Primal-dual regularized SQP and SQCQP type methods for convex programming and their complexity analysis

M. Marques Alves * R.D.C. Monteiro † Benar F. Svaiter‡

May 9, 2014

Abstract

This paper presents and studies the iteration-complexity of two new inexact variants of Rockafellar's proximal method of multipliers (PMM) for solving convex programming (CP) problems with a finite number of functional inequality constraints. In contrast to the first variant which solves convex quadratic programming (QP) subproblems at every iteration, the second one solves convex constrained quadratic QP subproblems. Their complexity analysis are performed by: a) viewing the original CP problem as a monotone inclusion problem (MIP); b) proposing a large-step inexact higher-order proximal extragradient framework for MIPs, and; c) showing that the above two PMM variants are just instances of this framework.

2000 Mathematics Subject Classification: 90C25, 90C30, 47H05.

Key words: convex programming, nonlinear programming, complexity, extragradient, maximal monotone operator, proximal method of multipliers, ergodic convergence, hybrid, SQP, SQCQP.

1 Introduction

In this paper we study the computational complexity of two new inexact variants of Rockafellar's proximal method of multipliers (PMM) [16] for solving the convex programming problem

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & g(x) \leq 0 \end{array} \tag{1}$$

where $f_0 : \mathbb{R}^n \to \mathbb{R}$ and the components of $g = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth and convex everywhere functions.

*Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (maicon.alves@ufsc.br). The work of this author was partially supported by CNPq grants no. 305414/2011-9 and 406250/2013-8.

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by CNPq Grant 406250/2013-8 and NSF Grant CMMI-1300221.

‡IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil (benar@impa.br). The work of this author was partially supported by CNPq grants no. 474996/2013-1, 302962/2011-5, FAPERJ grant E-26/102.940/2011, and PRONEX-Optimization.

If $x^* \in \mathbb{R}^n$ is an optimal solution of (1) and $y^* \in \mathbb{R}^m$ is a Lagrange multiplier, then the pair (x^*, y^*) is known to be a saddle-point of

$$\min_{x' \in \mathbb{R}^n} \max_{y' \in \mathbb{R}^m_+} f_0(x') + \langle g(x'), y' \rangle. \tag{2}$$

PMM is essentially the proximal point method applied to (2). More specifically, given a sequence of stepsizes $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an initial point $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$, the PMM computes the k-th iterate (x_k, y_k) as being unique saddle-point of

$$\min_{x' \in \mathbb{R}^n} \max_{y' \in \mathbb{R}^m_+} f_0(x') + \langle g(x'), y' \rangle + \frac{1}{2\lambda} \left[\|x' - x\|^2 - \|y' - y\|^2 \right] \tag{3}$$

where $(x, y) = (x_{k-1}, y_{k-1})$. Clearly, apart from its objective function being strongly convex-concave, the above problem is as hard as (2).

The two variants studied in this paper are inexact versions of the PPM in the sense that they solve inexact versions of (3) in which the functions f_i's are approximated by either affine and/or quadratic functions. More specifically, the first variant solves at every iteration the inexact prox-subproblem obtained by replacing the functions f_0 and g in (3) by

$$\begin{aligned} Q_z(x') &:= f_0(x) + \langle \nabla f_0(x), x' - x \rangle + \frac{1}{2} \left\langle x' - x, \left(\nabla^2 f_0(x) + \sum_{i=1}^m y_i^+ \nabla^2 f_i(x) \right) (x' - x) \right\rangle, \\ \ell_x(x') &:= g(x) + \nabla g(x)^T (x' - x), \end{aligned}$$

respectively, where $z := (x, y)$, $y_i^+ := \max\{0, y_i\}$ and $\nabla g(x) := [\nabla f_1(x), \ldots, \nabla f_m(x)]$. Note that the first variant approximates f_0 by a quadratic function and f_i, $i = 1, \ldots, m$, by their first-order affine approximation at x. On the other hand, the second variant solves at every iteration the (inexact) prox-subproblem obtained by approximating all the functions f_i's in (3) by their second-order quadratic approximations at x, i.e.,

$$f_i(x') \approx f_i(x) + \langle \nabla f_i(x), x' - x \rangle + \frac{1}{2} \langle x' - x, \nabla^2 f_i(x)(x' - x) \rangle \quad \forall i = 0, \ldots, m.$$

Both variants then use the saddle-point of the inexact prox-subproblem to obtain the next iterate by performing an extragradient step as prescribed by the hybrid proximal extragradient (HPE) method which we will discuss in more detail below.

The first variant above is related to the stabilized sequential quadratic programming (sSQP) method introduced in [25] and thoroughly studied in [10] (see also the survey [8]). More specifically, the sSQP method computes the next iterate as the unique saddle-point of

$$\min_{x' \in \mathbb{R}^n} \max_{y' \in \mathbb{R}^m_+} Q_z(x') + \langle \ell_x(x'), y' \rangle - \frac{1}{2\lambda} \|y' - y\|^2. \tag{4}$$

Note that the above subproblem only adds a quadratic regularization term to the dual space while the one for the first PPM variant adds quadratic regularization terms to both the primal and the dual spaces. Note also that the sSQP method does not perform an extragradient step as the above PMM variants do. Finally, we note that, in contrast to the sSQP method which computes λ as function of the iterate $z = (x, y)$, the above two PMM variants computes λ by using a large step strategy as described in [14] (see also [13]). The fact that both the sSQP method and the first PMM variant use the same type of approximation to construct the inner subproblems leads us to refer to the latter one as the primal-dual regularized extragradient SQP (re-SQP) method.

On the other hand, the second PMM variant above is closer to the sequential quadratically constrained quadratic programming (SQCQP) type methods (see [1, 2, 9, 19] and the references therein) in that it also solves quadratically constrained quadratic programming subproblems at every iteration. However, in contrast to the second PMM variant, the above proposed SQCQP-type methods do not use regularizations or extragradient steps. In view of these observations, we will refer to the second PMM variant as the primal-dual regularized extragradient SQCQP (re-SQCQP).

A convenient setting for analyzing the aforementioned methods/variants is that of the *monotone inclusion problem (MIP)*, namely: finding z such that

$$0 \in T(z) \tag{5}$$

where T is a maximal monotone point-to-set operator. Our problem of interest, namely (1), is well-known to be equivalent to the MIP

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in T(x,y) = \begin{pmatrix} \nabla f_0(x) + \nabla g(x) y \\ -g(x) + N_{\mathbb{R}^m_+}(y) \end{pmatrix}. \tag{6}$$

A broad class of optimization, saddle-point (SP), equilibrium and variational inequality problems can be posed as MIPs. The proximal point method (PPM), proposed by Rockafellar [17], is a classical iterative scheme for solving the MIP which generates a sequence $\{z_k\}$ according to

$$\|z_k - (\lambda_k T + I)^{-1}(z_{k-1})\| \le e_k, \qquad \sum_{k=1}^{\infty} e_k < \infty.$$

This method has been used as a generic framework for the design and analysis of several implementable algorithms.

New inexact versions of the proximal point method which uses instead relative error criteria were proposed by Solodov and Svaiter [20, 21, 22, 23]. In this article, we will use one of these variants, namely, the hybrid proximal-extragradient (HPE) method in [20], to develop and analyze new algorithms for (1). We now briefly discuss this framework. The *exact* proximal point iteration from z with step-size $\lambda > 0$ is given by $z_+ = (\lambda T + I)^{-1}(z)$, which is equivalent to

$$v \in T(z_+), \qquad \lambda v + z_+ - z = 0. \tag{7}$$

In each step of the HPE method, the above *proximal system* is to be solved inexactly with $(z, \lambda) = (z_{k-1}, \lambda_k)$ to obtain $z_k = z_+$ as follows. For a given constant $\sigma \in [0, 1)$, a triple $(\tilde{z}, v, \varepsilon) = (\tilde{z}_k, v_k, \varepsilon_k)$ is found such that

$$v \in T^{\varepsilon}(\tilde{z}), \qquad \|\lambda v + \tilde{z} - z\|^2 + 2\lambda\varepsilon \le \sigma^2 \|\tilde{z} - z\|^2, \tag{8}$$

where T^{ε} denotes the ε-enlargement [4] of T. (It has the property that $T^{\varepsilon}(z) \supset T(z)$ for each z.) Note that this construction relaxes both the inclusion and the equation in (7). Finally, instead of choosing $\tilde{z}$ as the next iterate z_+, the HPE method computes the next iterate z_+ by means of the *extragradient* step $z_+ = z - \lambda v$.

Iteration complexity bounds for the HPE method were established in [13] and they depend on the distance of the initial iterate to the solution set instead of the diameter of the feasible set. Applications of the HPE method as a framework to the iteration-complexity analysis of several zero-order (resp., first-order) methods for solving monotone variational inequalities and MIPs (resp., saddle-point problems) are discussed in [13] and in the subsequent papers [12, 15]. More specifically, by viewing Korpelevich's method [11] as well as Tseng's modified forward-backward splitting (MF-BS) method [24] as special cases of the HPE framework, the authors have established in [12, 13] the pointwise and ergodic iteration-complexities of these methods applied to either: monotone variational inequalities, MIPs consisting of the sum of a Lipschitz continuous monotone map and a maximal monotone operator with an easily computable resolvent, and convex-concave saddle-point problems.

The HPE framework was also used to study the iteration-complexity of first-order (second-order in the context of optimization and/or saddle-point problems) methods for solving the monotone nonlinear equation (see Section 7 of [13]), the monotone smooth variational inequality and, more generally, the MIP (5) where T is the sum of a smooth monotone map F and a maximal monotone point-to-set operator B (see [14]). The latter paper introduces a large-step subclass of the HPE framework and derives iteration-complexity results for it. It also discusses a first-order inexact (Newton-like) variant of the PPM which is shown to be an instance of the latter large-step HPE framework. More specifically, the latter variant computes at each iteration an approximate solution of the first-order approximation of (7) with $T = F+B$, and F and B as mentioned above, obtained by linearizing F at z and uses it to perform an extragradient step as prescribed by the HPE framework as long as the large-step condition is satisfied; otherwise, the stepsize is revised and the above procedure is repeated. Finally, pointwise and ergodic iteration-complexity results are derived for this first-order method as a by-product of the general complexity results for the large-step HPE framework.

This paper extends the analysis of [14], which only considers methods based on first-order approximations of F, to methods based on approximations of F of order larger than zero. This is accomplished in a unified manner by considering a more general large-step inexact higher-order proximal extragradient framework based on the latter approximations of F. The latter extended framework is then used to derive pointwise and ergodic iteration-complexity bounds for the two PMM variants mentioned earlier. Although the bounds derived for both methods depend on the tolerances in a similar manner, it is shown that the one for the second variant does not depend on the magnitude of the Hessians of the objective and constraints functions of (1). On the other hand, this advantage of the second PMM variant comes at the expense of solving slightly more complicated subproblems which involve the Hessians of the constraint functions.

Organization of the paper. Section 2 reviews some basic properties of ε-enlargements of maximal monotone operators and analyzes the pointwise and ergodic iteration complexity of a generalized large-step HPE (GLS-HPE) framework for finding approximate solutions of monotone inclusion problems. Section 3 is concerned with the problem of finding zeros of a sum of a continuous point-to-point monotone operator F and a point-to-set maximal monotone operator B. A notion of approximation of F at a point z is introduced and the GLS-IPE framework is derived as a special instance of the GLS-HPE framework of Section 2. As in the case of the latter method, pointwise and ergodic complexity results are obtained. Section 4 presents a search procedure to compute a stepsize satisfying a (generalized) large step condition required by the method of Section 3. In Section 5, the theory developed in the previous sections is applied for finding approximate solutions of (inequality) constrained convex programming problems. The primal-dual re-SQP and re-SQCQP methods are studied in Subsections 5.1 and 5.2, respectively.

2 The generalized large-step hybrid proximal extragradient framework

This section is devoted to the introduction and analysis of a generalized large-step HPE framework. It is presented in Subsection 2.2 along with pointwise and ergodic-complexity estimates, in the spirit of the ones obtained in [13] for the large-step HPE method. The first subsection introduces some basic facts and concepts about maximal monotone operators, subdifferentials of convex functions and enlargments.

Notation. Throughout this paper, we assume that $\mathbb{E}$ is a (real) finite dimensional vector space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. For $t>0$, we let $\log^+(t) := \max\{\log(t), 0\}$.

2.1 The ε-subdifferential and ε-enlargement of monotone operators

Given a set-valued operator $S : \mathbb{E} \rightrightarrows \mathbb{E}$, its *graph* and *domain* are taken respectively as

$$\mathrm{Gr}(S) = \{(z,v) \in \mathbb{E}\times\mathbb{E} \,|\, v \in S(z)\}, \quad \mathrm{Dom}(S) = \{z \in \mathbb{E} \,|\, S(z) \neq \emptyset\}.$$

In this work we are concerned with algorithms for solving inclusion problems modeled by maximal monotone operators, a special class of set-valued operators with several applications in applied mathematics and optimization.

An operator $T : \mathbb{E} \rightrightarrows \mathbb{E}$ is *monotone* if

$$\langle v - \tilde{v}, z - \tilde{z}\rangle \geq 0 \quad \text{whenever} \quad (z,v), (\tilde{z},\tilde{v}) \in \mathrm{Gr}(T).$$

It is *maximal monotone* if it is monotone and maximal in the following sense: if $S : \mathbb{E} \rightrightarrows \mathbb{E}$ is monotone and $\mathrm{Gr}(T) \subset \mathrm{Gr}(S)$, then $T = S$. Important examples of such operators are given by the subdifferentials of convex functions (in the sense of convex analysis) and by certain set-valued operators derived from saddle-point formulations of convex programming problems. The latter are discussed and used in Section 5.

Recall that the *ε-subdifferential* of a closed convex function $f : \mathbb{E} \to \overline{\mathbb{R}}$ is defined at $z \in \mathbb{E}$ as

$$\partial_\varepsilon f(z) = \{v \in \mathbb{E} \,|\, f(\tilde{z}) \geq f(z) + \langle v, \tilde{z} - z\rangle - \varepsilon \ \ \forall \tilde{z} \in \mathbb{E}\}\,.$$

When $\varepsilon = 0$, then $\partial f_0(z)$ is denoted by $\partial f(z)$ and is called the *subdifferential* of f at z. The set-valued operator $\partial f_\varepsilon : \mathbb{E} \rightrightarrows \mathbb{E}$ with $\varepsilon > 0$ is an enlargement of ∂f (in the sense that $\partial f(z) \subset \partial f_\varepsilon(z)$ for every $z \in \mathbb{E}$) which has better topological properties than ∂f (see [3]). The simplest example of subdifferential is given by considering indicator functions of closed convex sets. Given a closed convex set $X \subset \mathbb{E}$, its *indicator function* is denoted by δ_X and is defined as

$$\delta_X(z) = \begin{cases} 0, & z \in X \\ \infty, & \text{otherwise,} \end{cases}$$

and its *ε-normal cone* is defined as $N_X^\varepsilon = \partial_\varepsilon \delta_X$. When $\varepsilon = 0$, then N_X^0 is denoted by N_X and is called the *normal cone* of X. The following result will be useful in Section 5.

