Let w be such a sequential cluster point. There thus
564 exists a subsequence $(x_{k_n})_{n \in \mathbb{N}}$ of $(x_k)_{k \in \mathbb{N}}$ such that $x_{k_n} \rightarrow w$. It follows from (4.15)
565 and (4.21) that

$$566 \quad p_{k_n} \rightarrow w \quad \text{and} \quad u_{k_n} \rightarrow 0$$

568 Since A and B are continuous operators on $\text{dom } C$, $u_{k_n} - Ap_{k_n} - Bp_{k_n} \rightarrow -Aw - Bw$. It
569 follows from (4.8) that $(p_{k_n}, u_{k_n} - Ap_{k_n} - Bp_{k_n})$ lies in $\text{gra } C$. Maximally monotonicity
570 of C implies that $(w, -Aw - Bw) \in \text{gra } C$ [5, Proposition 20.33(iii)]. Thus, $w \in$
571 $\text{zer}(A + B + C)$. Hence $x_k \rightarrow w$ and, since $x_k - p_k \rightarrow 0$, $(p_k)_{k \in \mathbb{N}}$ has the same limit.
572 This concludes our proof. \square

573 In [3, 10, 33], the problem of finding a zero of the sum of two operators B and C is
574 considered when B is Lipschitz, C is maximally monotone, and $B + C$ satisfies the weak
575 Minty condition. Next, we analyze the case when we replace the pseudo-monotonicity
576 assumption with the weak Minty condition. Let us first recall this condition.

577 **DEFINITION 4.3.** *An operator $T : \mathbb{H} \rightarrow 2^{\mathbb{H}}$ satisfies the weak Minty condition on*
578 *$\mathcal{Z} \subset \mathbb{H}$ if there exists some $\rho \geq 0$ such that the following holds:*

$$579 \quad (4.22) \quad \langle \hat{w}, w - z \rangle \geq -\rho \|\hat{w}\|^2 \quad \text{for every } z \in \mathcal{Z}, w \in \mathbb{H} \text{ and, } \hat{w} \in Tw.$$

580 Note that pseudo-monotone operators (see Definition 2.4) satisfy the weak Minty condi-
581 tion on their set of zeros \mathcal{Z} with $\rho = 0$. Weak Minty condition covers, in particular,
582 minimization problems having star-convex or quasar-convex differentiable objective
583 functions [22].

584 *Remark 4.4.*

- 585 i) First, one can notice that our proof works with a weak Minty type condition,
586 where $\mathcal{Z} = \text{zer}(A + B + C)$ and $\rho = 0$, instead of Assumption 1.iv. Indeed, in the
587 proof of Theorem 4.2, the pseudo-monotonicity of $A + B + C$ has been used to
588 derive inequality (4.13), which can also be derived from the weak Minty condition
589 with $\rho = 0$.
- 590 ii) Second, let us replace the pseudo-monotone condition in Assumption 1.iv with the
591 assumption that $A + B + C$ satisfies the weak Minty condition on $\text{zer}(A + B + C)$
592 with $\rho > 0$ and, additionally, assume $\text{dom } C$ bounded. From the continuity of
593 A , B , a , b , and c , and the boundedness of $\text{dom } C$, there exists $(R_a, R_b, R_c) \in$
594 $(0, +\infty)^3$ such that, for every $z \in \text{dom } C$,

$$595 \quad (4.23) \quad a(z)d(z) \leq R_a, \quad b(z)d(z) \leq R_b, \quad \text{and} \quad c(z)d(z) \leq R_c.$$

597 Then, the results from the last theorem hold as long as the following conditions
598 are satisfied:

$$599 \quad (4.24) \quad \rho < \frac{2^{-\frac{3}{2}}\sqrt{\alpha_{\min}}(1 - \sqrt{\alpha_{\max}})}{(1 + \sqrt{\alpha_{\max}})\sqrt{L_B^2 + R_a\eta^{\mu-2} + R_b\eta^{\theta-2} + R_c\eta^{\beta-2}}} := \rho_{\max},$$

600 and either

$$601 \quad (4.25) \quad \sigma \geq \sqrt{\frac{\alpha_{\min}}{2(L_B^2 + R_a\eta^{\mu-2} + R_b\eta^{\theta-2} + R_c\eta^{\beta-2})}}, \text{ or } (\forall k \geq 0) \gamma_k = \bar{\gamma}_k.$$

602 Indeed, from (4.11) and (4.22),

$$603 \quad (\forall k \in \mathbb{N}) \quad \|x_{k+1} - \bar{z}\|^2 \leq \|x_k - \bar{z}\|^2 - (1 - \alpha_k)\|x_k - p_k\|^2 + 2\gamma_k\rho\|u_k\|^2.$$

605 Using (4.17), we obtain

$$606 \quad (4.26) \quad \|x_{k+1} - \bar{z}\|^2 \leq \|x_k - \bar{z}\|^2 - (1 - \alpha_k - 2\gamma_k^{-1}\rho(1 + \sqrt{\alpha_k})^2)\|x_k - p_k\|^2.$$

608 On other hand, (4.2), (4.18), (4.23), (4.24), and (4.25) yield

$$609 \quad \gamma_k \geq \min\{\sigma, \bar{\gamma}_k\} \geq \sqrt{\frac{\alpha_{\min}}{2(L_B^2 + R_a\eta^{\mu-2} + R_b\eta^{\theta-2} + R_c\eta^{\beta-2})}} \\ 610 \quad \geq \frac{2\rho_{\max}(1 + \sqrt{\alpha_{\max}})}{1 - \sqrt{\alpha_{\max}}} \geq \frac{2\rho_{\max}(1 + \sqrt{\alpha_k})}{1 - \sqrt{\alpha_k}}.$$

612 Hence, it follows that

$$613 \quad 1 - \alpha_k - 2\gamma_k^{-1}\rho(1 + \sqrt{\alpha_k})^2 \geq \left(1 - \frac{\rho}{\rho_{\max}}\right)(1 - \alpha_k) \geq \left(1 - \frac{\rho}{\rho_{\max}}\right)(1 - \alpha_{\min}).$$

615 The inequality above and (4.26) show that $(x_k)_{k \in \mathbb{N}}$ is a Fejèr monotone sequence
616 with respect to $\text{zer}(A + B + C)$ and $\sum_{k=0}^{+\infty} \|x_k - p_k\|^2 < +\infty$. By proceeding
617 similarly to the proof of Theorem 4.2, the convergence of $(x_k)_{k \in \mathbb{N}}$ to a zero of
618 $A + B + C$ can be proved.

619 Now, we show a sublinear convergence rate result for the iterates of AFBF algorithm.

620 **THEOREM 4.5.** *Under the same assumptions as in Theorem 4.2, the following*
621 *hold: for every $k_0 \in \mathbb{N}$ and $k \in \mathbb{N}^*$,*

$$622 \quad \gamma_{\min} \min_{k_0 \leq j \leq k_0+k-1} \|u_j\| \leq \min_{k_0 \leq j \leq k_0+k-1} \|x_j - \hat{x}_j\| \\ 623 \quad \leq (1 + \sqrt{\alpha_{\max}}) \min_{k_0 \leq j \leq k_0+k-1} \|x_j - p_j\| \leq \frac{\varepsilon_{k_0}}{\sqrt{k}},$$

625 where u_k is defined in (4.8) and $\varepsilon_{k_0} \rightarrow 0$ as $k_0 \rightarrow +\infty$.

626 *Proof.* According to (4.12), (4.17), and (4.19),

$$627 \quad (4.27) \quad (\forall j \in \mathbb{N}) \quad \gamma_{\min} \|u_j\| \leq \|x_j - \hat{x}_j\| \leq (1 + \sqrt{\alpha_{\max}})\|x_j - p_j\|.$$

628 Let \bar{z} be the limit of $(x_j)_{j \in \mathbb{N}}$. It follows from (4.14) that

$$629 \quad (1 - \alpha_{\max}) \sum_{j=k_0}^{k_0+k-1} \|x_j - p_j\|^2 \leq \|x_{k_0} - \bar{z}\|^2,$$

630 which leads to

$$631 \quad \min_{k_0 \leq j \leq k_0 + k - 1} \|x_j - p_j\|^2 \leq \frac{1}{(1 - \alpha_{\max})k} \|x_{k_0} - \bar{z}\|^2.$$

632 The result follows from the latter equation and (4.27), by setting

$$633 \quad \varepsilon_{k_0} = \frac{1}{\sqrt{1 - \alpha_{\max}}} \|x_{k_0} - \bar{z}\|. \quad \square$$

634 Note that convergence results in Theorems 4.2 and 4.5 are consistent with those
635 obtained in the literature on (non)monotone inclusion problems [8, 15, 17, 24, 37].

636 **4.2.2. Convergence results under uniform pseudo-monotonicity.** In this
637 section, we refine our convergence results when the operator $A + B + C$ is uniformly
638 pseudo-monotone. Next, we present the definition of a uniformly monotone/pseudo-
639 monotone operator.

640 **DEFINITION 4.6.** *Let $T: \mathbb{H} \rightarrow 2^{\mathbb{H}}$.*

641 *i) T is said to be uniformly monotone with modulus $q \geq 1$ if there exists a constant
642 $\nu > 0$ such that, for every $(x, y) \in \mathbb{H}^2$ and $(\hat{x}, \hat{y}) \in Tx \times Ty$,*

$$643 \quad \langle \hat{x} - \hat{y}, x - y \rangle \geq \frac{\nu}{2} \|x - y\|^q.$$

644 *ii) T is said to be uniformly pseudo-monotone with modulus $q \geq 1$ if there exists a
645 constant $\nu > 0$ such that, for every $(x, y) \in \mathbb{H}$ and $(\hat{x}, \hat{y}) \in Tx \times Ty$,*

$$647 \quad \langle \hat{x}, y - x \rangle \geq 0 \quad \implies \quad \langle \hat{y}, y - x \rangle \geq \frac{\nu}{2} \|x - y\|^q.$$

649 When $q = 2$ in the definition above, we say that operator T is strongly monotone /
650 pseudo-monotone. Note that, if T is uniformly monotone, then T is also uniformly
651 pseudo-monotone.

652 *Example 4.7.* Consider a proper uniformly convex function $f: \mathbb{R}^n \rightarrow (-\infty, +\infty)$.
653 The subdifferential ∂f of f is uniformly monotone [5, Example 22.5]

654 Below we give an example of a strongly pseudo-monotone map that is not monotone.

655 *Example 4.8.* Consider the unit ball $U = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ and the map
656 $F: U \setminus \{0\} \rightarrow \mathbb{R}^n$ such that

$$657 \quad (\forall x \in U \setminus \{0\}) \quad F(x) = \left(\frac{2}{\|x\|} - 1 \right) x.$$

658 Note that F is not monotone on $U \setminus \{0\}$. For example, setting $y = (1, 0, \dots, 0)$ and
659 $w = (1/2, 0, \dots, 0)$ yields

$$660 \quad \langle F(y) - F(w), y - w \rangle = -\frac{1}{4}.$$

661 However, F is strongly pseudo-monotone on $U \setminus \{0\}$. Indeed, for every $(x, y) \in$
662 $(U \setminus \{0\})^2$, if $\langle F(x), y - x \rangle \geq 0$, then $\langle x, y - x \rangle \geq 0$, and consequently:

$$663 \quad \langle F(y), y - x \rangle = (2\|y\|^{-1} - 1)\langle y, y - x \rangle \geq (2\|y\|^{-1} - 1)\langle y - x, y - x \rangle \geq \|y - x\|^2.$$

664 Next, considering operators A, B, C satisfying Assumption 1 with $\mu \geq 2$ and stepsizes
665 $(\gamma_k)_{k \in \mathbb{N}}$ computed as in (4.2), we derive linear convergence rates when $A + B + C$ is
666 uniformly pseudo-monotone with modulus $q \in [1, 2]$, and sublinear rates when $q > 2$.

667 THEOREM 4.9. *Suppose that Assumption 1 holds with $\mu \geq 2$. Let $(x_k)_{k \in \mathbb{N}}$, and*
 668 *$(p_k)_{k \in \mathbb{N}}$ be sequences generated by AFBF algorithm with stepsizes $(\gamma_k)_{k \in \mathbb{N}}$ given by*
 669 *(4.2). Assume that $A + B + C$ is uniformly pseudo-monotone with modulus $q \geq 1$ and*
 670 *constant $\nu > 0$. Then, for some $\bar{z} \in \text{zer}(A + B + C)$ and constants*

$$671 \quad (4.28) \quad R = \sup_{k \in \mathbb{N}} \|p_k - \bar{z}\| < +\infty \quad \text{and} \quad r = \min\{1 - \alpha_{\max}, \gamma_{\min} \nu R^{q-2}\} < 1,$$

672
 673 *the following hold:*

674 *i) For $q \in [1, 2]$, x_k converges to \bar{z} linearly:*

$$675 \quad (4.29) \quad (\forall k \in \mathbb{N}) \quad \|x_k - \bar{z}\| \leq \left(1 - \frac{r}{2}\right)^{k/2} \|x_0 - \bar{z}\|.$$

676 *ii) For $q > 2$ and $\bar{r} = \frac{r}{2^{q-1} R^{q-2}}$, x_k converges to \bar{z} sublinearly:*

$$677 \quad (\forall k \in \mathbb{N}) \quad \|x_k - \bar{z}\| \leq \frac{\|x_0 - \bar{z}\|}{\left(\frac{q-2}{2} \bar{r} \|x_0 - \bar{z}\|^{q-2} k + 1\right)^{\frac{1}{q-2}}}.$$

679 *Proof.* From Theorem 4.2, the sequence $(p_k)_{k \in \mathbb{N}}$ generated by AFBF algorithm is
 680 convergent. Hence, for some $\bar{z} \in \text{zer}(A + B + C)$, we have $R = \sup_{k \in \mathbb{N}} \|p_k - \bar{z}\| < +\infty$.
 681 Since $A + B + C$ is uniformly pseudo-monotone with modulus q and constant $\nu > 0$,

$$682 \quad (\forall k \in \mathbb{N}) \quad \langle u_k, p_k - \bar{z} \rangle \geq \frac{\nu}{2} \|p_k - \bar{z}\|^q.$$

684 It follows from (4.11) that

$$685 \quad \|x_{k+1} - \bar{z}\|^2$$

$$686 \quad \leq \|x_k - \bar{z}\|^2 - \|x_k - p_k\|^2 + \gamma_k^2 \|Ap_k + Bp_k - Ax_k - Bx_k\|^2 - \gamma_k \nu \|p_k - \bar{z}\|^q.$$