Proposition 2.1. *For a closed convex cone $K \subset \mathbb{R}^n$ and scalar $\rho \geq 0$, we have:*

$$u \in N_K^\rho(y) \iff y \in K, -u \in K^* \ \ \text{and} \ \ \langle u, y\rangle \geq -\rho$$

where $K^* := \{v \in \mathbb{R}^n \,|\, \langle z, v\rangle \geq 0 \ \ \forall z \in K\}$ is the dual cone of K.

A generalization of the concept of ε-enlargement for arbitrary maximal monotone operators was introduced and studied in [4]. Given $T : \mathbb{E} \rightrightarrows \mathbb{E}$ maximal monotone and $\varepsilon > 0$, the ε-enlargment of T is defined as

$$T^{\varepsilon} : \mathbb{E} \rightrightarrows \mathbb{E}, \quad T^{\varepsilon}(z) = \{v \in \mathbb{E} \,|\, \langle \tilde{v} - v, \tilde{z} - z\rangle \geq -\varepsilon \;\; \forall (\tilde{z}, \tilde{v}) \in \mathrm{Gr}(T)\}.$$

If $f : \mathbb{E} \to \overline{\mathbb{R}}$ is closed and convex, then it holds that $\partial_{\varepsilon} f(z) \subset (\partial f)^{\varepsilon}(z)$ for all $z \in \mathbb{E}$.

The following summarizes some useful properties of T^{ε}.

Proposition 2.2. *Let $T, T' : \mathbb{E} \rightrightarrows \mathbb{E}$. Then,*

(a) if $\varepsilon_1 \leq \varepsilon_2$, then $T^{\varepsilon_1}(z) \subseteq T^{\varepsilon_2}(z)$ for every $z \in \mathbb{E}$;

(b) $T^{\varepsilon}(z) + (T')^{\varepsilon'}(z) \subseteq (T + T')^{\varepsilon + \varepsilon'}(z)$ for every $z \in \mathbb{E}$ and $\varepsilon, \varepsilon' \in \mathbb{R}$;

(c) T is monotone if, and only if, $T \subseteq T^0$;

(d) T is maximal monotone if, and only if, $T = T^0$;

(e) if T is maximal monotone, then $\mathrm{Dom}\, T^{\varepsilon} \subseteq \mathrm{cl}\,(\mathrm{Dom}\, T)$ for any $\varepsilon \geq 0$.

Proof. Statements (a)-(d) can be easily proved using the definition of T^{ε}. The proof of (e) can be found in [5, Corollary 3.8(ii)]. □

2.2 The generalized large-step HPE framework

Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ be maximal monotone operator. The monotone inclusion problem for T consists of finding $z \in \mathbb{E}$ such that

$$0 \in T(z)\,. \tag{9}$$

We also assume throughout this subsection that this problem has a solution, that is, $T^{-1}(0) \neq \emptyset$.

The purpose of this subsection is to study a framework of inexact proximal point methods based on a relative error criterion and a large step condition for approximately solving (9) in the sense that it generates a sequence of triples $\{(z'_k, v'_k, \varepsilon'_k)\} \subset \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ such that $v'_k \in T^{\varepsilon'_k}(z'_k)$ and the measure $\max\{\|v'_k\|\,, \varepsilon'_k\}$ converges to zero. Moreover, by considering two ways of generating the above sequence, namely the pointwise way or the ergodic way, it derives two convergence rate results establishing how fast the quantities $\|v'_k\|$ and ε'_k converge to zero. Clearly, given tolerances $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, these results can be easily translated into iteration-complexity results to compute a point $(z, v, \varepsilon) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ satisfying

$$v \in T^{\varepsilon}(z), \quad \|v\| \leq \bar{\rho}, \quad \varepsilon \leq \bar{\varepsilon}. \tag{10}$$

For the sake of shortness, we have omitted the statements of these iteration-complexity results on the presentation of this subsection and Section 3. However, some iteration-complexity results will be stated in Section 5 in the context of two specific applications of the method presented in Section 3.

For the MIP (9), the exact proximal point iteration from z with stepsize $\lambda > 0$ is the unique solution $\tilde{z}$ of the inclusion

$$0 \in (\lambda T + I)(\tilde{z}) - z = \lambda T(\tilde{z}) + \tilde{z} - z, \tag{11}$$

or equivalently, $\tilde{z} = (\lambda T + I)^{-1}(z)$. Note that $\tilde{z}$ and $v = (z - \tilde{z})/\lambda$ is also the unique solution $(\tilde{z}, v)$ of the inclusion/equation

$$v \in T(\tilde{z}), \qquad \lambda v + \tilde{z} - z = 0. \tag{12}$$

The method we are interested in studying in this subsection is based on the following notion of approximate solution of (12) which was introduced in [20] and extensively used in the analysis of proximal-point based methods.

Definition 2.3. *Given $\hat\sigma \geq 0$, the triple $(\tilde z, v, \varepsilon)$ is said to be a $\hat\sigma$-approximate solution of* (12) *if*

$$v \in T^{\varepsilon}(\tilde z), \quad \|\lambda v + \tilde z - z\|^2 + 2\lambda\varepsilon \leq \hat\sigma^2 \|\tilde z - z\|^2. \tag{13}$$

We now make two remarks about Definition 2.3. First, in view of the discussion preceding the above definition, if the resolvent operator $(\lambda T + I)^{-1}(\cdot)$ can be evaluated exactly then a $\hat\sigma$-solution $(\tilde z, v, \varepsilon)$ of (12) for any $\hat\sigma \geq 0$ can be obtained by setting $\tilde z = (\lambda T + I)^{-1}(z)$, $v = (z - \tilde z)/\lambda$ and $\varepsilon = 0$. Second, the error criterion (13) is relative in the sense that it requires the residual $\lambda v + \tilde z - z$ and the tolerance ε to be small relative to $\|\tilde z - z\|$.

Our generalized large-step HPE framework depends of a special family of functions which we now introduce.

Definition 2.4. *Denote by Ψ the class of all continuous functions $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying the following conditions:*

(a) ψ is strictly increasing;

(b) $\psi(0) = 0$ and $\psi(t) > 0$ for every $t > 0$;

(c) there exists $p \geq 1/2$ (possibly depending on ψ) such that the function $t > 0 \mapsto 1/\psi(t^p)$ is convex.

We will now state the framework that will be the main subject of study in this subsection.

Generalized large-step (GLS) HPE framework (GLS-HPE framework):

(0) Let $z_0 \in \mathbb{E}$, $\theta > 0$, $0 \leq \sigma < 1$ and $\psi \in \Psi$ be given and set $k = 1$;

(1) if $0 \in T(z_{k-1})$, then **STOP**. Otherwise, choose stepsize $\lambda_k > 0$, tolerance $\sigma_k \in [0, \sigma]$ and σ_k-approximate solution $(\tilde z_k, v_k, \varepsilon_k)$ of (12) with $(\lambda, z) = (\lambda_k, z_{k-1})$, i.e.,

$$v_k \in T^{\varepsilon_k}(\tilde z_k), \quad \|\lambda_k v_k + \tilde z_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma_k^2 \|\tilde z_k - z_{k-1}\|^2, \tag{14}$$

such that

$$\lambda_k \psi(\|\tilde z_k - z_{k-1}\|) \geq \theta; \tag{15}$$

(2) define $z_k = z_{k-1} - \lambda_k v_k$, set $k \leftarrow k + 1$ and go to step 1.

end

We now make some remarks about the GLS-HPE framework. First, the above framework with condition (15) removed, or equivalently with $\theta = 0$, is equivalent to the HPE method introduced in [20] and whose global convergence rate is studied in [13]. Hence, the generalized large-step HPE framework is a special case of the HPE framework. Second, the large-step HPE framework of [13] corresponds to the particular case of the above framework in which $\psi(t) = t$ for all $t \in \mathbb{R}_+$. Third, in view of the first remark following Definition 2.3, relations (44) and (45), and Lemma 4.3(c), we easily see that a stepsize λ_k and a triple $(\tilde{z}_k, v_k, \varepsilon_k)$ satisfying (14) and (15) always exist. Fourth, step 1 does not specify how to compute such λ_k and $(\tilde{z}_k, v_k, \varepsilon_k)$. Their computation depends on the instance of the framework under consideration. Fifth, Definition 2.4(c) is needed only for the iteration-complexity analysis of the GLS-HPE framework. Hence, the third remark (as well as all the results of Section 4) holds without assuming Definition 2.4(c).

Our goal in the remaining part of this subsection is to describe global convergence rate bounds for the whole GLS-HPE framework. We first review a useful result (see for example [13] for its proof) which holds for the whole HPE framework, and hence for the above framework.

Proposition 2.5. *Let $\{z_k\}$ and $\{\tilde{z}_k\}$ be generated by the GLS-HPE framework, or more generally, by the HPE framework (i.e., the more general case of the above framework in which the large-step condition (15) is removed). Then, for any $z^* \in T^{-1}(0)$, the sequence $\{\|z^* - z_k\|\}$ is nonincreasing and*

$$\|z^* - z_0\|^2 \geq \sum_{k=1}^{\infty}(1-\sigma_k^2)\,\|\tilde{z}_k - z_{k-1}\|^2 \geq (1-\sigma^2)\sum_{k=1}^{\infty}\|\tilde{z}_k - z_{k-1}\|^2\,. \tag{16}$$

As a consequence, for every $k \geq 1$, we have

$$\|z_k - z_0\| \leq 2d_0 \tag{17}$$

where d_0 denotes the distance of z_0 to $T^{-1}(0)$.

We are now ready to present two convergence rate results for the GLS-HPE framework. The first one, namely Theorem 2.6, which we refer to as the pointwise one, describes global convergence rate bounds for the best among the iterates $\tilde{z}_1, \ldots, \tilde{z}_k$. The second one, namely Theorem 2.9, which we refer to as the ergodic one, describes global convergence rate bounds for an ergodic sequence associated with $\{\tilde{z}_k\}$.

Theorem 2.6. *(pointwise convergence rate) Consider the sequences $\{z_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$ and $\{\varepsilon_k\}$ generated by the generalized large-step HPE framework. Then, for every $k \geq 1$, $v_k \in T^{\varepsilon_k}(\tilde{z}_k)$ and there exists an index $i \leq k$ such that*

$$\|v_i\| \leq \frac{(1+\sigma)d_0}{\theta\sqrt{1-\sigma^2}\sqrt{k}}\,\psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}}\right),$$

$$\varepsilon_i \leq \frac{\sigma^2 d_0^2}{2\theta(1-\sigma^2)k}\,\psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}}\right),$$

where d_0 is the distance of z_0 to $T^{-1}(0)$.

Proof. Let z^* be such that $d_0 = \|z_0 - z^*\|$. First observe that Proposition 2.5 implies that, for every $k \in \mathbb{N}$, there exists $i \leq k$ such that

$$\|\tilde{z}_i - z_{i-1}\|^2 \leq \frac{d_0^2}{k(1-\sigma^2)}. \tag{18}$$

Using the fact that $(\tilde{z}_i, v_i, \varepsilon_i)$ is a σ_i-approximate solution of (12) with $(\lambda, z) = (\lambda_i, z_{i-1})$, i.e., that (14) holds with $i = k$, and the fact that $\sigma_i \leq \sigma$, we conclude that

$$\lambda_i\|v_i\| \leq \|\lambda_i v_i + \tilde{z}_i - z_{i-1}\| + \|\tilde{z}_i - z_{i-1}\| \leq (1+\sigma)\|\tilde{z}_i - z_{i-1}\|$$

and

$$2\lambda_i\varepsilon_i \le \sigma^2\|\tilde z_i - z_{i-1}\|^2.$$

Multiplying the above two inequalities by $\psi(\|\tilde z_i - z_{i-1}\|)$ and using (15) we get

$$\theta\|v_i\| \le (1+\sigma)\psi(\|\tilde z_i - z_{i-1}\|)\|\tilde z_i - z_{i-1}\|, \qquad 2\theta\varepsilon_i \le \sigma^2\psi(\|\tilde z_i - z_{i-1}\|)\|\tilde z_i - z_{i-1}\|^2.$$

The conclusion of the theorem now follows immediately from (18), Definition 2.4(a) and the above two inequalities. □

We will now derive convergence rate bounds for an ergodic sequence associated with the sequence $\{\tilde z_k\}$ generated by the GLS-HPE framework. The ergodic sequence $\{\tilde z_k^a\}$ associated with $\{\tilde z_k\}$ is defined as

$$\tilde z_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \tilde z_i, \qquad \text{where} \qquad \Lambda_k := \sum_{i=1}^k \lambda_i. \tag{19}$$

Define also

$$v_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i v_i, \qquad \varepsilon_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i(\varepsilon_i + \langle \tilde z_i - \tilde z_k^a, v_i - v_k^a\rangle). \tag{20}$$

We will make use of the following general result which holds for any instance of the HPE (and hence GLS-HPE) framework and whose proof follows immediately from Theorem 4.7 of [13].

Proposition 2.7. *For every $k \ge 1$,*

$$v_k^a \in T^{\varepsilon_k^a}(\tilde z_k^a)$$

and the following bounds hold:

$$\|v_k^a\| \le \frac{2d_0}{\Lambda_k}, \qquad 0 \le \varepsilon_k^a \le \frac{2\eta d_0^2}{\Lambda_k}, \tag{21}$$

where d_0 is the distance of z_0 to $T^{-1}(0)$, and

$$\eta := 1 + \frac{\sigma}{\sqrt{(1-\sigma^2)}}\,. \tag{22}$$

The convergence rate bounds stated in Proposition 2.7 are in terms of Λ_k. In what follows, we will derive the desired ones in terms of k by obtaining a lower on Λ_k in terms of an appropriate function of k. With this goal in mind, we now establish the following technical result.