688 Since $\alpha_k \leq \alpha_{\max}$ and $\gamma_{\min} \leq \gamma_k$, we deduce from Lemma 4.1 that

$$689 \quad (4.30) \quad \|x_{k+1} - \bar{z}\|^2 \leq \|x_k - \bar{z}\|^2 - (1 - \alpha_{\max}) \|x_k - p_k\|^2 - \gamma_{\min} \nu \|p_k - \bar{z}\|^q.$$

691 i) If $q \in [1, 2]$, using the definition of R , we get

$$692 \quad (1 - \alpha_{\max}) \|x_k - p_k\|^2 + \gamma_{\min} \nu \|p_k - \bar{z}\|^q$$

$$693 \quad \stackrel{(4.28)}{\geq} (1 - \alpha_{\max}) \|x_k - p_k\|^2 + \gamma_{\min} \nu R^{q-2} \|p_k - \bar{z}\|^2$$

$$694 \quad \geq \min\{1 - \alpha_{\max}, \gamma_{\min} \nu R^{q-2}\} (\|x_k - p_k\|^2 + \|p_k - \bar{z}\|^2) \stackrel{(4.28)}{\geq} \frac{r}{2} \|x_k - \bar{z}\|^2.$$

696 Combining the two last inequalities we obtain

$$697 \quad \|x_{k+1} - \bar{z}\|^2 \leq \left(1 - \frac{r}{2}\right) \|x_k - \bar{z}\|^2.$$

699 Therefore, unrolling the above inequality allows us to prove the first statement.

700 ii) If $q > 2$, it follows from (4.28) that

$$701 \quad (1 - \alpha_{\max}) \|x_k - p_k\|^2 + \gamma_{\min} \nu \|p_k - \bar{z}\|^q$$

$$702 \quad \stackrel{(4.28)}{\geq} \frac{(1 - \alpha_{\max})}{R^{q-2}} \|x_k - p_k\|^q + \gamma_{\min} \nu \|p_k - \bar{z}\|^q$$

$$703 \quad \geq \min \left\{ \frac{(1 - \alpha_{\max})}{R^{q-2}}, \gamma_{\min} \nu \right\} (\|x_k - p_k\|^q + \|p_k - \bar{z}\|^q) \geq \bar{r} \|x_k - \bar{z}\|^q.$$

704

705 Therefore, using (4.30), we obtain $\|x_{k+1} - \bar{z}\|^2 \leq \|x_k - \bar{z}\|^2 - \bar{r}\|x_k - \bar{z}\|^q$. Multiplying
706 the inequality above by $\bar{r}^{\frac{2}{q-2}}$, we obtain

$$707 \quad \bar{r}^{\frac{2}{q-2}} \|x_{k+1} - \bar{z}\|^2 \leq \bar{r}^{\frac{2}{q-2}} \|x_k - \bar{z}\|^2 - \left(\bar{r}^{\frac{2}{q-2}} \|x_k - \bar{z}\|^2 \right)^{\frac{q}{2}}.$$

709 Applying [30, Lemma 8(i)] for $\zeta = \frac{q-2}{2} > 0$, we get

$$710 \quad \|x_k - \bar{z}\| \leq \frac{\|x_0 - \bar{z}\|}{\left(\frac{q-2}{2} \bar{r} \|x_0 - \bar{z}\|^{q-2} k + 1 \right)^{\frac{1}{q-2}}}.$$

712 This proves the second statement of the theorem.

713 *Remark 4.10.* In Theorem 4.9, we can replace the assumption of uniform pseudo-
714 monotonicity with the following one: there exists $\nu > 0$ and $q \geq 1$, such that, for every
715 $w \in \mathbb{H}$, $\hat{w} \in (A + B + C)w$, and $\bar{z} \in \text{zer}(A + B + C)$, the following inequality holds:

$$716 \quad (4.31) \quad \langle \hat{w}, w - \bar{z} \rangle \geq \nu \|w - \bar{z}\|^q.$$

717 Proceeding similarly to the proof of Theorem 4.9, linear and sublinear rates can be
718 derived under this condition. Condition (4.31), with $q = 2$, covers, e.g., minimization
719 problems with strongly star-convex differentiable objective function or strongly quasi-
720 convex objective functions [22].

721 **4.3. Second adaptive choice for the stepsize.** In this section, we present
722 another possible *adaptive* choice for the stepsize when the operator A satisfies As-
723 sumption 1.vi with $\mu \in (0, 2)$. Let $\epsilon \in (0, 1)$ be the desired accuracy for solving
724 problem (1.1), i.e., to obtain u in the range of $A + B + C$ such that $\|u\| \leq \epsilon$. The
725 procedure is described below.

Stepsize Choice 2:

1. Choose $\epsilon \in (0, 1)$, $0 < \alpha_{\min} \leq \alpha_{\max} < 1$, and $\sigma > 0$.

2. For $k \geq 0$ do:

(a) Choose $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ and compute $d(x_k) = \zeta \|Ax_k + Bx_k\| + \tau$.

(b) Compute $\bar{\gamma}_k^{(1)} > 0$ as the solution to the equation

$$(4.32) \quad L_B^2 \gamma^2 + b(x_k) d(x_k)^{\theta-2} \gamma^\theta + c(x_k) d(x_k)^{\beta-2} \gamma^\beta + 2^{2-\mu} a(x_k) \gamma^\mu \epsilon^{\mu-2} = \frac{\alpha_k}{2}$$

(c) Compute $\bar{\gamma}_k^{(2)} > 0$ as the solution to the equation

$$(4.33) \quad L_B^2 d(x_k)^{2-\mu} \gamma^2 + b(x_k) d(x_k)^{\theta-\mu} \gamma^\theta + c(x_k) d(x_k)^{\beta-\mu} \gamma^\beta + a(x_k) \gamma^\mu = \frac{\epsilon^{2-\mu}}{2^{3-\mu}} \alpha_k$$

(d) Update

$$(4.34) \quad \bar{\gamma}_k = \min \left\{ \bar{\gamma}_k^{(1)}, \bar{\gamma}_k^{(2)} \right\}$$

(e) Choose γ_k such that

$$(4.35) \quad \gamma_k \in \begin{cases} [\sigma, \bar{\gamma}_k] & \text{if } \sigma \leq \bar{\gamma}_k \\ \bar{\gamma}_k & \text{otherwise.} \end{cases}$$

727 Note that γ is well-defined in Steps 2.(b) and 2.(c) of this second procedure for the
728 choice of the stepsize, i.e., there exist unique $\bar{\gamma}_k^{(1)}, \bar{\gamma}_k^{(2)}$ satisfying (4.32) and (4.33),
729

730 respectively. Indeed, consider the functions

$$731 \quad h(\gamma) = \gamma^2 L_B^2 + b(x_k)d(x_k)^{\theta-2}\gamma^\theta + c(x_k)d(x_k)^{\beta-2}\gamma^\beta + 2^{2-\mu}a(x_k)\gamma^\mu\epsilon^{\mu-2} - \frac{\alpha_k}{2}$$

$$732 \quad r(\gamma) = L_B^2 d(x_k)^{2-\mu}\gamma^2 + b(x_k)d(x_k)^{\theta-\mu}\gamma^\theta + c(x_k)d(x_k)^{\beta-\mu}\gamma^\beta + a(x_k)\gamma^\mu - \frac{\epsilon^{2-\mu}}{2^{3-\mu}}\alpha_k,$$