Lemma 2.8. *For any $\psi \in \Psi$ and $C > 0$, the optimal value of the minimization problem*

$$\begin{array}{ll} \min & \sum_{i=1}^k \dfrac{1}{\psi(t_i)} \\ \text{s.t.} & \sum_{i=1}^k t_i^2 \le C, \quad t_i > 0, \; i = 1,\ldots,k \end{array} \tag{23}$$

is equal to

$$\frac{k}{\psi\left(\sqrt{C/k}\right)}.$$

Proof. Problem (23) clearly has the same optimal value as the problem

$$\begin{array}{ll} \min & \sum_{i=1}^k \dfrac{1}{\psi(t_i^p)} \\ \text{s.t.} & \sum_{i=1}^k t_i^{2p} \le C, \quad t_i > 0,\ i = 1, \dots, k, \end{array} \tag{24}$$

which is convex due to Definition 2.4(c). It is easy to see that the continuity of ψ and Definition 2.4(b) imply that (24) has an optimal solution. Using the fact that the objective and constraint functions in (24) are convex and invariant under permutations on $(t_1, t_2, \dots, t_k)$, we easily see that that it has an optimal solution of the form $(t_*, \dots, t_*) \in \mathbb{R}^k$, for some $t_* > 0$. Consequently, the minimum value of (24), and hence of (23), is equal to $k/\psi(t_*^p)$. Moreover, in view of Definition 2.4(a), we may also assume that $kt_*^{2p} = C$. Hence, the conclusion of the lemma follows. □

We are now ready to establish the desired ergodic convergence rate bounds for the generalized large-step HPE framework in terms of an appropriate function of k.

Theorem 2.9. *(ergodic convergence rate) Let $\{\lambda_k\}$, $\{\varepsilon_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$ and $\{z_k\}$ be the sequences generated by the generalized large-step HPE framework and consider the ergodic sequences $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ defined according to (19) and (20). Then, for every $k \ge 1$, $v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a)$ and*

$$\|v_k^a\| \le \frac{2d_0}{k\theta}\, \psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}} \right), \qquad 0 \le \varepsilon_k^a \le \frac{2\eta d_0^2}{k\theta}\, \psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}} \right),$$

where η is defined in (22).

Proof. Let $k \in \mathbb{N}$ and define $\tilde{t}_i = \|\tilde{z}_i - z_{i-1}\|$ for $i = 1, \dots, k$. Using the fact that $0 \notin T(z_{i-1})$ and (14), it is easy to see that $\tilde{t}_i > 0$ for $i = 1, \dots, k$. By (16) of Proposition 2.5, we have

$$\sum_{i=1}^k \tilde{t}_i^2 = \sum_{i=1}^k \|\tilde{z}_i - z_{i-1}\|^2 \le \frac{d_0^2}{(1-\sigma^2)}\,, \tag{25}$$

or equivalently, $(\tilde{t}_1, \dots, \tilde{t}_k)$ is feasible for problem (23) with $C = d_0^2/(1-\sigma^2)$. Therefore, using Lemma 2.8, we have

$$\frac{k}{\psi\left(\dfrac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}} \right)} \le \sum_{i=1}^k \frac{1}{\psi(\tilde{t}_i)} \le \frac{1}{\theta} \sum_{i=1}^k \lambda_i = \frac{\Lambda_k}{\theta},$$

where the last inequality follows from (15) and the definition of $\tilde{t}_i$. The conclusion of the theorem now follows immediately from the definition of Λ_k in (19), Proposition 2.7 and the last inequality. □

We have observed in the paragraph following the statement of the GLS-HPE framework that the large-step HPE framework of [14] corresponds to the special case of the GLS-HPE framework in which $\psi(t) = t$ for all $t \ge 0$. In this regards, it is worth mentioning that Theorems 2.5 and 2.7 of [14] correspond to the special cases of Theorems 2.6 and 2.9, respectively, with ψ as described above. Moreover, noting that $\lim_{t\to 0} \psi(t) = 0$, in view of the the continuity of ψ and Definition 2.4(b), it follows that the pointwise and ergodic convergence rates derived in Theorems 2.6 and 2.9 are asymptotically better than the ones obtained in [12] for the HPE method, namely $(v_i, \varepsilon_i) = (\mathcal{O}(1/\sqrt{k}), \mathcal{O}(1/k))$ in the pointwise version, and $(v_k^{\mathrm{a}}, \varepsilon_k^{\mathrm{a}}) = (\mathcal{O}(1/k), \mathcal{O}(1/k))$ in the ergodic version.

3 The GLS-IPE framework

In this section, we consider the MIP

$$0 \in (F + B)(z), \tag{26}$$

where the following assumptions hold:

C.1) $B : \mathbb{E} \rightrightarrows \mathbb{E}$ is a maximal monotone operator;

C.2) $F : \mathrm{Dom}\, F \subseteq \mathbb{E} \to \mathbb{E}$ is monotone and continuous on a closed convex set Ω such that $\mathrm{Dom}\, B \subseteq \Omega \subseteq \mathrm{Dom}\, F$;

C.3) the solution set of (26) is nonempty.

We note also that, under the above assumptions, $F + B$ is a maximal monotone operator whose domain is equal to $\mathrm{Dom}\, B \subseteq \Omega$ (see Proposition A.1 in Appendix A of [12]).

Recall that, for the MIP (26), the exact proximal iteration from z with stepsize $\lambda > 0$ is the unique solution $\tilde{z}$ of the inclusion

$$0 \in \lambda(F + B)(\tilde{z}) + \tilde{z} - z, \tag{27}$$

or equivalently, the $\tilde{z}$-component of the unique solution $(\tilde{z}, v)$ of the inclusion/equation

$$v \in (F + B)(\tilde{z}), \qquad \lambda v + \tilde{z} - z = 0. \tag{28}$$

Our GLS-IPE framework is based on a notion of approximation of F on Ω which we now describe.

First, we introduce the following subclass of the class of functions Ψ.

Definition 3.1. *For a positive scalar $\beta > 0$, we let $\Psi(\beta)$ denote the class of all functions $\psi \in \Psi$ satisfying*

(a) $\psi(\tau t) \le \tau^{\beta}\psi(t)$ for every $t > 0$ and $\tau \ge 1$

We note that for every $\psi \in \Psi(\beta)$ the following relation holds whenever $0 \le \tau \le 1$:

$$\psi(\tau t) \ge \tau^{\beta}\psi(t) \qquad \forall t > 0. \tag{29}$$

The following definition introduces the notion of an approximation of F on Ω.

Definition 3.2. *A triple $(\beta_z, \phi_z, \mathcal{A}_z)$ consisting of a scalar $\beta_z > 0$ and functions $\mathcal{A}_z : \mathrm{Dom}\, A_z \to \mathbb{E}$ and $\phi_z : \mathbb{R}_+ \to \mathbb{R}_+$ is called an approximation of F at z if the following conditions are satisfied:*

a) $\phi_z \in \Psi(\beta_z)$ and $\mathrm{Dom}\, A_z \supset \Omega$ and $\mathcal{A}_z$ is monotone and continuous on Ω;

b) for every $z' \in \Omega$,

$$\|F(z') - \mathcal{A}_z(z')\| \le \phi_z(\|z' - z\|)\|z' - z\|. \tag{30}$$

We now state an instance of the GLS-HPE framework for solving (26) based on the above notion of approximate solution.

GLS-IPE framework: An approximation-based GLS inexact proximal extragradient framework

(0) Let $z_0 \in \mathbb{E}$ and scalars $0 < \sigma_\ell < \sigma_u \le \sigma < 1$ be given and set $k = 1$;

(1) if $0 \in (F + B)(z_{k-1})$, then **STOP**;

(2) otherwise, choose $\hat\sigma_k \ge 0$ and an approximation $(\beta_k, \phi_k, \mathcal{A}_k) = (\beta_{z_{k-1}}, \phi_{z_{k-1}}, \mathcal{A}_{z_{k-1}})$ of F at z_{k-1} such that

$$\sigma_u + \hat\sigma_k \le \sigma, \quad \sigma_\ell(1 + \hat\sigma_k)^{\beta_k} < \sigma_u(1 - \hat\sigma_k)^{\beta_k}; \tag{31}$$

(3) compute stepsize $\lambda_k > 0$ and a triple $(\tilde z_k, u_k, \varepsilon_k) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ such that

$$u_k \in (\mathcal{A}_k + B^{\varepsilon_k})(\tilde z_k), \qquad \|\lambda_k u_k + \tilde z_k - z_{k-1}\|^2 + 2\lambda_k\varepsilon_k \le \hat\sigma_k^2 \|\tilde z_k - z_{k-1}\|^2 \tag{32}$$

and

$$\sigma_\ell \le \lambda_k \phi_k(\|\tilde z_k - z_{k-1}\|) \le \sigma_u; \tag{33}$$

(4) set

$$v_k := F(\tilde z_k) + u_k - \mathcal{A}_k(\tilde z_k), \qquad z_k := z_{k-1} - \lambda_k v_k, \tag{34}$$

let $k \leftarrow k + 1$ and go to step 1.

end

We now make some remarks about the GLS-IPE framework. First, it will be shown in Lemma 3.4 below that, under some mild conditions on $\{\phi_k\}$ (see (40)), the GLS-IPE framework is a special case of the GLS-HPE framework of Section 2. Second, the GLS-IPE framework is a generalization of the inexact NPE method of [14]. Indeed, if F has L-Lipschitz continuous derivative F' on Ω as assumed in [14], then the latter method can be viewed as an instance of the first one by letting the corresponding approximation of F at $z \in \mathbb{E}$ as

$$\beta_z = 1, \quad \phi_z(t) = Lt/2, \quad \mathcal{A}_z(z') = F(z_\Omega) + F'(z_\Omega)(z' - z_\Omega), \quad \forall\, (z', t) \in \mathbb{E} \times \mathbb{R}_+,$$

where z_Ω denotes the orthogonal projection of z onto Ω. Third, two specific ways of choosing the triple $(\beta_z, \phi_z, \mathcal{A}_z)$ as in step 2 of the GLS-IPE framework will be described in Subsections 5.1 and 5.2 in the context of SQP-type methods for constrained convex programming problems. Fourth, Section 4 presents a stepsize search procedure which, given an approximation $(\beta_k, \phi_k, \mathcal{A}_k)$ of F at z_{k-1}, searches for a stepsize $\lambda_k > 0$ and a triple $(\tilde z_k, u_k, \varepsilon_k)$ such that (32) and (33) hold. More specifically, using a black-box (assumed given) which, given $(\beta_k, \phi_k, \mathcal{A}_k)$ and a trial stepsize $\lambda_k > 0$, outputs a triple $(\tilde z_k, u_k, \varepsilon_k)$ satisfying (32) only, the procedure eventually outputs the required $\lambda_k > 0$ and a triple $(\tilde z_k, u_k, \varepsilon_k)$ satisfying (32) and (33) in a finite number of trials. Fifth, if the resolvent operator $(I + \lambda(\mathcal{A}_k + B))^{-1}(\cdot)$ can be exactly evaluated for any $\lambda > 0$, then a black-box which yields a triple satisfying (32) consists of simply choosing the triple $(\tilde z_k, u_k, \varepsilon_k)$ where

$$\tilde z_k = (I + \lambda_k(\mathcal{A}_k + B))^{-1}(z_{k-1}), \quad u_k = \frac{z_{k-1} - \tilde z_k}{\lambda_k}, \quad \varepsilon_k = 0$$

(see also the first remark after Definition 2.3). Sixth, the second inequality in (31) and the fact that $\phi_k \in \Psi(\beta_k)$ (due to step 2 of the GLS-IPE framework and Definition 3.2(a)) will be used only in

Section 4 to derive the complexity of the stepsize search procedure studied there.

Lemma 3.3. *Let* $(\lambda, z) \in \mathbb{R}_{++} \times \mathbb{E}$ *and* $(\beta_z, \phi_z, \mathcal{A}_z)$ *be an approximation of* F *at* z. *Let also* $(\tilde{z}, u, \varepsilon) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ *be such that*

$$u \in (\mathcal{A}_z + B^\varepsilon)(\tilde{z}), \qquad \|\lambda u + \tilde{z} - z\|^2 + 2\lambda\varepsilon \le \hat{\sigma}^2 \|\tilde{z} - z\|^2 \tag{35}$$

and define $v := F(\tilde{z}) + u - \mathcal{A}_z(\tilde{z})$. *Then,*

$$v \in (F + B^\varepsilon)(\tilde{z}) \subseteq (F + B)^\varepsilon(\tilde{z}), \qquad \|\lambda v + \tilde{z} - z\|^2 + 2\lambda\varepsilon \le \Big(\hat{\sigma} + \lambda\phi_z(\|\tilde{z} - z\|)\Big)^2 \|\tilde{z} - z\|^2. \tag{36}$$

Proof. First note that $\tilde{z} \in \Omega \subset \operatorname{Dom} F$ in view of the first inclusion in (35), Proposition 2.2(e) and Assumption C.2, from which we conclude that v is well-defined. The assumption on $(\tilde{z}, u, \varepsilon)$ imply that $u - \mathcal{A}_z(\tilde{z}) \in B^\varepsilon(\tilde{z})$. This together with the definition of v imply the first inclusion in (36), while Proposition 2.2 implies the second inclusion in (36). To simplify the proof of the inequality in (36), define

$$r := \lambda u + \tilde{z} - z, \qquad \tilde{r} := \lambda[F(\tilde{z}) - \mathcal{A}_z(\tilde{z})],$$

and note that the definition of v implies that

$$\lambda v + \tilde{z} - z = r + \tilde{r}. \tag{37}$$

Note that the inequality in (35), the assumption that $(\beta_z, \phi_z, \mathcal{A}_z)$ is an approximation of F at z and (30) imply that

$$\|r\|^2 + 2\lambda\varepsilon \le \hat{\sigma}^2 \|\tilde{z} - z\|^2, \qquad \|\tilde{r}\| \le \lambda\phi_z(\|\tilde{z} - z\|)\|\tilde{z} - z\|.$$

Using the three last relations, we then conclude that

$$\begin{aligned}\|\lambda v + \tilde{z} - z\|^2 + 2\lambda\varepsilon &= \|r + \tilde{r}\|^2 + 2\lambda\varepsilon \le \|r\|^2 + \|\tilde{r}\|^2 + 2\|r\|\,\|\tilde{r}\| + 2\lambda\varepsilon \\ &\le \Big[\hat{\sigma}^2 + (\lambda\phi_z(\|\tilde{z} - z\|))^2 + 2\lambda\hat{\sigma}\phi_z(\|\tilde{z} - z\|)\Big]\,\|\tilde{z} - z\|^2,\end{aligned}$$

which clearly implies the inequality in (36). ☐

The following result shows that the GLS-IPE framework is a special case of the HPE framework, and also gives a sufficient condition for it to be a special case of the GLS-HPE framework.