734 and variables $w_k = \frac{\sqrt{\alpha_k}}{L_B}$ and $\bar{w}_k = \frac{\sqrt{\alpha_k}}{L_B d(x_k)^{\frac{2-\mu}{2}}}$. Note that $h(0) < 0$ and $h(w_k) \geq$

735 $\alpha_k/2 > 0$. Since h is continuous on $[0, w_k]$ there exist $\bar{\gamma}_k^{(1)} \in (0, w_k)$ such that
 736 $h(\bar{\gamma}_k^{(1)}) = 0$. Moreover, since $h'(\gamma) \geq 2\gamma L_B^2 > 0$ for every $\gamma \in (0, +\infty)$, then h is
 737 strictly increasing in $(0, +\infty)$. Hence, there exists exactly one $\bar{\gamma}_k^{(1)} > 0$ such that
 738 $h(\bar{\gamma}_k^{(1)}) = 0$. Using the same arguments, we can conclude that r is strictly increasing
 739 on $(0, +\infty)$ and there exist only one $\bar{\gamma}_k^{(2)} \in (0, \bar{w}_k)$ such that $r(\bar{\gamma}_k^{(2)}) = 0$. Since both
 740 functions h and r are strictly increasing in $(0, +\infty)$, $h(0) < 0$ and $r(0) < 0$, γ_k defined
 741 in (4.35) satisfies the following two inequalities:

$$742 \quad (4.36) \quad h(\gamma_k) \leq 0 \quad \text{and} \quad r(\gamma_k) \leq 0.$$

744 Note that

$$745 \quad (4.37) \quad \bar{\gamma}_k^{(1)} \leq \eta \quad \text{and} \quad \bar{\gamma}_k^{(2)} \leq \bar{\eta} := \left(\frac{\epsilon^{2-\mu}\alpha_{\max}}{2^{3-\mu}L_B^2\tau^{2-\mu}} \right)^{\frac{1}{2}},$$

746 with η defined in (4.4). The theorem below provides a bound on the number of
 747 iterations required, for a given $\epsilon > 0$, to generate $\|u_k\| \leq \epsilon$, with u_k defined in (4.8).

748 **THEOREM 4.11.** *Let $\epsilon \in (0, 1)$. Suppose that Assumption 1 holds with $\mu \in (0, 2)$.
 749 Let $(x_k)_{k \in \mathbb{N}}$ and $(p_k)_{k \in \mathbb{N}}$ be the sequences generated by AFBF algorithm with stepsizes
 750 $(\gamma_k)_{k \in \mathbb{N}}$ given by (4.35). Then, for $u_k = \gamma_k^{-1}(x_k - p_k) + Ap_k + Bp_k - Ax_k - Bx_k \in$
 751 $Ap_k + Bp_k + Cp_k$ and $\gamma_{\min}(\epsilon) = \mathcal{O}(\epsilon^{(2-\mu)/\mu})$, performing*

$$752 \quad K \geq \frac{1}{\epsilon^2} \left(\frac{(1 + \sqrt{\alpha_{\max}})^2}{\gamma_{\min}^2(\epsilon)(1 - \alpha_{\max})} \right) \|x_0 - \bar{z}\|^2$$

754 iterations ensures that there exists $k \in \{0, \dots, K-1\}$ such that $\|u_k\| \leq \epsilon$.

755 *Proof.* i) First, consider the case when, for every $k \in \{0, \dots, K-1\}$, $\gamma_k^{-1}\|x_k -$
 756 $p_k\| > \epsilon/2$. We deduce from (4.6) and (4.36) that

$$757 \quad \gamma_k^2 \|Ax_k + Bx_k - Ap_k - Bp_k\|^2$$

$$758 \quad \stackrel{(4.6)}{\leq} 2 \left(\gamma_k^2 L_B^2 + b(x_k)d_k^{\theta-2}\gamma_k^\theta + c(x_k)d_k^{\beta-2}\gamma_k^\beta + 2^{2-\mu}a(x_k)\gamma_k^\mu\epsilon^{\mu-2} \right) \|x_k - p_k\|^2$$

$$759 \quad \stackrel{(4.36)}{\leq} \alpha_k \|x_k - p_k\|^2.$$

761 Let $\bar{z} \in \text{zer}(A + B + C)$. Since $\alpha_k \leq \alpha_{\max}$, using a similar reasoning as in (4.10),
 762 the inequality (4.14) also holds when $k \in \{0, \dots, K-1\}$, for this second choice
 763 of the stepsize. This implies that

$$764 \quad (4.38) \quad (1 - \alpha_{\max}) \sum_{k=0}^{K-1} \|x_k - p_k\|^2 \leq \|x_0 - \bar{z}\|^2,$$

$$765 \quad (\forall k \in \{0, \dots, K\}) \quad \|x_k - \bar{z}\| \leq \|x_0 - \bar{z}\|.$$

767 Let D be the closed ball of center \bar{z} and radius $\|x_0 - \bar{z}\|$. Since A , B , a , b , and c
768 are continuous on $\text{dom } C$, the quantities define below take finite values:

$$769 \quad R_1 = \sup_{x \in D} a(x), \quad R_2 = \sup_{x \in D} b(x)d(x)^{\theta-2}, \quad R_3 = \sup_{x \in D} c(x)d(x)^{\beta-2}$$

$$770 \quad R_4 = \sup_{x \in D} d(x)^{2-\mu}, \quad R_5 = \sup_{x \in D} b(x)d(x)^{\theta-\mu}, \quad R_6 = \sup_{x \in D} c(x)d(x)^{\beta-\mu}.$$

772 From (4.37), (4.32) and (4.33), one can lower-bound the stepsize as:

$$773 \quad (\forall k \in \{0, \dots, K-1\}) \quad \gamma_k \geq \gamma_{\min}(\epsilon) := \min\{\gamma_{\min}^{(1)}(\epsilon), \gamma_{\min}^{(2)}(\epsilon), \sigma\},$$

775 with

$$776 \quad (4.39) \quad \gamma_{\min}^{(1)}(\epsilon) := \left(\frac{\alpha_{\min}}{2(L_B^2 \eta^{2-\mu} + 2^{2-\mu} R_1 \epsilon^{\mu-2} + R_2 \eta^{\theta-\mu} + R_3 \eta^{\beta-\mu})} \right)^{\frac{1}{\mu}}$$

778 and

$$779 \quad (4.40) \quad \gamma_{\min}^{(2)}(\epsilon) := \left(\frac{\epsilon^{2-\mu} \alpha_{\min}}{2^{3-\mu} (R_1 + L_B^2 \bar{\eta}^{2-\mu} R_4 + R_5 \bar{\eta}^{\theta-\mu} + R_6 \bar{\eta}^{\beta-\mu})} \right)^{\frac{1}{\mu}}.$$

781 Note that, if ϵ is sufficiently small, then $\gamma_{\min}(\epsilon) = \mathcal{O}(\epsilon^{(2-\mu)/\mu})$. Using (4.17), we
782 finally obtain

$$783 \quad (\forall k \in \{0, \dots, K-1\}) \quad \|u_k\| \leq (\gamma_{\min}(\epsilon))^{-1} (1 + \sqrt{\alpha_{\max}}) \|x_k - p_k\|,$$