Lemma 3.4. *Consider the scalars* σ_ℓ *and* σ *and sequences* $\{\lambda_k\}$, $\{\varepsilon_k\}$, $\{\hat{\sigma}_k\}$, $\{z_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$ *and* $\{\phi_k\}$ *as in the GLS-IPE framework applied to* (26), *and define*

$$\theta := \sigma_\ell, \quad \sigma_k := \hat{\sigma}_k + \lambda_k\phi_k(\|\tilde{z}_k - z_{k-1}\|), \quad \forall k \ge 1. \tag{38}$$

Then, the following statements hold:

(a) for each $k \ge 1$, $\sigma_k \in [0, \sigma]$ *and* (14) *hold and*

$$v_k \in (F + B^{\varepsilon_k})\,(\tilde{z}_k); \tag{39}$$

as a consequence, the GLS-IPE framework is a special case of the HPE framework, and hence the conclusions of Proposition 2.5 hold;

(b) if, in addition, there exists $\psi \in \Psi$ (see Definition 2.4) such that

$$\phi_k \leq \psi \quad \forall k \geq 1, \tag{40}$$

then (15) *hold for every* $k \geq 1$.

As a consequence of (a) and (b), every instance of the GLS-IPE framework satisfying condition (40) *is also an instance of the GLS-HPE framework of Section 2 with* $\theta = \sigma_\ell$.

Proof. (a) The statement that $\sigma_k \in [0, \sigma]$ follows from the second inequality in (33), the first inequality in (31) and the definition of σ_k in (38). Condition (14) and relation (39) follow from (32), Lemma 3.3 with $(\lambda, z) = (\lambda_k, z_{k-1})$, $(\beta_z, \phi_z, \mathcal{A}_z) = (\beta_k, \phi_k, \mathcal{A}_k)$ and $(\tilde{z}, u, \varepsilon) = (\tilde{z}_k, u_k, \varepsilon_k)$, where the sequences $\{\beta_k\}$, $\{\mathcal{A}_k\}$ and $\{u_k\}$ are as in the GLS-IPE framework and the definition of σ_k in (38). As a consequence of the previous facts, condition (34) and the first remark after the GLS-HPE framework we conclude the proof of (a).

(b) If there exists $\psi \in \Psi$ satisfying (40), then condition (15) follows from the definition of θ in (38), relation (40) and the first inequality in (33).

The last statement of the lemma follows directly from (a), (b) and the definition of θ in (38). □

The next result gives pointwise and ergodic convergence rate bounds for the GLS-IPE framework under the sufficient condition (40).

Theorem 3.5. *Consider the scalars σ_ℓ and σ and sequences $\{\lambda_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$ and $\{\phi_k\}$ as in the GLS-IPE framework applied to* (26) *and assume that condition* (40) *holds for some $\psi \in \Psi$. Moreover, define the ergodic sequences $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ according to (19) and (20) and let d_0 denote the distance of z_0 to the solution set of* (26). *Then, for every $k \geq 1$:*

(a) (pointwise iteration-complexity) $v_k \in (F + B^{\varepsilon_k})(\tilde{z}_k)$ and there exists an index $i \leq k$ such that

$$\|v_i\| \leq \frac{(1+\sigma)d_0}{\sigma_\ell\sqrt{1-\sigma^2}\sqrt{k}}\,\psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}}\right), \qquad \varepsilon_i \leq \frac{\sigma^2 d_0^2}{2\sigma_\ell(1-\sigma^2)k}\,\psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}}\right);$$

(b) (ergodic iteration-complexity) $v_k^a \in (F + B)^{\varepsilon_k^a}(\tilde{z}_k^a)$ and

$$\|v_k^a\| \leq \frac{2d_0}{k\sigma_\ell}\,\psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}}\right), \qquad 0 \leq \varepsilon_k^a \leq \frac{2\eta d_0^2}{k\sigma_\ell}\,\psi\left(\frac{d_0}{\sqrt{1-\sigma^2}\sqrt{k}}\right),$$

where $\eta > 0$ is as in (22).

Proof. The proof is a direct consequence of Lemma 3.4, Theorem 2.6 and Theorem 2.9. □

We now make a few remarks about Theorem 3.5. First, note that the bound on $\|v_i\|$ is $\mathcal{O}(\psi(1/\sqrt{k})/\sqrt{k})$ while the one on $\|v_k^a\|$ is $\mathcal{O}(\psi(1/\sqrt{k})/k)$, which is better than the first one by a factor of $\sqrt{k}$. However, note that the pointwise residual inclusion $v_k \in (F + B^{\varepsilon_k})(\tilde{z}_k)$ is stronger than the ergodic residual inclusion $v_k^a \in (F + B)^{\varepsilon_k^a}(\tilde{z}_k^a)$, in the sense that $v_k \in (F + B^{\varepsilon_k})(\tilde{z}_k)$ implies $v_k \in (F + B)^{\varepsilon_k}(\tilde{z}_k)$. Second, recalling that the inexact NPE method of [14] can be viewed as a special case of the GLS-IPE framework in which $\phi_z(t) = Lt/2$ (see the paragraph preceding Lemma 3.3), the bounds in Theorem 3.5 with $\psi(t) = Lt/2$ apply to the inexact NPE method of [14]. It turns out that these specialized bounds are identical to the ones obtained in Theorems 3.5 and 3.6 of [14].

4 Search procedure for the stepsize λ

The main goal of this section is to present a search procedure for computing a stepsize $\lambda_k > 0$ and a triple $(\tilde{z}_k, u_k, \varepsilon_k)$ as in step 3 of the GLS-IPE framework. Its presentation is rather technical and can be skipped in a first reading without any loss of continuity. This section starts by describing a slightly more general version of the line search problem which we are interested in solving. It contains two subsections. Subsection 4.1 describes (see the Bracketing and Bisection stages before and after Lemma 4.5) and derives complexity bounds (Lemma 4.6) for a procedure which solves the latter more general line search problem. Subsection 4.2 describes a specialization of the latter procedure (see the Bracketing/Bisection procedure there) for computing a stepsize $\lambda_k > 0$ and a triple $(\tilde{z}_k, u_k, \varepsilon_k)$ as in step 3 of the GLS-IPE framework and obtains more specialized complexity bounds for it (see Theorem 4.11).

Assume that $0 \notin (F + B)(z_{k-1})$ and let $\hat{\sigma}_k \geq 0$ and an approximation $(\beta_k, \phi_k, \mathcal{A}_k)$ of F at z_{k-1} satisfy (31). Our goal in this section is to present a search procedure which computes a stepsize $\lambda_k > 0$ and a triple $(\tilde{z}_k, u_k, \varepsilon_k) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ satisfying (32) and (33). For the given $\hat{\sigma}_k \geq 0$, it is assumed that there exists a black-box such that:

Assumption B-B: for any given tentative $\lambda_k > 0$, the black-box finds a triple $(\tilde{z}_k, u_k, \varepsilon_k) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ satisfying condition (32).

After a finite number of calls to the black-box, the procedure will eventually find the desired stepsize $\lambda_k > 0$ and a corresponding triple $(\tilde{z}_k, u_k, \varepsilon_k) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ satisfying both (32) and (33).

For a fixed $\lambda_k > 0$, the inclusion $(\mathcal{A}_k + B^{\varepsilon_k})(\tilde{z}_k) \subseteq (\mathcal{A}_k + B)^{\varepsilon_k}(\tilde{z}_k)$ implies that any triple $(\tilde{z}_k, u_k, \varepsilon_k)$ satisfying (32) is a $\hat{\sigma}$-approximate solution (see Definition 2.3) of

$$u \in T(\tilde{z}), \qquad \lambda u + \tilde{z} - z = 0 \tag{41}$$

where $T = \mathcal{A}_k + B$, $\lambda = \lambda_k$, $\hat{\sigma} = \hat{\sigma}_k$ and $z = z_{k-1}$. Hence, the above black-box is also a black-box which, for a given $\lambda > 0$, finds a $\hat{\sigma}$-approximate solution of (41). Thus, to accomplish the goal mentioned above, it suffices to present a stepsize search procedure which accomplishes the following more general goal.

General Goal: *Let a maximal monotone operator $T : \mathbb{E} \rightrightarrows \mathbb{E}$, a point $z \in \mathbb{E}$ such that $0 \notin T(z)$, scalars $\sigma_\ell, \sigma_u \geq 0$, $\hat{\sigma} \in [0, 1)$ and $\beta > 0$ satisfying*

$$\sigma_\ell (1 + \hat{\sigma})^\beta < \sigma_u (1 - \hat{\sigma})^\beta, \tag{42}$$

and a function $\phi \in \Psi(\beta)$ be given. Assuming that we have at our disposal a black-box which, for any given $\lambda > 0$, finds a $\hat{\sigma}$-approximate solution $(\tilde{z}, u, \varepsilon)$ of (41), the goal is to find a specific $\lambda > 0$ and an associated $\hat{\sigma}$-approximate solution $(\tilde{z}, u, \varepsilon)$ of (41) such that

$$\sigma_\ell \leq \lambda \phi(\|\tilde{z} - z\|) \leq \sigma_u. \tag{43}$$

Consider now the goal mentioned at the beginning of this section. Note that by assumption $\hat{\sigma} = \hat{\sigma}_k$ and $\beta = \beta_k$ satisfies (31), and hence (42). Hence, any procedure which solves the General Goal above with $T = \mathcal{A}_k + B$, $\lambda = \lambda_k$, $\hat{\sigma} = \hat{\sigma}_k$ and $\beta = \beta_k$ will also accomplish the goal mentioned at the beginning of this section.

4.1 A generic search procedure

The main goal of this subsection is to describe a search procedure for computing λ and an associated $\hat{\sigma}$-approximate solution $(\tilde{z}, u, \varepsilon)$ of (41) as described in the General Goal posed at the beginning of this section.

We assume throughout this subsection that $T : \mathbb{E} \rightrightarrows \mathbb{E}$ is a maximal monotone operator and $\phi \in \Psi(\beta)$ for some $\beta > 0$. As in subsection 4.1 of [14], define for each $\lambda > 0$ and $z \in \mathbb{E}$,

$$y_T(\lambda; z) := (I + \lambda T)^{-1}(z), \qquad \varphi_T(\lambda; z) := \lambda \|y_T(\lambda; z) - z\|. \tag{44}$$

The procedure proposed in [14] for computing an appropriate stepsize in the context of the large-step HPE method studied there heavily relies on some properties of the function $\varphi_T(\lambda; z)$. Since in this paper we are using a different large-step condition, namely (43), it is more useful in the current context to examine properties of the function $\Phi_T : \mathbb{R}_{++} \times \mathbb{E} \to \mathbb{R}_+$ defined as

$$\Phi_T(\lambda; z) := \lambda \phi(\|y_T(\lambda; z) - z\|) = \lambda \phi\left(\frac{\varphi_T(\lambda; z)}{\lambda}\right) \quad \forall (\lambda, z) \in \mathbb{R}_{++} \times \mathbb{E}. \tag{45}$$

The main motivation for considering this function is due to the fact that the term $\lambda\phi(\|\tilde{z} - z\|)$ in (43) reduces to $\Phi_T(\lambda; z)$ when $(\tilde{z}, u, \varepsilon)$ is a 0-approximate solution of (41). Moreover, the latter function will be shown to provide a good approximation of the above term when $(\tilde{z}, u, \varepsilon)$ is a $\hat{\sigma}$-approximate solution of (41) for some $\hat{\sigma} > 0$ (see Lemma 4.3(d) below).

We first present some known properties of $\varphi_T(\lambda; z)$ and $y_T(\lambda; z)$ which will be useful in our analysis.

Proposition 4.1. *For every $\lambda > 0$ and $z, w \in \mathbb{E}$, there hold:*

$$\|y_T(\lambda; z) - y_T(\lambda; w)\| \le \|z - w\|, \qquad \|[z - y_T(\lambda; z)] - [w - y_T(\lambda; w)]\| \le \|z - w\|. \tag{46}$$

As a consequence, if $z^* \in T^{-1}(0)$, then

$$\max\left\{\|y_T(\lambda; z) - z^*\|, \|y_T(\lambda; z) - z\|\right\} \le \|z - z^*\|. \tag{47}$$

Proof. The proof can be found in Proposition 4.5 of [14]. □

Proposition 4.2. *Consider the function φ_T defined in* (44) *and $z \in \mathbb{E}$. Then, the following statements hold:*

(a) for every $0 < \tilde{\lambda} < \lambda$,

$$\frac{\lambda}{\tilde{\lambda}} \varphi_T(\tilde{\lambda}; z) \le \varphi_T(\lambda; z) \le \left(\frac{\lambda}{\tilde{\lambda}}\right)^2 \varphi_T(\tilde{\lambda}; z);$$

(b) if $0 \notin T(z)$, then $\varphi_T(\lambda; z) > 0$ for every $\lambda > 0$;

(c) if, for some $\lambda > 0$ and $\hat{\sigma} \in [0, 1]$, the triple $(\tilde{z}, u, \varepsilon)$ is a $\hat{\sigma}$-approximate solution of (41)*, then*

$$(1 - \hat{\sigma})\lambda\|\tilde{z} - z\| \le \varphi_T(\lambda; z) \le (1 + \hat{\sigma})\lambda\|\tilde{z} - z\|;$$

as a consequence, if $0 \notin T(z)$, then $\tilde{z} \ne z$.

Proof. Statements (a), (b) and (c) are proved in Lemmas 4.3(b), 4.4(a) and 4.1 of [14], respectively. □

Using the above two technical results, we can now derive some useful properties about Φ_A.

Lemma 4.3. *For a given $z \in \mathbb{E}$, the following statements hold:*

(a) for every $0 < \tilde{\lambda} < \lambda$,

$$\frac{\lambda}{\tilde{\lambda}}\,\Phi_T(\tilde{\lambda}; z) \leq \Phi_T(\lambda; z) \leq \left(\frac{\lambda}{\tilde{\lambda}}\right)^{1+\beta} \Phi_T(\tilde{\lambda}; z); \tag{48}$$

(b) $\lambda > 0 \mapsto \Phi_T(\lambda, z)$ is continuous and nondecreasing;

(c) if $0 \notin T(z)$, then $\Phi_T(\lambda; z) > 0$ for every $\lambda > 0$; moreover, $\lambda > 0 \mapsto \Phi_T(\lambda; z)$ is a strictly increasing function which converges to 0 and ∞ as λ tends to 0 and ∞, respectively;

(d) if, for some $\lambda > 0$ and $\hat{\sigma} \in [0, 1]$, the triple $(\tilde{z}, u, \varepsilon)$ is a $\hat{\sigma}$-approximate solution of (41), then

$$(1-\hat{\sigma})^{\beta}\lambda\phi(\|\tilde{z}-z\|) \leq \Phi_T(\lambda; z) \leq (1+\hat{\sigma})^{\beta}\lambda\phi(\|\tilde{z}-z\|). \tag{49}$$

Proof. (a) For $0 < \tilde{\lambda} < \lambda$, the second inequality in Proposition 4.2(a) gives

$$\frac{\varphi_T(\lambda; z)}{\lambda} \leq \frac{\lambda}{\tilde{\lambda}}\frac{\varphi_T(\tilde{\lambda}; z)}{\tilde{\lambda}}.$$

Using (45), the above inequality, Definition 2.4(a) and Definition 3.1(a) we obtain

$$\begin{aligned}\Phi_T(\lambda; z) = \lambda\phi\left(\frac{\varphi_T(\lambda; z)}{\lambda}\right) &\leq \lambda\phi\left(\frac{\lambda}{\tilde{\lambda}}\frac{\varphi_T(\tilde{\lambda}; z)}{\tilde{\lambda}}\right) \\ &\leq \lambda\left(\frac{\lambda}{\tilde{\lambda}}\right)^{\beta}\phi\left(\frac{\varphi_T(\tilde{\lambda}; z)}{\tilde{\lambda}}\right) = \left(\frac{\lambda}{\tilde{\lambda}}\right)^{1+\beta}\Phi_T(\tilde{\lambda}; z),\end{aligned}$$

which gives the second inequality in (48). The proof of the first inequality uses a similar reasoning and the first inequality in Proposition 4.2(a) instead of the second one.