785 which, by virtue of (4.38), yields

$$786 \quad \min_{0 \leq k \leq K-1} \|u_k\|^2 \leq \frac{1}{K} \left(\frac{(1 + \sqrt{\alpha_{\max}})^2}{\gamma_{\min}^2(\epsilon)(1 - \alpha_{\max})} \right) \|x_0 - \bar{z}\|^2.$$

788 ii) Second, consider the case when, there exists $k \in \{0, \dots, K-1\}$ such that
789 $\gamma_k^{-1} \|x_k - p_k\| \leq \epsilon/2$. Let us prove that $\|Ap_k + Bp_k - Ax_k - Bx_k\| \leq \epsilon/2$.
790 Indeed, we deduce from (4.5) that

$$791 \quad \|Ap_k + Bp_k - Ax_k - Bx_k\|^2$$

$$792 \quad \leq 2L_B^2 \|x_k - p_k\|^2 + 2a(x_k) \|x_k - p_k\|^\mu + 2b(x_k) \|x_k - p_k\|^\theta + 2c(x_k) \|x_k - p_k\|^\beta$$

$$793 \quad \stackrel{(4.6)}{\leq} 2(L_B^2 d(x_k)^{2-\mu} \gamma_k^{2-\mu} + b(x_k) d(x_k)^{\theta-\mu} \gamma_k^{\theta-\mu} + c(x_k) d(x_k)^{\beta-\mu} \gamma_k^{\beta-\mu}$$

$$794 \quad + a(x_k)) \|x_k - p_k\|^\mu$$

$$795 \quad \leq 2(L_B^2 d(x_k)^{2-\mu} \gamma_k^2 + b(x_k) d(x_k)^{\theta-\mu} \gamma_k^\theta + c(x_k) d(x_k)^{\beta-\mu} \gamma_k^\beta + a(x_k) \gamma_k^\mu) \frac{\epsilon^\mu}{2^\mu}$$

$$796 \quad \stackrel{(4.36)}{\leq} \frac{\epsilon^2 \alpha_k}{4} \leq \frac{\epsilon^2}{4}.$$

798 Hence, from the definition of u_k , applying the triangle inequality leads to $\|u_k\| \leq$
799 ϵ . Hence, the statement of the theorem is proved. \square

800 It can be noticed that the literature on convergence rates for the general inclusion
801 problem addressed in this section is scarce. Existing results predominantly focus on
802 the composite problem outlined in Example 3.1, particularly when $g = 0$ and $L = I_n$,
803 spanning both the convex case [32] and the nonconvex one [38].

804 **5. Simulations.** In this section, we evaluate the performance of our algorithm
 805 on convex quadratically constrained quadratic programs (QCQPs), see (3.5), using
 806 synthetic and real data. Then, we also test our algorithm on a pseudo-convex prob-
 807 lem using synthetic data. We compare our Adaptive Forward-Backward-Forward
 808 (AFBF) algorithm to Tseng’s algorithm [37], and one dedicated commercial opti-
 809 mization software packages, Gurobi [20] (which has a specialized solver for QCQPs).
 810 We implemented the algorithm AFBF as follows: at each iteration $k \in \mathbb{N}$, the stepsize
 811 $\gamma_k = \bar{\gamma}_k$, where $\bar{\gamma}_k$ is computed as in (4.3), $b(x_k)$ and $c(x_k)$ are computed as in (3.22),
 812 and $\alpha_k = 0.99$. The code was implemented using MATLAB R2020a on a computer
 813 equipped with an AMD Ryzen CPU operating at 3.4 GHz and 64 GB of RAM.

814 **5.1. Solving convex QCQPs.** We consider the following convex QCQP

$$815 \quad \min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^\top Q_0 x + b^\top x$$

$$816 \quad (5.1) \quad \text{s.t.} \quad g_i(x) = \frac{1}{2} x^\top Q_i x + l_i^\top x - r_i \leq 0, \quad \forall i \in \{1, \dots, m\},$$

818 where $(Q_i)_{0 \leq i \leq m}$ are symmetric positive semidefinite matrices in $\mathbb{R}^{n \times n}$, $(l_i)_{1 \leq i \leq m}$ and
 819 b are vectors in \mathbb{R}^n , and $(r_i)_{1 \leq i \leq m}$ are nonnegative reals. Note that the operator A
 820 defined in (3.6) for QCQPs fits (3.22). For every $i \in \{0, \dots, m\}$, Q_i was generated as
 821 $Q_i = R_i^\top R_i$, where R_i is a sparse random matrix whose element are drawn indepen-
 822 dently from a uniform distribution over $[0, 1]$. Moreover, the components of vectors
 823 b and $(l_i)_{1 \leq i \leq m}$ were generated from a standard normal distribution $\mathcal{N}(0, 1)$. Con-
 824 stants $(r_i)_{1 \leq i \leq m}$ and the components of the algorithm starting point were generated
 825 from a uniform distribution over $[0, 1]$. For the algorithm in [37], named **Tseng**, the
 826 line-search is computed as in [37, equation (2.4)], with $\theta = 0.995$, $\sigma = 1$, and $\beta = 0.5$.
 827 We consider the following stopping criteria for AFBF and Tseng’s algorithms:

$$828 \quad \|u_k\| \leq 10^{-2}, \text{ with } u_k \text{ defined in (4.8).}$$

n	p	m	AFBF		Tseng [37]		
			ITER	CPU	ITER	LSE	CPU
10^3	10^3	250	3914	36.09	15298	91513	387.4
10^3	10^3	500	7563	131.8	23400	140070	1179.3
10^3	10^3	10^3	19044	597.6	37932	227029	3570.4
10^3	10^3	$2 \cdot 10^3$	44039	2900.1	63143	377963	12990
10^4	10^4	125	4705	195.5	3351	19963	418.6
10^4	10^4	250	6131	475.2	4888	29209	1178
10^4	10^4	500	8862	1329	7240	43319	3398
10^4	10^4	750	11380	1821	8670	51893	4251
10^3	500	250	4992	66.9	14750	88223	590.9
10^3	500	500	11069	288.7	25741	154114	2068.7
10^3	500	10^3	24460	1192.7	45654	273360	7010.4
10^3	500	$2 \cdot 10^3$	59762	5939	*	*	*
10^4	$5 \cdot 10^3$	125	5318	336	3428	20412	689.8
10^4	$5 \cdot 10^3$	250	7445	895.3	4762	28452	1864
10^4	$5 \cdot 10^3$	500	11515	2711	11271	67514	8647
10^4	$5 \cdot 10^3$	750	15719	3655.4	14073	84324	10462

TABLE 1

CPU time (sec) and number of iterations (ITER) for solving synthetic QCQPs of the form (5.1) with AFBF and Tseng’s [37] algorithms: strongly convex case (top) and convex case (bottom).

830 The CPU time (in seconds) and the number of iterations (ITER) required by each
 831 algorithm for solving problem (5.1) are given in Table 1, where “*” means that the
 832 corresponding algorithm needs more than 5 hours to solve the problem. Moreover,
 833 for Tseng’s algorithm, we also report the number of line-search evaluations (LSE).
 834 The first half of the table corresponds to strongly convex functions ($Q_i \succ 0$, for
 835 every $i \in \{0, \dots, m\}$) and the other half is for convex functions ($Q_i \succeq 0$, for every

836 $i \in \{0, \dots, m\}$). As we can notice in Table 1, AFBF outperforms Tseng's algorithm
 837 (sometimes even $10\times$ faster). Comparisons with Gurobi software are not included in
 838 Table 1, since we observed that its performance is quite poor on these large test cases.

839 **5.2. Solving multiple kernel learning in support vector machine.** In
 840 this section, we test AFBF on Support Vector Machine (SVM) with multiple kernel
 841 learning using real data, which can also be formulated as a convex QCQP. Let us
 842 briefly describe the problem (our presentation follows [14]). Given a set of n_{dat} data
 843 points $\mathcal{S} = \{(d_j, l_j)\}_{1 \leq j \leq n_{\text{dat}}}$ where, for every $j \in \{1, \dots, n_{\text{dat}}\}$ $d_j \in \mathbb{R}^{n_d}$ is the input
 844 vector and $l_j \in \{-1, 1\}$ is its class label, SVM searches for a hyperplane that can best
 845 separate the points from the two classes. When the data points cannot be separated
 846 in the original space \mathbb{R}^{n_d} , we can search in a feature space \mathbb{R}^{n_f} , by mapping the input
 847 data space \mathbb{R}^{n_d} to the feature space through a function $\varphi : \mathbb{R}^{n_d} \rightarrow \mathbb{R}^{n_f}$. Using function
 848 φ , we can define a kernel function $\kappa : \mathbb{R}^{n_d} \times \mathbb{R}^{n_d} \rightarrow \mathbb{R}$ as $\kappa(d_j, d_{j'}) := \langle \varphi(d_j), \varphi(d_{j'}) \rangle$
 849 for every $(d_j, d_{j'}) \in (\mathbb{R}^{n_d})^2$, where $\langle \cdot, \cdot \rangle$ denotes the inner product of \mathbb{R}^{n_f} . One popular
 850 choice of the kernel function in the SVM literature is the Gaussian kernel:

$$851 \quad \kappa_{\text{GAU}}(d_j, d_{j'}) = \exp\left(-\frac{\|d_j - d_{j'}\|^2}{2\sigma^2}\right), \quad \forall (j, j') \in \{1, \dots, n_{\text{dat}}\}^2$$

852 with $\sigma > 0$. We separate the given set \mathcal{S} into a training set, $\mathcal{S}_{\text{tr}} = \{(d_j, l_j)\}_{1 \leq j \leq n_{\text{tr}}}$
 853 and a testing set, $\mathcal{S}_{\text{te}} = \{(d_j, l_j)\}_{1 \leq j \leq n_{\text{te}}}$, such that $n_{\text{tr}} + n_{\text{te}} = n_{\text{dat}}$. Choosing a set
 854 of kernel functions $(\kappa_i)_{1 \leq i \leq m}$, the SVM classifier is learned by solving the following
 855 convex QCQP problem on the training set \mathcal{S}_{tr} :

$$856 \quad \min_{x \in \mathbb{R}^{n_{\text{tr}}}, x_0 \in \mathbb{R}, x \geq 0} \frac{1}{2} x^\top Q_0 x - e^\top x + R x_0$$

$$857 \quad \text{s.t.} \quad \frac{1}{2} x^\top \left(\frac{1}{R_i} G_i(K_{i,\text{tr}}) \right) x - x_0 \leq 0 \quad \forall i \in \{1, \dots, m\}, \quad \sum_{j=1}^{n_{\text{tr}}} l_j x_j = 0,$$

858 where $Q_0 = C^{-1} I_{n_{\text{tr}}}$, C being a parameter related the soft margin criteria, and the
 859 vector e denotes a vector of all ones. In addition, for every $i \in \{1, \dots, m\}$, $K_{i,\text{tr}} \in$
 860 $\mathbb{R}^{n_{\text{tr}} \times n_{\text{tr}}}$ is a symmetric positive semidefinite matrix, whose (j, j') element is defined
 861 by the kernel function: $[K_{i,\text{tr}}]_{j,j'} := \kappa_i(d_j, d_{j'})$. The matrix $G_i(K_{i,\text{tr}}) \in \mathbb{R}^{n_{\text{tr}} \times n_{\text{tr}}}$ in
 862 the i -th quadratic constraint of (5.2) is a symmetric positive semidefinite matrix, its
 863 (j, j') element being $[G_i(K_{i,\text{tr}})]_{j,j'} = l_j l_{j'} [K_{i,\text{tr}}]_{j,j'}$. Moreover, R and $(R_i)_{1 \leq i \leq m}$ are
 864 given positive constants. Clearly, (5.2) is an instance of problem (3.5). In our exper-
 865 iments, we employed a predefined set of Gaussian kernel functions $(\kappa_i)_{1 \leq i \leq m}$, with
 866 the corresponding $(\sigma_i^2)_{1 \leq i \leq m}$ values. Following the pre-processing strategy outlined
 867 in [14], we normalized each matrix $K_{i,\text{tr}}$ such that $R_i = \text{trace}(K_{i,\text{tr}})$ was set to 1, thus
 868 restricting $R = \sum_{i=1}^m R_i = m$. For each dataset, the σ_i^2 's were set to m different grid
 869 points within the interval $[10^{-1}, 10]$ for the first five datasets and $[10^{-2}, 10^2]$ for the
 870 last one, with two different values for the number of grid points, namely $m = 3$ and
 871 $m = 5$. Additionally, we set $C = 1$. In order to give a better overview of the advan-
 872 tages offered by the multiple kernel SVM approach, we also learn a single Gaussian
 873 kernel SVM classifier with σ^2 set a priori to 7, by solving the following QP problem:

$$874 \quad \min_{x \in [0, C]^{n_{\text{tr}}}} \frac{1}{2} x^\top G(K_{\text{tr}}) x - e^\top x, \quad \text{s.t.} \quad \sum_{j=1}^{n_{\text{tr}}} l_j x_j = 0.$$

875 We consider the following stopping criteria for AFBF and Tseng's algorithms:

$$876 \quad |f(x) - f^*| \leq 10^{-4}, \quad \left| \sum_{j=1}^{n_{\text{tr}}} l_j x_j \right| \leq 10^{-4} \quad \text{and} \quad \max(0, g_i(x)) \leq 10^{-4}, \quad \forall i \in \{1, \dots, m\},$$

880

881 with f^* computed by Gurobi solver and the starting point chosen as the null vector.
 882 Moreover, for Tseng’s algorithm the line-search was computed as in [37, equation
 883 (2.4)], with $\theta = 0.99$, $\sigma = 1$ and $\beta = 0.1$. Table 2 presents a comparison between
 884 AFBF algorithm, Tseng’s algorithm [37], and Gurobi solver [20] (specialized solver for
 885 QCQPs) in terms of CPU time for solving the QCQP of the form (5.2) using 6 real
 886 datasets `Ozone-level-8hr`, `mfeat-fourier`, `USPS`, `isolet`, `semeion` and `Ovarian`
 887 from <https://www.openml.org>. Each dataset was divided into a training set comprising
 888 80% of the data and a testing set of the remaining 20%. For each dataset,
 889 we also provided the nonzero optimal dual multiplier value y^* corresponding to the
 890 unique active quadratic inequality constraint and the corresponding value of σ^2 corresponding
 891 to that active constraint. Finally, the table presents a comparison between
 892 the Testing Set Accuracies on the remaining testing datasets obtained by the multiple
 893 Gaussian kernel SVM classifier with σ^2 derived from (5.2), named TSA, and the single
 894 Gaussian kernel SVM classifier with $\sigma^2 = 7$, named TSA0.