(b) This statement follows immediately from (a).

(c) The first assertion follows from (45), Proposition 4.2(b) and Definition 2.4(b). The other assertion follows immediately from the first one and (a).

(d) Assume that $(\tilde{z}, u, \varepsilon)$ is a $\hat{\sigma}$-approximate solution of (41) for some $\lambda > 0$ and $\hat{\sigma} \in [0, 1]$. From the second inequality in Proposition 4.2(c), we have

$$\frac{\varphi_T(\lambda; z)}{\lambda} \leq (1+\hat{\sigma})\|\tilde{z}-z\|.$$

Using the above inequality, (45), Definition 2.4(a) and Definition 3.1(a) we obtain $\Phi_T(\lambda; z) \leq \lambda\phi\left((1+\hat{\sigma})\|\tilde{z}-z\|\right) \leq (1+\hat{\sigma})^{\beta}\lambda\phi(\|\tilde{z}-z\|)$, which gives the second inequality in (49). The proof of the first inequality in (49) uses a similar reasoning, (29) with $\tau = 1-\hat{\sigma}$ instead of Definition 3.1(a), and the first inequality in Proposition 4.2(c). □

Lemma 4.4. *Let $z \in \mathbb{E}$ such that $0 \notin T(z)$ and scalars $\sigma_\ell, \sigma_u \geq 0$, $\hat{\sigma} \in [0,1)$ and $\beta > 0$ satisfying* (42) *be given. Then, the set of all scalars $\lambda > 0$ such that*

$$\sigma_\ell (1+\hat{\sigma})^\beta \leq \Phi_T(\lambda; z) \leq \sigma_u (1-\hat{\sigma})^\beta \tag{50}$$

is an closed interval $[\lambda_\ell, \lambda_u] \subset \mathbb{R}_{++}$ such that

$$\frac{\lambda_u}{\lambda_\ell} \geq \left[\left(\frac{1-\hat{\sigma}}{1+\hat{\sigma}} \right)^\beta \left(\frac{\sigma_u}{\sigma_\ell} \right) \right]^{\frac{1}{1+\beta}} > 1. \tag{51}$$

Moreover, if $\lambda > 0$ and $(\tilde{z}, u, \varepsilon)$ is a $\hat{\sigma}$-approximate solution of (41)*, then the following statements hold:*

(a) if $\lambda \phi(\|\tilde{z} - z\|) < \sigma_\ell$, then $\lambda < \lambda_\ell$ and $\lambda_u \leq \sigma_u / \phi(\|\tilde{z} - z\|)$;

(b) if $\lambda \phi(\|\tilde{z} - z\|) > \sigma_u$, then $\sigma_\ell / \phi(\|\tilde{z} - z\|) \leq \lambda_\ell$ and $\lambda_u < \lambda$.

As a consequence, if $\lambda \in [\lambda_\ell, \lambda_u]$ then $(\lambda, z, \tilde{z})$ satisfies condition (43).

Proof. First note that (42) and Lemma 4.3(c) imply the existence of unique scalars $0 < \lambda_\ell < \lambda_u$ such that

$$\Phi_T(\lambda_\ell; z) = \sigma_\ell (1+\hat{\sigma})^\beta, \qquad \Phi_T(\lambda_u; z) = \sigma_u (1-\hat{\sigma})^\beta, \tag{52}$$

and that the set of $\lambda > 0$ such that (50) holds is equal to the interval $[\lambda_\ell, \lambda_u]$. The last two identities and the second inequality in (48) with $\tilde{\lambda} = \lambda_\ell$ and $\lambda = \lambda_u$ imply

$$\sigma_u (1-\hat{\sigma})^\beta = \Phi_T(\lambda_u; z) \leq \left(\frac{\lambda_u}{\lambda_\ell} \right)^{1+\beta} \Phi_T(\lambda_\ell; z) = \left(\frac{\lambda_u}{\lambda_\ell} \right)^{1+\beta} \sigma_\ell (1+\hat{\sigma})^\beta,$$

which implies (51).

To prove (a), assume that $\lambda > 0$ is such that $\lambda \phi(\|\tilde{z} - z\|) < \sigma_\ell$. The second inequality in (49) and the first identity in (52) then imply that $\Phi_T(\lambda; z) < (1+\hat{\sigma})^\beta \sigma_\ell = \Phi_T(\lambda_\ell; z)$. The claim that $\lambda < \lambda_\ell$ now follows immediately from Lemma 4.3(c). Multiplying the first inequality in (49) by λ_u / λ and using the first inequality in (48) with $\tilde{\lambda} = \lambda$ and $\lambda = \lambda_u$, we conclude that

$$(1-\hat{\sigma})^\beta \lambda_u \phi(\|\tilde{z} - z\|) \leq \frac{\lambda_u}{\lambda} \Phi_T(\lambda; z) \leq \Phi_T(\lambda_u; z) = \sigma_u (1-\hat{\sigma})^\beta,$$

where the last equality follows from (52). Hence, the second inequality in (a) holds. The proof of (b) follows in a similar manner.

The last conclusion of the lemma follows immediately from (a) and (b). □

The interval $[\lambda_\ell, \lambda_u]$ of Lemma 4.4 plays an important role in the analysis of the search procedure that will be presented next. Observe that, in view of statements (a) and (b) of Lemma 4.4, the fact that the triple $(\lambda, z, \tilde{z})$ fails to satisfy (43) gives you the relevant information that the interval determined by λ and another easily computable scalar contains the interval $[\lambda_\ell, \lambda_u]$. Moreover, (51) ensures that the length of the latter interval on the logarithmic scale is bounded below by a positive constant which depends only on $\hat{\sigma}$, σ_ℓ, σ_u and β.

We now present the generic search procedure which accomplishes the General Goal described at the beginning of this section. The procedure consists of two stages. The first one, namely the *bracketing stage*, either computes a stepsize $\lambda > 0$ and a $\hat{\sigma}$-approximate solution $(\tilde{z}, u, \varepsilon)$ of (41) such that the triple $(\lambda, z, \tilde{z})$ satisfies (43), or finds an interval $[t_\ell, t_u]$ containing $[\lambda_\ell, \lambda_u]$. The second one, namely the *bisection stage*, iteratively performs a geometric bisection scheme on the interval $[t_\ell, t_u]$ until a triple $(\lambda, z, \tilde{z})$ as described above is found.

We start by describing the bracketing stage.

Bracketing stage:

Input: $z \in \mathbb{E}$ such that $0 \notin T(z)$, scalars $\sigma_\ell, \sigma_u \geq 0$, $\hat{\sigma} \in [0, 1)$ and $\beta > 0$ satisfying (42), and initial guess $\lambda^0 > 0$;

Output: either a scalar $\lambda > 0$ and a $\hat{\sigma}$-approximate solution $(\tilde{z}, u, \varepsilon)$ of (41) such that the triple $(\lambda, z, \tilde{z})$ satisfies (43), or the endpoints of an interval $[t_\ell, t_u]$ containing the interval $[\lambda_\ell, \lambda_u]$ as in Lemma 4.4.

(1) use the black-box to compute a $\hat{\sigma}$-approximate solution $(\tilde{z}^0, u^0, \varepsilon^0)$ of (41) with $\lambda = \lambda^0$;

(2) (a) if $\lambda^0 \phi(\|\tilde{z}^0 - z\|) \in [\sigma_\ell, \sigma_u]$, then output $\lambda = \lambda^0$ and $(\tilde{z}, u, \varepsilon) = (\tilde{z}^0, u^0, \varepsilon^0)$;
(b) if $\lambda^0 \phi(\|\tilde{z}^0 - z\|) < \sigma_\ell$, then output $t_\ell := \lambda^0$ and $t_u := \sigma_u / \phi(\|\tilde{z}^0 - z\|)$;
(c) if $\lambda^0 \phi(\|\tilde{z}^0 - z\|) > \sigma_u$, then output $t_\ell := \sigma_\ell / \phi(\|\tilde{z}^0 - z\|)$ and $t_u := \lambda^0$.

end

The following result establishes the correctness of the bracketing stage.

Lemma 4.5. *The Bracketing stage produces the desired output at the end of step 2.*

Proof. If (2a) holds then $(\tilde{z}, u, \varepsilon)$ is clearly a $\hat{\sigma}$-approximate solution of (41) such that $(\lambda, z, \tilde{z})$ satisfies (43). Otherwise, if either (2b) or (2c) holds then $[t_\ell, t_u] \supset [\lambda_\ell, \lambda_u]$ in view of (a) and (b) of Lemma 4.4 and the definitions of t_ℓ and t_u in (2b) and (2c). □

We next describe the bisection stage which is the one that accounts for the overall complexity of the stepsize search procedure.

Bisection stage:

Input: $z \in \mathbb{E}$ such that $0 \notin T(z)$, scalars $\sigma_\ell, \sigma_u \geq 0$, $\hat{\sigma} \in [0, 1)$ and $\beta > 0$ satisfying (42), and interval $[t_\ell, t_u]$ containing the interval $[\lambda_\ell, \lambda_u]$ as in Lemma 4.4;

Output: a scalar λ and a $\hat{\sigma}$-approximate solution $(\tilde{z}, u, \varepsilon)$ of (41) such that the triple $(\lambda, z, \tilde{z})$ satisfies (43).

(1) set $\lambda = \sqrt{t_\ell t_u}$ and use the black-box to compute a $\hat{\sigma}$-approximate solution $(\tilde{z}, u, \varepsilon)$ of (41);

(2) if $\lambda \phi(\|\tilde{z} - z\|) \in [\sigma_\ell, \sigma_u]$, then output λ and $(\tilde{z}, u, \varepsilon)$, and **stop**;

(3) if $\lambda \phi(\|\tilde{z} - z\|) > \sigma_u$, then set $t_u \leftarrow \lambda$; else set $t_\ell \leftarrow \lambda$;

(4) go to step 1.

end

We now make some remarks about the bisection stage. If the stage stops at step 2, then clearly λ and $(\tilde{z}, u, \varepsilon)$ are such that the triple $(\lambda, z, \tilde{z})$ satisfies (43). On the other hand, if the stage enters into step 3, then the interval $[t_\ell, t_u]$ obtained after this step is executed can be easily seen, by induction argument, to satisfy the invariant condition $[t_\ell, t_u] \supset [\lambda_\ell, \lambda_u]$ in view of Lemma 4.4.

Lemma 4.6. *Assume that the bracketing stage outputs an interval $[t_\ell, t_u]$ (which therefore contains $[\lambda_\ell, \lambda_u]$) and that this interval is given as input to the bisection stage. Then, the bisection stage terminates with the desired output after performing no more than*

$$1 + \log\left(\frac{(1+\beta)\log\tau^0}{\log[(1-\hat{\sigma})^\beta \sigma_u / ((1+\hat{\sigma})^\beta \sigma_\ell)]}\right)$$

iterations (and hence black-box calls), where

$$\tau^0 := \max\left\{\frac{\sigma_u}{\lambda^0 \phi(\|\tilde{z}^0 - z\|)}, \frac{\lambda^0 \phi(\|\tilde{z}^0 - z\|)}{\sigma_\ell}\right\} > 1 \tag{53}$$

and $\tilde{z}^0$ is obtained as in step 1 of the bracketing stage.

Proof. Consider the interval $[\lambda_\ell, \lambda_u]$ as in Lemma 4.4 and let $[t_\ell^0, t_u^0]$ denote the interval output by the bracketing stage. In view of the bracketing stage, we have $[t_\ell^0, t_u^0] \supset [\lambda_\ell, \lambda_u]$. Moreover, (53) and step 2 of the bracketing stage can be easily seen to imply that

$$\tau^0 = \frac{t_u^0}{t_\ell^0}. \tag{54}$$

Assume that the bisection stage does not terminate at the j-th iteration and let t_ℓ^j and t_u^j denote the scalars t_ℓ and t_u obtained at the end of this iteration. Using the second observation following the bisection stage, and steps 1 and 3 of the bisection stage, we conclude that

$$[t_\ell^j, t_u^j] \supset [\lambda_\ell, \lambda_u], \quad \log\left(\frac{t_u^j}{t_\ell^j}\right) = \frac{1}{2^j}\log\left(\frac{t_u^0}{t_\ell^0}\right) = \frac{1}{2^j}\log\tau^0,$$

which together with (51) then imply that

$$\frac{1}{2^j}\log\tau^0 = \log\left(\frac{t_u^j}{t_\ell^j}\right) \geq \log\left(\frac{\lambda_u}{\lambda_\ell}\right) \geq \frac{1}{1+\beta}\log\left(\frac{(1-\hat{\sigma})^\beta \sigma_u}{(1+\hat{\sigma})^\beta \sigma_\ell}\right),$$

and hence that

$$j \leq \log\left(\frac{(1+\beta)\log\tau^0}{\log[(1-\hat{\sigma})^\beta \sigma_u / ((1+\hat{\sigma})^\beta \sigma_\ell)]}\right).$$

The conclusion of the lemma now immediately follows from the above inequality. □

4.2 Computation of the stepsize λ_k

The goal of this subsection is to present a search procedure for computing a stepsize λ_k and a triple $(\tilde{z}_k, u_k, \varepsilon_k)$ as in step 3 of the GLS-IPE framework, and to derive its computational complexity.

We start by stating the stepsize search procedure, which is essentially the brackecting and the bisection stages of Subsection 4.1 combined and specialized to the case in which $z = z_{k-1}$, $\hat{\sigma} = \hat{\sigma}_k$,

$T = \mathcal{A}_k + B$ and $(\beta, \phi) = (\beta_k, \phi_k)$ (see the notation of the GLS-IPE framework). The procedure assumes that we have at our disposal a black-box which, for any given tentative $\lambda_k > 0$, is able to compute a triple $(\tilde{z}_k, u_k, \varepsilon_k) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ satisfying (32).