Dataset (n, n_d)	m	TSA0	TSA	σ^2	AFBF		TSENG		Gurobi	
					CPU	y^*	CPU	y^*	CPU	y^*
Ozone-level-8hr (2534, 72)	3	52.7	91.7	5.05	31.18	3.1	58.09	2.99	95.61	3
	5		91.7	2.575	49.9	5.04	61.38	5	339.88	5
mfeat-fourier (2000, 76)	3	87.7	89	5.05	11.82	3.04	21.5	2.99	40.56	3
	5		89	2.575	20.54	5.02	35.06	4.99	170.06	5
USPS (1424, 256)	3	60.2	91.5	10	4	3	5.23	3	232.98	3
	5		92.2	10	3.95	5	8.33	5	1106.7	5
isolet (600, 617)	3	57.5	95	10	0.59	3	1.35	3	10.8	3
	5		95.8	10	0.68	4.97	2.23	5	25.09	5
semeion (319, 256)	3	47.6	77.8	10	0.75	2.98	1.43	2.97	1.37	3
	5		84.1	10	0.89	5.02	3.19	4.99	4.12	5
Ovarian (253, 15154)	3	66	78	100	0.38	3.04	1.72	2.99	0.82	3
	5		88	100	0.47	4.96	2.48	4.99	2.31	5

TABLE 2

Comparison between our algorithm AFBF, Tseng’s algorithm [37] and Gurobi solver [20] in terms of CPU time (in seconds) to solve QPQCs of the form (5.2) for various real datasets and two different choices of $m = 3, 5$. Additionally, TSA’s are provided for (5.2) and (5.3).

895 **5.3. Fractional programming.** In this final set of experiments, we consider the
 896 linear fractional program (3.10), where the objective function is pseudo-convex. We
 897 compare our algorithm with [35, Algorithm 1] developed for solving non-Lipschitzian
 898 and pseudo-monotone variational inequalities. We implemented [35, Algorithm 1]
 899 with the parameters $\mu = 0.995$, $\gamma = 1$, and $l = 0.001$. From Theorem 1 in [28],
 900 when the vector $r = \eta d$ with $\eta \geq 0$, the objective function f in (3.10) is pseudo-convex on D . In our simulations, the components of the vector d and the constant h_0

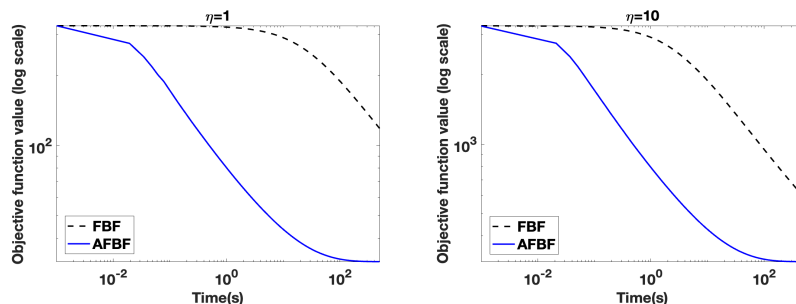


FIGURE 1. Evolution of Algorithm 1 in [35] (called here FBF) and our AFBF algorithm in function values along time for two linear fractional programs of the form (3.10) with data generated randomly, $\eta = 1$ and $\eta = 10$, and dimension $n = 10^6$.

901

902 were drawn independently from a standard normal distribution $\mathcal{N}(0, 1)$, vector r was

903 chosen as $r = \eta d$, with $\eta > 0$, vector h was taken as a perturbation of vector d , i.e.,
 904 $h = d + 0.01\nu$. Vector ν and constant d_0 were generated from a uniform distribution.
 905 Moreover, we chose the starting point x_0 as $x_0 = \text{proj}_D(t)$, vector t being generated
 906 from a standard normal distribution $\mathcal{N}(0, 1)$. The results are displayed in Figure 1,
 907 where we plot the evolution of function values along time (in sec). Note that, AFBF
 908 is faster than Algorithm 1 from [35] (named here FBF) for chosen values of η .

909 **6. Conclusions.** In this paper, we have addressed the problem of finding a zero
 910 of a pseudo-monotone operator. We have made the assumption that this operator
 911 can be split as a sum of three operators: the first continuous operator A satisfies a
 912 generalized Lipschitz inequality, the second operator B is Lipschitzian, and the third
 913 one C is maximally monotone. For solving this challenging problem, our solution
 914 relied upon the forward-backward-forward algorithm, which requires however the use
 915 of an iteration-dependent stepsize. In this context, we designed two novel adaptize
 916 stepsize strategies. We also derived asymptotic sublinear convergence properties un-
 917 der the considered assumptions. Additionally, when $A + B + C$ satisfies a uniform
 918 pseudo-monotonicity condition, the convergence rate becomes even linear. Prelimi-
 919 nary numerical results confirm the good performance of our algorithm.

920 For future research, it would be intriguing to investigate the possibility of achiev-
 921 ing more precise convergence rates. For instance, in Example 3.1, when $g = 0$ and
 922 $L = I_n$, [32] introduces a universal gradient method with a convergence rate of order
 923 $O(\epsilon^{-2/(1+\nu)})$ for the convex (i.e., maximally monotone) case, where ν is the constant
 924 from Definition 2.2 (note that $\mu = 2\nu$ in this scenario). Conversely, in the noncon-
 925 vex (i.e., nonmonotone) case under the same settings, [38] examines a gradient-type
 926 method with an adaptive stepsize and achieves a convergence rate of order $O(\epsilon^{-\frac{1+\nu}{\nu}})$
 927 in the norm of the gradient. On the other hand, the convergence rate obtained in
 928 Theorem 4.11 within the general nonmonotone framework we considered is of order
 929 $O(\epsilon^{-2/\nu})$ in the norm of the gradient, which is not as favorable as the rate in [38].

930

REFERENCES

- 931 [1] S. ADLY, L. BOURDIN, AND F. CAUBET, *On a decomposition formula for the proximal operator*
 932 *of the sum of two convex functions*, Journal of Convex Analysis, 26 (2019), pp. 699–718.
 933 [2] C. AHOLT, S. AGARWAL, AND R. THOMAS, *A qcqp approach to triangulation*, in Computer
 934 Vision – ECCV 2012, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, eds.,
 935 Berlin, Heidelberg, 2012, Springer Berlin Heidelberg, pp. 654–667.
 936 [3] A. ALACA OGLU, D. KIM, AND S. WRIGHT, *Extending the reach of first-order algorithms for*
 937 *nonconvex min-max problems with cohypomonotonicity*, 2024, [https://arxiv.org/abs/2402.](https://arxiv.org/abs/2402.05071)
 938 05071. Preprint.
 939 [4] D. AUSSSEL, *Subdifferential properties of quasiconvex and pseudoconvex functions: unified ap-*
 940 *proach*, Journal of Optimization Theory and Applications, 97 (1998), pp. 29–45.
 941 [5] H. BAUSCHKE AND P. COMBETTES, *Convex analysis and monotone operator theory in Hilbert*
 942 *spaces*, Springer, second ed., 2019.
 943 [6] J. M. BORWEIN AND A. S. LEWIS, *Convex analysis and nonlinear optimization: theory and*
 944 *examples*, Springer Science and Business Media, 2006. New York.
 945 [7] R. I. BOT, E. R. CSETNEK, AND P. T. VUONG, *The forward-backward-forward method from*
 946 *continuous and discrete perspective for pseudo-monotone variational inequalities in hilbert*
 947 *spaces*, European Journal of Operational Research, 287 (2020), pp. 49–60.
 948 [8] L. BRICENO-ARIAS AND D. DAVIS, *Forward-backward-half forward algorithm for solving mono-*
 949 *tone inclusions*, SIAM Journal on Optimization, 28 (2018), pp. 2839–2871.
 950 [9] F. E. BROWDER, *Convergence theorems for sequences of nonlinear operators in banach spaces*,
 951 *Mathematische Zeitschrift*, 100 (1967), pp. 201–225.
 952 [10] Y. CAI AND W. ZHENG, *Accelerated single-call methods for constrained min-max optimization*,
 953 in International Conference on Learning Representations, 2022.
 954 [11] A. CAMBINI, J.-P. CROUZEIX, AND L. MARTEIN, *On the pseudoconvexity of a quadratic frac-*
 955 *tional function*, Optimization, 51 (2002), pp. 677–687.