Bracketing/Bisection Procedure:

Input: $z_{k-1} \in \mathbb{E}$ such that $0 \notin (F + B)(z_{k-1})$, scalars $\sigma_\ell, \sigma_u \geq 0$, $\hat{\sigma}_k \in [0, 1)$ and an approximation $(\beta_k, \phi_k, \mathcal{A}_k)$ of F at z_{k-1} satisfying the second inequality in (31), and an initial guess $\lambda_k^0 > 0$.

Output: stepsize $\lambda_k > 0$ and a triple $(\tilde{z}_k, u_k, \varepsilon_k)$ satisfying (32) and (33).

(1) (**Bracketing stage**) use the black-box to compute a triple $(\tilde{z}_k^0, u_k^0, \varepsilon_k^0)$ satisfying (32) with $\lambda_k = \lambda_k^0$.

- (1.a) if $\lambda_k^0 \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|) \in [\sigma_\ell, \sigma_u]$ then output $\lambda_k := \lambda_k^0$ and $(\tilde{z}_k, u_k, \varepsilon_k) := (\tilde{z}_k^0, u_k^0, \varepsilon_k^0)$ and **stop**;
- (1.b) if $\lambda_k^0 \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|) < \sigma_\ell$, then set $t_\ell := \lambda_k^0$ and $t_u := \sigma_u / \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|)$;
- (1.c) if $\lambda_k^0 \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|) > \sigma_u$, then set $t_\ell := \sigma_\ell / \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|)$ and $t_u := \lambda_k^0$;

(2) (**Bisection stage**)

- (2.a) set $\lambda_k = \sqrt{t_\ell \, t_u}$ and use the Black-Box to compute a triple $(\tilde{z}_k, u_k, \varepsilon_k)$ satisfying (32);
- (2.b) if $\lambda_k \phi_k(\|\tilde{z} - z_{k-1}\| \in [\sigma_\ell, \sigma_u]$, then output λ_k and $(\tilde{z}_k, u_k, \varepsilon_k)$, and **stop**;
- (2.c) if $\lambda_k \phi_k(\|\tilde{z} - z_{k-1}\|) > \sigma_u$, then set $t_u \leftarrow \lambda_k$; else set $t_\ell \leftarrow \lambda_k$;
- (2.d) go to step 2.a.

end

Observe that step 0 of the above procedure assumes that $0 \notin (F + B)(z_{k-1})$, which equivalent to assuming that $0 \notin (\mathcal{A}_k + B)(z_{k-1})$ due to the fact that Definition 3.2(b) implies that $\mathcal{A}_k(z_{k-1}) = F(z_{k-1})$.

The following result is a specialization of Lemma 4.6 to the particular context of the above procedure.

Lemma 4.7. *The number of black-box calls performed by the Bracketing/Bisection procedure at the k-th iteration of the GLS-IPE framework is bounded by*

$$2 + \log^+ \left(\frac{(1 + \beta_k) \log \tau_k^0}{\log[(1 - \hat{\sigma}_k)^{\beta_k} \sigma_u / ((1 + \hat{\sigma}_k)^{\beta_k} \sigma_\ell)]} \right),$$

where

$$\tau_k^0 := \max \left\{ \frac{\sigma_u}{\lambda_k^0 \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|)} \, , \, \frac{\lambda_k^0 \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|)}{\sigma_\ell} \right\} > 1.$$

Proof. If the Bracketing/Bisection procedure terminates at the brackecting stage, the conclusion of the lemma immediately holds due to the fact that $\log_+(t) \geq 0$ for all $t > 0$. Otherwise, the conclusion of the lemma follows from Lemma 4.6 with $\phi = \phi_k$, $\beta = \beta_k$, $\hat\sigma = \hat\sigma_k$, $z = z_{k-1}$, $\tilde z^0 = \tilde z_k^0$ and $\lambda^0 = \lambda_k^0$, and the fact that $\log(t) \le \log^+(t)$ for all $t > 0$. □

The above bound on the number of black-box calls depends on the term $\phi_k(\|\tilde z_k^0 - z_{k-1}\|)$ and its inverse. In what follows, we derive lower and upper bounds on this quantity. Lemma 4.9 establishes an upper bound on the latter quantity while Lemma 4.10 derives a lower bound on it. Finally, Theorem 4.11 states an alternative bound on the number of black-box calls of the Bracketing/Bisection procedure given in terms of the lower and upper bounds derived in these two lemmas.

We start by stating the following technical result.

Lemma 4.8. *Assume that $z^* \in (F+B)^{-1}(0)$ and let $(\lambda, z) \in \mathbb{R}_+ \times \mathbb{E}$ be given. Let also $(\beta_z, \phi_z, \mathcal{A}_z)$ be an approximation of F at z and define $T_z : \mathbb{E} \rightrightarrows \mathbb{E}$ as $T_z := \mathcal{A}_z + B$. Then,*

$$\|z - (I + \lambda T_z)^{-1}(z)\| \le \|z - z^*\| \left[1 + \lambda \phi_z\left(\|z - z^*\|\right)\right].$$

Proof. Let

$$r := F(z^*) - \mathcal{A}_z(z^*), \qquad \tilde T_z := T_z + r. \tag{55}$$

Using the assumption that $0 \in (F+B)(z^*)$ and the definition of T_z and r, we easily see that $0 \in \tilde T_z(z^*)$. Hence, by relation (47) of Proposition 4.1 with $T = \tilde T_z$, we have

$$\|z - (I + \lambda \tilde T_z)^{-1}(z)\| \le \|z - z^*\|. \tag{56}$$

Using the observation that

$$(I + \lambda \tilde T_z)^{-1}(z + \lambda r) = [I + \lambda(T_z + r)]^{-1}(z + \lambda r) = (I + \lambda T_z)^{-1}(z),$$

the triangle inequality for norms, relation (56) and the fact that by Proposition 4.1 the resolvent is non-expansive, we then conclude that

$$\begin{aligned}
\|z - (I + \lambda T_z)^{-1}(z)\| =& \|z - (I + \lambda \tilde T_z)^{-1}(z + \lambda r)\| \\
\le& \|z - (I + \lambda \tilde T_z)^{-1}(z)\| + \|(I + \lambda \tilde T_z)^{-1}(z) - (I + \lambda \tilde T_z)^{-1}(z + \lambda r)\| \\
\le& \|z - z^*\| + \lambda \|r\| \le \|z - z^*\| \left[1 + \lambda \phi_z(\|z - z^*\|)\right],
\end{aligned}$$

where the last inequality is due to the definition of r in (55) and relation (30) with $z' = z^*$. □

Lemma 4.9. *Consider the sequences $\{z_k\}$, $\{\phi_k\}$, $\{\hat\sigma_k\}$ and $\{\beta_k\}$ generated by the GLS-IPE framework, the sequence $\{\lambda_k^0\}$ of initial stepsizes for the Bracketing/Bisection procedure and the corresponding sequence of triples $\{(\tilde z_k^0, u_k^0, \varepsilon_k^0)\}$ generated at its step 1. Let also d_0 denote the distance of z_0 to $(F+B)^{-1}(0)$. Then, for every $k \ge 1$,*

$$\phi_k\left(\|\tilde z_k^0 - z_{k-1}\|\right) \le \frac{1}{(1 - \hat\sigma_k)^{\beta_k}} \left\{\phi_k\left([1 + \lambda_k^0 \phi_k(d_0)] d_0\right)\right\}.$$

Proof. Letting $T_k = \mathcal{A}_k + B$ and noting that $(\tilde z_k^0, u_k^0, \varepsilon_k^0)$ is a $\hat\sigma_k$-approximate solution of (41) with $\lambda = \lambda_k^0$ and $T = T_k$, we conclude from Proposition 4.2(c) and (44) with $\lambda = \lambda_k^0$, $T = T_k$, $\hat\sigma = \hat\sigma_k$ and $z = z_{k-1}$ that

$$(1 - \hat\sigma_k)\|\tilde z_k^0 - z_{k-1}\| \le \|(I + \lambda_k^0 T_k)^{-1}(z_{k-1}) - z_{k-1}\|.$$

Now, let $z^* \in (F+B)^{-1}(0)$ be such that $\|z^*-z_0\| = d_0$ and note that $\|z_{k-1}-z^*\| \le \|z_0 - z^*\| = d_0$, in view of Lemma 3.4(a). Using the latter observation, the fact that the assumption $\phi_k \in \Psi(\beta_k)$ implies that ϕ_k is increasing, and Lemma 4.8 with $(\lambda, z) = (\lambda_k^0, z_{k-1})$ and $(\beta_z, \phi_z, \mathcal{A}_z) = (\beta_k, \phi_k, \mathcal{A}_k)$, we conclude that

$$\|(I+\lambda_k^0 T_k)^{-1}(z_{k-1}) - z_{k-1}\| \le \left[1+\lambda_k^0\phi_k(d_0)\right] d_0.$$

The conclusion of the lemma now follows the fact that ϕ_k is increasing, the assumption that $\phi_k \in \Psi(\beta_k)$ implies inequality (29) with $\psi = \phi_k$, $\tau = 1-\hat{\sigma}_k$ and $\beta = \beta_k$, and the above two displayed relations. □

Lemma 4.10. *Let $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$ be given and let ζ_k be defined as*

$$\zeta_k := \min\left\{1, \lambda_k^0 \phi_k(\mu_k)\right\}, \tag{57}$$

where

$$\mu_k := \min\left\{\frac{(2\lambda_k^0\bar{\varepsilon})^{1/2}}{\hat{\sigma}_k}, \frac{\lambda_k^0\bar{\rho}}{2+\hat{\sigma}_k}\right\} \tag{58}$$

and λ_k^0 and $\hat{\sigma}_k$ are as in the input of the Bracketing/Bisection procedure. Also, consider the triple $(\tilde{z}_k^0, u_k^0, \varepsilon_k^0)$ computed at step 1 of the Bracketing/Bisection procedure and define

$$v_k^0 := F(u_k^0) + u_k^0 - \mathcal{A}_k(\tilde{z}_k^0).$$

Then, either one of the following statements hold:

(a) $\lambda_k^0\phi_k(\|\tilde{z}_k^0 - z_{k-1}\|) > \zeta_k$, in which case

$$\zeta_k < \frac{\sigma_u}{\sigma_\ell}\left\{\lambda_k^0\phi_k\left([1+\lambda_k^0\phi_k(d_0)]d_0\right)\right\}; \tag{59}$$

(b) $v_k^0 \in (F+B^{\varepsilon_k^0})(\tilde{z}_k^0) \subseteq (F+B)^{\varepsilon_k^0}(\tilde{z}_k^0)$ and the following error bounds hold:

$$\|v_k^0\| \le \bar{\rho}, \qquad \varepsilon_k^0 \le \bar{\varepsilon}. \tag{60}$$

Proof. Assume that (a) does not hold, i.e.,

$$\lambda_k^0\phi_k(\|\tilde{z}_k^0 - z_{k-1}\|) \le \zeta_k. \tag{61}$$

Using the fact that $(\tilde{z}_k^0, u_k^0, \varepsilon_k^0)$ satisfies (32) with $\lambda_k = \lambda_k^0$ in view of step 1 of the Bracketing/Bisection procedure, we conclude from relation (61) and Lemma 3.3 with $(\lambda, z) = (\lambda_k^0, z_{k-1})$ and $(\tilde{z}, u, \varepsilon) = (\tilde{z}_k^0, u_k^0, \varepsilon_k^0)$ that $v_k^0 \in (F+B^{\varepsilon_k^0})(\tilde{z}_k^0)$,

$$\|\lambda_k^0 v_k^0 + \tilde{z}_k^0 - z_{k-1}\| \le (\hat{\sigma}_k + \zeta_k)\|\tilde{z}_k^0 - z_{k-1}\|, \quad 2\lambda_k^0\varepsilon_k^0 \le \hat{\sigma}_k\|\tilde{z}_k^0 - z_{k-1}\|^2. \tag{62}$$

Now the latter conclusion together with the fact that $\zeta_k \le 1$ easily imply that

$$\max\left\{\frac{\|\lambda_k^0 v_k^0\|}{2+\hat{\sigma}_k}, \frac{(2\lambda_k^0\varepsilon_k^0)^{1/2}}{\hat{\sigma}_k}\right\} \le \|\tilde{z}_k^0 - z_{k-1}\|,$$

The previous inequality, Definition 2.4(a) and relations (57) and (61), then imply that

$$\phi_k\left(\max\left\{\frac{\|\lambda_k^0 v_k^0\|}{2+\hat{\sigma}_k}, \frac{(2\lambda_k^0\varepsilon_k^0)^{1/2}}{\hat{\sigma}_k}\right\}\right) \leq \phi_k(\|\tilde{z}_k^0 - z_{k-1}\|) \leq \frac{\zeta_k}{\lambda_k^0} \leq \phi_k\left(\mu_k\right).$$

Using Definition 2.4(a) again, we then conclude that that

$$\max\left\{\frac{\|\lambda_k^0 v_k^0\|}{2+\hat{\sigma}_k}, \frac{(2\lambda_k^0\varepsilon_k^0)^{1/2}}{\hat{\sigma}_k}\right\} \leq \mu_k,$$

which, in view of the definition of μ_k and a simple algebraic manipulation, immediately implies (60). Finally, note that the second inclusion in (b) follows from Proposition 2.2(b). We have thus shown that (b) holds. To finish the proof, note that (59) follows from the first statement in (a), Lemma 4.9 and (31). □

Theorem 4.11. *Let tolerances $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$ be given and consider the GLS-IPE framework where the stepsize λ_k and the triple $(\tilde{z}_k, u_k, \varepsilon_k)$ in its step 3 is found by means of the Bracketing/Bisection procedure. Then, either the triple $(\tilde{z}_k^0, v_k^0, \varepsilon_k^0)$ as in Lemma 4.10 is a $(\bar{\rho}, \bar{\varepsilon})$-approximate solution of* (26) *in the sense of statement (b) of Lemma 4.10, or the Bracketing/Bisection procedure performs at most*

$$2 + \log^+\left[(1+\beta_k)\left[\log\frac{(1-\hat{\sigma}_k)^{\beta_k}\sigma_u}{(1+\hat{\sigma}_k)^{\beta_k}\sigma_\ell}\right]^{-1}\log\left(\sigma_u \max\left\{\frac{\lambda_k^0\phi_k\left([1+\lambda_k^0\phi_k(d_0)]d_0\right)}{\sigma_\ell^2}, \zeta_k^{-1}\right\}\right)\right] \tag{63}$$

black-box calls during the k-th iteration of the GLS-IPE framework, where ζ_k is as in (57) *and λ_k^0 is the initial stepsize for the Bracketing/Bisection procedure during the k-th iteration of the GLS-IPE framework.*

Proof. First note that the second inequality in (31) and relation (59) imply that the arguments for the two log terms in (63) is greater than one, and hence (63) is well-defined. Now, the conclusion of the theorem and the bound (63) immediately follow from Lemmas 4.7, 4.9 and 4.10, and the fact that the assumption that $\hat{\sigma}_k \geq 0$ and the second inequality in (31) imply that $(1-\hat{\sigma}_k)^{\beta_k} > \sigma_\ell/\sigma_u$. □

Even though the black-box-complexity bound for the stepsize search procedure given by Theorem 4.11 depends on k, it easily leads to one which does not depend on k under some mild assumptions on $\hat{\sigma}_k$, β_k, λ_k^0 and ϕ_k. Instead of stating a result along this direction in this section, we will derive uniform black-box-complexity bounds (i.e., ones that do not depend on k) for the stepsize search procedure only in the context of the two applications described in Section 5 by making direct use of Theorem 4.11.

5 Applications to SQP type methods

In this section, we consider two instances of the GLS-IPE framework studied in Section 3 for solving the convex optimization problem

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & g(x) \leq 0, \end{array} \tag{64}$$

where $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $g = (f_1, \ldots, f_m)^T : \mathbb{R}^n \to \mathbb{R}^m$ and the following assumptions are made:

O.1) $f_0, f_1, \ldots, f_m$ are convex and twice continuously differentiable;

O.2) the Hessian of f_i is L_i-Lipzchitz continuous for every $i = 0, 1, \ldots, m$;

O.3) there exists $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $(x, y) = (x^*, y^*)$ satisfies

$$\nabla_x \mathcal{L}(x, y) = 0, \quad g(x) \leq 0, \quad y \geq 0, \quad \langle y, g(x) \rangle = 0, \tag{65}$$

where $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, \infty)$ is the canonical *Lagrangian* defined as

$$\mathcal{L}(x, y) = f_0(x) + \langle y, g(x) \rangle - \delta_{\mathbb{R}^m_+}(y). \tag{66}$$

We now make some remarks about the above assumptions. First, it is well-known that a pair (x^*, y^*) satisfies (65) if and only if x^* is an optimal solution of (64), y^* is an optimal solution for the Lagrangian dual $\max_{y \geq 0} \{\inf_{x \in \mathbb{R}^n} \mathcal{L}(x, y)\}$ and the duality gap between the two problems is zero. In this case, y^* is said to be a Lagrange multiplier associated with the optimal solution x^* of (64). Second, (x^*, y^*) satisfies (65) if and only if (x^*, y^*) is a solution of

$$\nabla_x \mathcal{L}(x, y) = 0, \quad -g(x) + N_{\mathbb{R}^m_+}(y) \ni 0, \tag{67}$$

or equivalently (x^*, y^*) is a solution of (26) with F and B defined by

$$F(x, y) = \begin{bmatrix} \nabla f_0(x) + \nabla g(x) y \\ -g(x) \end{bmatrix}, \qquad B(x, y) = N_{\mathbb{R}^n \times \mathbb{R}^m_+}(x, y) \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^m. \tag{68}$$

Clearly,

$$F(x, y) = \begin{bmatrix} \nabla_x \mathcal{L}(x, y) \\ -g(x) \end{bmatrix} \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+. \tag{69}$$

Third, in view of Corollary 12.18 in [18], B is a maximal monotone operator. Fourth, assumption O.1 implies that F is continuously differentiable on $\mathbb{E}$ and

$$F'(x, y) = \begin{bmatrix} \nabla^2_{xx} \mathcal{L}(x, y) & \nabla g(x) \\ -\nabla g(x)^T & 0 \end{bmatrix} \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+. \tag{70}$$

Fifth, since the linear operator $F'(x, y)$ is positive semidefinite for every $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$ and the latter set is obviously convex, it follows from [7, Proposition 2.3.2] that F is monotone on $\mathbb{R}^n \times \mathbb{R}^m_+$.

In view of the above remarks and assumptions O.1 and O.3, we conclude that the operators F and B satisfy conditions C.1-C.3 of Section 3 with

$$\Omega := \mathbb{R}^n \times \mathbb{R}^m_+, \qquad \mathbb{E} := \mathbb{R}^n \times \mathbb{R}^m, \tag{71}$$

and $\mathbb{E}$ endowed with the canonical inner product, namely, $\langle (x, y), (x', y') \rangle = \langle x, x' \rangle + \langle y, y' \rangle$.

Our goal in this section will be to present two instances of the GLS-IPE framework to compute approximate solutions of the maximal monotone inclusion (26). Since instances of the GLS-IPE framework are inexact proximal point methods, it is interesting to discuss the different characterizations of an exact proximal point iteration applied to (26). This will also give insight on the different equivalent forms of an iteration of the two GLS-IPE instances discussed in Subsections 5.1 and 5.2.

With the later goal in mind, it is useful to introduce, for a given constant $\lambda > 0$, the *augmented Lagrangian* $\mathcal{L}_\lambda : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined as

$$\mathcal{L}_\lambda(x, y) = f_0(x) + \frac{1}{2\lambda} \left[\left\| (y + \lambda g(x))^+ \right\|^2 - \|y\|^2 \right]. \tag{72}$$

Proposition 5.1. *Let f_0 and $g = (f_1, \ldots, f_m)$ be as in* (64) *and F and B be as in* (68)*. For any $(x, y) \in \mathbb{E}$ and $\lambda > 0$, the following conditions on a given pair $(\tilde{x}, \tilde{y}) \in \mathbb{E}$ are equivalent:*

(a) $(\tilde{x}, \tilde{y}) = (I + \lambda (F + B))^{-1}(x, y)$;

(b) $\tilde{x} = \arg\min_{x' \in \mathbb{R}^n} \mathcal{L}_\lambda(x', y) + (1/2\lambda)\|x' - x\|^2$ *and* $\tilde{y} = (y + \lambda g(\tilde{x}))^+$;

(c) $(\tilde{x}, \tilde{y})$ *is the unique solution of*

$$\begin{aligned} &\min_{(x',y') \in \mathbb{E}} f_0(x') + \frac{1}{2\lambda}\left[\|y'\|^2 + \|x' - x\|^2\right] \\ &\text{s. t. } g(x') - \frac{1}{\lambda}(y' - y) \le 0. \end{aligned} \tag{73}$$

Moreover, $\tilde{y}$ is the unique Lagrange multiplier of (73)*.*

Proof. (a)$\Leftrightarrow$(b) By (68) and (69), we have that (a) is equivalent to

$$0 = \lambda \nabla_x \mathcal{L}(\tilde{x}, \tilde{y}) + \tilde{x} - x, \quad 0 \in \lambda\left(-g(\tilde{x}) + N_{\mathbb{R}^m_+}(\tilde{y})\right) + \tilde{y} - y\,. \tag{74}$$

Clearly, the inclusion in (74) is equivalent to $\tilde{y} = (y + \lambda g(\tilde{x}))^+$ so that that (74), and hence (a), is equivalent to

$$0 = \lambda \nabla_x \mathcal{L}(\tilde{x}, \tilde{y}) + \tilde{x} - x, \quad \tilde{y} = (y + \lambda g(\tilde{x}))^+\,. \tag{75}$$

Since the second equality in the above relation together with (72) imply that

$$\nabla_x \mathcal{L}_\lambda(\tilde{x}, y) = \nabla f_0(\tilde{x}) + \nabla g(\tilde{x})(y + \lambda g(\tilde{x}))^+ = \nabla_x \mathcal{L}(\tilde{x}, \tilde{y}),$$

we conclude that (75) is obviously equivalent to (b).

(a)$\Leftrightarrow$(c) Let $\mathcal{Q} : \mathbb{E} \times \mathbb{R}^m \to \mathbb{R}$ denote the canonical Lagrangian of (73), i.e.,

$$\mathcal{Q}(z', w') = \mathcal{L}(x', w') + \frac{1}{2\lambda}\left[\|y'\|^2 + \|x' - x\|^2\right] - \frac{1}{\lambda}\langle w', y' - y\rangle \quad \forall z' = (x', y') \in \mathbb{E}, \forall w' \in \mathbb{R}^m$$

and note that for $\tilde{z} := (\tilde{x}, \tilde{y})$, we have

$$\nabla_{z'} \mathcal{Q}(\tilde{z}, w') = \left(\nabla_x \mathcal{L}(\tilde{x}, w') + (1/\lambda)(\tilde{x} - x), \frac{1}{\lambda}(\tilde{y} - w')\right) \quad \forall w' \in \mathbb{R}^m_+. \tag{76}$$

Since (a) is equivalent to (74), it follows from the last identity that (a) is equivalent to

$$(0, 0) = \nabla_{z'} \mathcal{Q}(\tilde{z}, \tilde{y}), \quad 0 \in -\left(g(\tilde{x}) - \frac{1}{\lambda}(\tilde{y} - y)\right) + N_{\mathbb{R}^m_+}(\tilde{y}).$$

which is a necessary and sufficient condition for $\tilde{z}$ to be an optimal solution of (73) and $\tilde{y}$ to be an associated Lagrange multiplier. Noting that the conditions $w' \in \mathbb{R}^m_+$ and $\nabla_{z'} \mathcal{Q}(\tilde{z}, w') = 0$ necessarily imply that $w' = \tilde{y}$ in view of (76), we conclude that (a) and (c) are equivalent, and that the last statement of the proposition holds. □

Another equivalent characterization of condition O.3 is that (x^*, y^*) be a saddle point of the Lagrangian function $\mathcal{L}$, i.e., $y^* \in \mathbb{R}^m_+$ and

$$(0,0) \in \partial \left[\mathcal{L}(\cdot, y^*) - \mathcal{L}(x^*, \cdot)\right](x^*, y^*).$$

The following notion of an approximate saddle-point can be used as an approximate solution of (64).

Definition 5.2. *Given $(\bar\rho, \bar\varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$, $v = (p, q) \in \mathbb{E}$ and $\varepsilon \in \mathbb{R}_+$, the triple (z, v, ε) is said to be a $(\bar\rho, \bar\varepsilon)$-saddle point of $\mathcal{L}$ if $\|v\| \le \bar\rho$, $\varepsilon \le \bar\varepsilon$ and*

$$v = (p, q) \in T(z; \varepsilon) := \partial_\varepsilon \left[\mathcal{L}(\cdot, y) - \mathcal{L}(x, \cdot)\right](x, y). \tag{77}$$

Clearly, when $\bar\rho = \bar\varepsilon = 0$, the above notion is equivalent to z being a saddle point of $\mathcal{L}$ and the residual pair (v, ε) being zero, i.e., $(v, \varepsilon) = (0, 0)$. We will see later that the HPE variants discussed in this paper generate a sequence of triples $(z_k, v_k, \varepsilon_k)$ satisfying (77) and hence they will find a $(\bar\rho, \bar\varepsilon)$-saddle point when the conditions $\|v_k\| \le \bar\rho$ and $\varepsilon_k \le \bar\varepsilon$ are met. Note that the ability of these instances to also generate the sequence $\{\varepsilon_k\}$ is important since the condition that $v_k \in T(z_k; \bar\varepsilon)$ can not be easily checked in general.

In the two subsections of this section, we will present algorithms and study their complexity for approximately solving (64) in the sense of Definition 5.2. The following technical but simple result gives another characterization of (77) which is closer to the optimality condition (65) for problem (64).

Lemma 5.3. *Let $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$, $v = (p, q) \in \mathbb{E}$ and $\varepsilon \ge 0$ be given and define*

$$\varepsilon' := \varepsilon + \langle q + g(x), y\rangle. \tag{78}$$

Then, the following conditions are equivalent:

(a) inclusion (77) *holds;*

(b) there hold

$$q + g(x) \le 0, \quad \langle q + g(x), y\rangle \ge -\varepsilon, \quad p \in \partial_{x,\varepsilon'} \mathcal{L}(x, y); \tag{79}$$

(c) $0 \le \varepsilon' \le \varepsilon$ and

$$q \in -g(x) + N^{\varepsilon - \varepsilon'}_{\mathbb{R}^m_+}(y), \quad p \in \partial_{x,\varepsilon'} \mathcal{L}(x, y). \tag{80}$$

Proof. First note that, in view of (78), we have $\varepsilon' \ge 0$ if and only if the second inequality in (79) holds.

(a) $\Longleftrightarrow$ (b). Inclusion (77) is equivalent to

$$\mathcal{L}(x', y) - \mathcal{L}(x, y') \ge \langle p, x' - x\rangle + \langle q, y' - y\rangle - \varepsilon \quad \forall (x', y') \in \mathbb{R}^n \times \mathbb{R}^m_+, \tag{81}$$

which, in view of (66) and (78), is clearly equivalent to

$$\mathcal{L}(x', y) - \mathcal{L}(x, y) \ge \langle p, x' - x\rangle - \varepsilon' + \sup_{y' \ge 0} \langle q + g(x), y'\rangle \quad \forall x' \in \mathbb{R}^n.$$

Since the above inequality is obviously equivalent to (79), the equivalence of (a) and (b) follows.

(b) $\Longleftrightarrow$ (c). Using Proposition 2.1 with $K = \mathbb{R}^m_+$, $u = q + g(x)$ and $\rho = \varepsilon - \varepsilon'$, we easily see that the first inequality in (79) is equivalent to the inequality $\varepsilon' \le \varepsilon$ and the first inclusion in (80). The equivalence now follows due to the observation made at the beginning of this proof. $\square$

Clearly, when $\varepsilon = 0$, condition (79) reduces to

$$p = \nabla_x \mathcal{L}(x,y), \quad q + g(x) \le 0, \quad \langle q + g(x), y\rangle = 0. \tag{82}$$

5.1 Primal-dual regularized extragradient SQP (re-SQP) method

This subsection presents the first instance of the GLS-IPE framework in which $\mathcal{A}_k$ is chosen as the first-order approximation of F at the projection of z_{k-1} onto Ω.