- 956 [12] L. CAROSI AND L. MARTEIN, *On the pseudoconvexity and pseudolinearity of some classes of*
957 *fractional functions*, Optimization, 56 (2007), pp. 385–398.
- 958 [13] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with*
959 *applications to imaging*, J. Mathematical Imaging and Vision, 40 (2011), pp. 120–145.
- 960 [14] R. CHEN AND A. L. LIU, *A distributed algorithm for high-dimension convex quadratically con-*
961 *strained quadratic programs*, Comput. Optim. and Appl., 80 (2021), pp. 781–830.
- 962 [15] P. L. COMBETTES AND J.-C. PESQUET, *Primal-dual splitting algorithm for solving inclusions*
963 *with mixtures of composite, lipschitzian, and parallel-sum type monotone operators*, Set-
964 *Valued*, 20 (2012), pp. 307–330.
- 965 [16] L. CONDAT, *A primal-dual splitting method for convex optimization involving Lipschitzian,*
966 *proximable and linear composite terms*, Journal of Optimization Theory and Applications,
967 158 (2013), pp. 460–479.
- 968 [17] D. DAVIS AND W. YIN, *A three-operator splitting scheme and its optimization applications*,
969 *Set-Valued and Variational Analysis*, 25 (2017), pp. 829–858.
- 970 [18] A. DE MAIO, Y. HUANG, D. P. PALOMAR, S. ZHANG, AND A. FARINA, *Fractional qcqp with*
971 *applications in ml steering direction estimation for radar detection*, IEEE Transactions on
972 *Signal Processing*, 59 (2011), pp. 172–185.
- 973 [19] J. DIAKONIKOLAS, C. DASKALAKIS, AND M. JORDAN, *Efficient methods for structured*
974 *nonconvex-nonconcave min-max optimization*, in International Conference on Artificial
975 *Intelligence and Statistics*, vol. 130, 2021, pp. 2746–2754.
- 976 [20] GUROBI, *Gurobi optimizer reference manual*. <https://www.gurobi.com>.
- 977 [21] A. HASSOUNI AND A. JADDAR, *On pseudoconvex functions and applications to global optimiza-*
978 *tion*, in ESAIM, vol. 20, 2007, pp. 138–148.
- 979 [22] O. HINDER, A. SIDFORD, AND N. SOHONI, *Near-optimal methods for minimizing star-convex*
980 *functions and beyond*, in Conference on Learning Theory, vol. 125, 2020, pp. 1894–1938.
- 981 [23] Y. HUANG AND D. P. PALOMAR, *Randomized algorithms for optimal solutions of double-sided*
982 *qcqp with applications in signal processing*, IEEE Transactions on Signal Processing, 62
983 (2014), pp. 1093–1108.
- 984 [24] X. JIANG AND L. VANDENBERGHE, *Bregman three-operator splitting methods*, Journal of Opti-
985 *mization Theory and Applications*, 196 (2023), pp. 936–972.
- 986 [25] S. KARAMARDIAN AND S. SCHAIBLE, *Seven kinds of monotone maps*, Journal of Optimization
987 *Theory and Applications*, 66 (1990), pp. 37–46.
- 988 [26] G. R. LANCKRIET, N. CRISTIANINI, P. L. BARTLETT, L. E. GHAOUI, AND M. JORDAN, *Learning*
989 *the kernel matrix with semi-definite programming*, Journal of Machine Learning Research,
990 5 (2004), pp. 27–72.
- 991 [27] O. L. MANGASARIAN, *Pseudo-convex functions*, SIAM Journal on Control and Optimization, 3
992 (1965), pp. 281–290.
- 993 [28] L. MARTEIN AND L. CAROSI, *The sum of a linear and a linear fractional function: pseudocon-*
994 *vexity on the nonnegative orthant and solution methods*, Bulletin of Malaysian Mathemat-
995 *ical Sciences Society*, 35 (2012), pp. 591–599.
- 996 [29] P. MERTIKOPOULOS, C. PAPADIMITRIOU, AND G. PILIOURAS, *Cycles in adversarial regularized*
997 *learning*, in ACM-SIAM Symposium on Discrete Algorithms, 2018, pp. 2703–2717.
- 998 [30] I. NECOARA AND F. CHOROBURA, *Efficiency of stochastic coordinate proximal gradient methods*
999 *on nonseparable composite optimization*, Mathematics of Operations Research, (2024).
- 1000 [31] A. NEMIROVSKI, *Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with*
1001 *Lipschitz continuous monotone operators and smooth convex-concave saddle point prob-*
1002 *lems*, SIAM Journal on Optimization, 15 (2004), pp. 229–251.
- 1003 [32] Y. NESTEROV, *Universal gradient methods for convex optimization problems*, Mathematical
1004 *Programming*, 152 (2015), pp. 381–404.
- 1005 [33] T. PETHICK, P. LATAFAT, P. PATRINOS, O. FERCOQ, AND V. CEVHER, *Escaping limit cycles:*
1006 *global convergence for constrained nonconvex-nonconcave minimax problems*, in Interna-
1007 *tional Conference on Learning Representations*, 2022.
- 1008 [34] S. SCHAIBLE, *Quasiconvex, pseudoconvex, and strictly pseudoconvex quadratic functions*, Jour-
1009 *nal of Optimization Theory and Application*, 35 (1981), pp. 303–338.
- 1010 [35] D. V. THONG AND P. T. VUONG, *Modified tseng’s extragradient methods for solving pseudo-*
1011 *monotone variational inequalities*, Optimization, 68 (2019), pp. 2207–2226.
- 1012 [36] B. TONGNOI, *A modified tseng’s algorithm with extrapolation from the past for pseudo-*
1013 *monotone variational inequalities*, Taiwanese J. of Mathematics, 28 (2024), pp. 187–210.
- 1014 [37] P. TSENG, *A modified forward-backward splitting method for maximal monotone mappings*,
1015 *SIAM Journal on Control and Optimization*, 38 (2000), pp. 431–446.
- 1016 [38] M. YASHTINI, *On the global convergence rate of the gradient descent method for functions with*
1017 *holder continuous gradients*, Optimization Letters, 10 (2016), pp. 1361–1370.