Throughout this subsection, we let

$$L_g := (L_1, \ldots, L_m)^T \tag{83}$$

and

$$\theta(x) = \sqrt{\sum_{i=1}^m \|\nabla^2 f_i(x)\|^2} \quad \forall x \in \mathbb{R}^n. \tag{84}$$

Lemma 5.4. *For any $x, x' \in \mathbb{R}^n$ and $y, y' \in \mathbb{R}^m_+$, we have:*

$$\|\nabla_x \mathcal{L}(x',y') - \nabla_x \mathcal{L}(x,y') - \nabla^2_{xx}\mathcal{L}(x,y')(x'-x)\| \le \frac{L_0 + \langle L_g, y'\rangle}{2}\|x'-x\|^2, \tag{85}$$

$$\|\nabla^2_{xx}\mathcal{L}(x,y') - \nabla^2_{xx}\mathcal{L}(x,y)\| \le \theta(x)\|y'-y\|, \tag{86}$$

$$|\theta(x') - \theta(x)| \le \|L_g\|\,\|x'-x\|. \tag{87}$$

Proof. Let $x, x' \in \mathbb{R}^n$ and $y, y' \in \mathbb{R}^m_+$ be given. Using (66) and conditions O.1 and O.2, we conclude that

$$\begin{aligned}\|\nabla^2_{xx}\mathcal{L}(x',y') - \nabla^2_{xx}\mathcal{L}(x,y')\| &= \left\|\nabla^2 f_0(x') - \nabla^2 f_0(x) + \sum_{i=1}^m y'_i\left[\nabla^2 f_i(x') - \nabla^2 f_i(x)\right]\right\| \\ &\le \big(L_0 + \langle L_g, y'\rangle\big)\ \|x'-x\|\end{aligned}$$

which, in view of Lemma A.1 with $H(\cdot) = \nabla_x \mathcal{L}(\cdot, y')$, immediately implies (85). Moreover, (66) and well-known norm properties imply that

$$\|\nabla^2_{xx}\mathcal{L}(x,y') - \nabla^2_{xx}\mathcal{L}(x,y)\| = \left\|\sum_{i=1}^m (y'_i - y_i)\nabla^2 f_i(x)\right\| \le \sum_{i=1}^m |y'_i - y_i|\ \left\|\nabla^2 f_i(x)\right\|$$

which, together with the Cauchy-Schwarz inequality and (84), yields (86). Finally, (84), (83), condition O.1 and the triangle inequality for norms imply that

$$|\theta(x') - \theta(x)| \le \sqrt{\sum_{i=1}^m \|\nabla^2 f_i(x') - \nabla^2 f_i(x)\|^2} \le \sqrt{\sum_{i=1}^m L_i^2\,\|x'-x\|^2} = \|L_g\|\|x'-x\|.$$

□

The following result gives some facts about the first-order approximation of F at a given point $z \in \mathbb{E}$ in light of our GLS-IPE framework.

Lemma 5.5. *Let Ω be as in* (71) *and let $z = (x, y) \in \mathbb{E}$ given. Define $\beta_z := 2$, the linear approximation $\mathcal{A}_z : \mathbb{E} \to \mathbb{E}$ of F at z with respect to Ω as*

$$\mathcal{A}_z(z') := F(x, y^+) + F'(x, y^+) \begin{pmatrix} x' - x \\ y' - y^+ \end{pmatrix} \quad \forall z' = (x', y') \in \mathbb{E} \tag{88}$$

and the map $\phi_z : \mathbb{R}_+ \to \mathbb{R}_+$ as

$$\phi_z(t) := \left(\frac{L_0 + \langle L_g, y^+ \rangle + 3\theta(x)}{2} \right) t + \frac{2\|L_g\|}{3} t^2 \quad \forall t \geq 0 \tag{89}$$

where L_0, L_g and $\theta(\cdot)$ are as in condition O.2, (83) *and* (84)*, respectively. Then, the triple $(\beta_z, \phi_z, \mathcal{A}_z)$ is an approximation of F at z and*

$$\mathcal{A}_z(z') = \begin{bmatrix} \nabla_x \mathcal{L}(x, y') + \nabla^2_{xx} \mathcal{L}(x, y^+)(x' - x) \\ -g(x) - \nabla g(x)^T (x' - x) \end{bmatrix} \quad \forall z' = (x', y') \in \Omega. \tag{90}$$

Proof. First note that (90) follows straightforwardly from (66), (69), (70) and (88).

It remains to show that the triple $(\beta_z, \phi_z, \mathcal{A}_z)$ is an approximation of F at z, or equivalently, it satisfies Definition 3.2. Indeed, using (89), it is easy to verify that ϕ_z satisfies conditions (a)-(c) of Definition 2.4 and condition (a) of Definition 3.1 with $p = 1$ and $\beta = 2$, and hence that $\phi_z \in \Psi(2) = \Psi(\beta_z)$ where the latter equality is due to the definition of β_z. In view of (88), $\mathcal{A}_z$ is a monotone affine (and hence continuous) map in view of the remark (i.e., the fifth one) following (70). We have thus shown that Definition 3.2(a) holds.

We will now show that Definition 3.2(b) holds, i.e., that (30) is satisfied. Let $z' = (x', y') \in \Omega$ be given. Letting $a_1 \in \mathbb{R}^n$ and $a_2 \in \mathbb{R}^m$ be defined as

$$a_1 = \nabla_x \mathcal{L}(x', y') - \nabla_x \mathcal{L}(x, y') - \nabla^2_{xx} \mathcal{L}(x, y^+)(x' - x), \tag{91}$$

$$a_2 = -\left(g(x') - g(x) - \nabla g(x)^T (x' - x) \right), \tag{92}$$

it follows from (90) and (69) that $F(z') - \mathcal{A}_z(z') = [a_1 \; a_2]^T$, and hence that

$$\|F(z') - \mathcal{A}_z(z')\| \leq \|a_1\| + \|a_2\|. \tag{93}$$

Since y^+ is the orthogonal projection of y onto $\mathbb{R}^m_+$, $y' \in \mathbb{R}^m_+$ and $\|z' - z\|^2 = \|x' - x\|^2 + \|y' - y\|^2$, it follows that

$$\max\{\|y' - y^+\|, \|x' - x\|\} \leq \|z' - z\| . \tag{94}$$

Using the fact that $y^+ \geq 0$, well-known norm properties, relations (91), (85), (86) with $y = y^+$, and (94), we obtain

$$\begin{aligned} \|a_1\| &\leq \|\nabla_x \mathcal{L}(x', y') - \nabla_x \mathcal{L}(x, y') - \nabla^2_{xx} \mathcal{L}(x, y')(x' - x)\| + \|[\nabla^2_{xx} \mathcal{L}(x, y') - \nabla^2_{xx} \mathcal{L}(x, y^+)](x' - x)\| \\ &\leq \frac{L_0 + \langle L_g, y' \rangle}{2} \|x' - x\|^2 + \theta(x) \|y' - y^+\| \|x' - x\| \\ &\leq \frac{L_0 + \langle L_g, y^+ \rangle + \|L_g\| \|y' - y^+\|}{2} \|x' - x\|^2 + \theta(x) \|y' - y^+\| \|x' - x\| \\ &\leq \left(\frac{L_0 + \langle L_g, y^+ \rangle}{2} + \theta(x) \right) \|z' - z\|^2 + \frac{\|L_g\|}{2} \|z' - z\|^3. \end{aligned} \tag{95}$$

Due to condition O.2 and Lemma A.2, we have for every $i = 1, \dots, m$,

$$|f_i(x') - f_i(x) - \langle \nabla f_i(x), x' - x \rangle| \leq \frac{L_i}{6}\|x' - x\|^3 + \frac{1}{2}\|\nabla^2 f_i(x)\|\|x' - x\|^2, \tag{96}$$

which, combined with the fact that $g = (f_1, \dots, f_m)^T$, relations (83), (84), (92) and (94), yields

$$\|a_2\| \leq \frac{\theta(x)}{2}\|z' - z\|^2 + \frac{\|L_g\|}{6}\|z' - z\|^3 \,. \tag{97}$$

To finish the proof, note that (30) follows immediately from (93), (95) and (97). $\square$

Proposition 5.6. *Let $z = (x, y) \in \mathbb{E}$ and $\lambda > 0$ be given and consider the map $\mathcal{A}_z$ as in* (88)*. For every $(x', y') \in \mathbb{E}$, define*

$$\mathcal{Q}_z(x') := f_0(x) + \langle \nabla f_0(x), x' - x \rangle + \frac{1}{2}\langle x' - x, \nabla^2_{xx}\mathcal{L}(x, y^+)(x' - x) \rangle, \tag{98}$$

$$\ell_x(x') := g(x) + \nabla g(x)^T(x' - x) \,. \tag{99}$$

Then the following conditions on a given pair $(\tilde{x}, \tilde{y}) \in \mathbb{E}$ are equivalent:

(a) $(\tilde{x}, \tilde{y}) = (I + \lambda(\mathcal{A}_z + B))^{-1}(x, y)$;

(b) $\tilde{x}$ *is the unique solution of*

$$\min_{x' \in \mathbb{R}^n} \; \mathcal{Q}_z(x') + \frac{1}{2\lambda}\left[\left\|(y + \lambda \ell_x(x'))^+\right\|^2 - \|y\|^2\right] + \frac{1}{2\lambda}\|x' - x\|^2 \tag{100}$$

and $\tilde{y} = (y + \lambda \ell_x(\tilde{x}))^+$;

(c) $(\tilde{x}, \tilde{y})$ *is the unique solution of*

$$\begin{aligned} \min_{(x', y') \in \mathbb{E}} \; & \mathcal{Q}_z(x') + \frac{1}{2\lambda}\left[\|y'\|^2 + \|x' - x\|^2\right] \\ \text{s.t.} \;\; & \ell_x(x') - \frac{1}{\lambda}(y' - y) \leq 0, \end{aligned} \tag{101}$$

in which case $\tilde{y}$ is the unique Lagrange multiplier of (101)*.*

Proof. Applying Proposition 5.1 with $f_0 = \mathcal{Q}_z$ and $g = \ell_x$, we conclude that (b) and (c) are equivalent and that they are both equivalent to (a) of Proposition 5.1 with $F = \tilde{F}$ where $\tilde{F}$ is given by

$$\tilde{F}(x', y') := \begin{bmatrix} \nabla \mathcal{Q}_z(x') + \nabla \ell_x(x') y' \\ -\ell_x(x') \end{bmatrix} = \begin{bmatrix} \nabla \mathcal{Q}_z(x') + \nabla g(x) y' \\ -\ell_x(x') \end{bmatrix} \quad \forall (x', y') \in \mathbb{E}. \tag{102}$$

The result now follows by noting that the above $\tilde{F}$ is equal to $\mathcal{A}_z$ due to (66), (90), (98) and (99). $\square$

We note that the proof of Proposition 5.6 implies that $\mathcal{A}_z$ can also be viewed as the "F map" (see relation (68)) for the quadratic programming (QP) approximation

$$\min\{\mathcal{Q}_z(x') : \ell_x(x') \leq 0\} \tag{103}$$

of (64). Note also that the constraint map of the above QP is the linearization of g at x and its quadratic objective function is obtained by adding the linearization of f_0 at x with the quadratic term centered at x whose Hessian equal to that of $\mathcal{L}(\cdot, y^+)$ evaluated at x.

We now state the first instance of the GLS-IPE framework based on the approximation $\mathcal{A}_k$ of Lemma 5.5.

Primal-dual regularized extragradient SQP (re-SQP) method:

(0) Choose $z_0 = (x_0, y_0) \in \mathbb{E}$, $0 < \sigma_\ell < \sigma_u < 1$, and set $k \leftarrow 1$;

(1) if $z_{k-1} = (x_{k-1}, y_{k-1})$ is a solution of (67), or equivalently $0 \in (F + B)(z_{k-1})$ where F and B are defined in (68), then STOP;

(2) otherwise, compute $\lambda_k > 0$ such that the unique solution of the quadratic programming (101) with $z = z_{k-1}$, or equivalently the pair $\tilde{z}_k = (\tilde{x}_k, \tilde{y}_k)$ given by

$$\tilde{z}_k = (I + \lambda_k(\mathcal{A}_{z_{k-1}} + B))^{-1}(z_{k-1}), \tag{104}$$

satisfies

$$\sigma_\ell \le \lambda_k \phi_{z_{k-1}}(\|\tilde{z}_k - z_{k-1}\|) \le \sigma_u \tag{105}$$

where $\mathcal{A}_z$ and ϕ_z are defined in (88) and (89), respectively;

(3) set $z_k = z_{k-1} - \lambda_k v_k$ where

$$v_k = \begin{pmatrix} p_k \\ q_k \end{pmatrix} := \begin{pmatrix} \nabla_x \mathcal{L}(\tilde{x}_k, \tilde{y}_k) \\ \lambda_k^{-1}(y_{k-1} - \tilde{y}_k) - \left[g(\tilde{x}_k) - g(x_{k-1}) - \nabla g(x_{k-1})^T (\tilde{x}_k - x_{k-1}) \right] \end{pmatrix};$$

(4) let $k \leftarrow k + 1$ and go to step 1.

end

We now make some remarks regarding the re-SQP method. First, the triple $(\tilde{z}_k, u_k, \varepsilon_k)$ where $\tilde{z}_k$ is given by (104) and $(u_k, \varepsilon_k) = ((z_{k-1} - \tilde{z}_k)/\lambda_k, 0)$ is easily seen to satisfy (32) with $\hat{\sigma}_k = 0$. Hence, the ability to compute $\tilde{z}_k$ as in (104) for any $\lambda_k > 0$ yields a black-box satisfying Assumption B-B with $\hat{\sigma}_k = 0$. Second, the task of finding $\lambda_k > 0$ and $\tilde{z}_k \in \mathbb{E}$ satisfying (104) and (105) corresponds to solving (32) and (33) with $\varepsilon_k = 0$ and $\hat{\sigma} = 0$. Hence, the stepsize λ_k and $\tilde{z}_k$ as in step 2 of the re-SQP method can be found via the Bracketing/Bisection procedure of Subsection 4.2, whose number of black-box calls (i.e., resolvent computations) is bounded according to Lemma 4.7. Third, the equivalence between (a), (b) and (c) of Proposition 5.6 provides alternative ways of computing $\tilde{z}_k$ in step 2 of the above method.

The following result shows that the re-SQP method is a special case of the GLS-IPE framework of Section 3 which satisfies condition (40) with ψ given by (107).

Proposition 5.7. *Consider the sequences* $\{\lambda_k\}$, $\{z_k\}$, $\{\tilde{z}_k\}$, $\{\phi_k = \phi_{z_{k-1}}\}$ *and* $\{\mathcal{A}_k = \mathcal{A}_{z_{k-1}}\}$

generated by the re-SQP method with F and B given by (68)*, and define*

$$\hat{\sigma}_k = 0, \quad \beta_k = 2, \quad \varepsilon_k = 0, \quad u_k = \frac{z_{k-1} - \tilde{z}_k}{\lambda_k} \qquad \forall k \geq 1 \tag{106}$$

and the function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ as

$$\psi(t) = \left(\frac{L_0 + \langle L_g, y_0^+ \rangle + 3\theta(x_0) + 2\sqrt{10}\|L_g\| d_0}{2} \right) t + \frac{2\|L_g\|}{3} t^2, \tag{107}$$

where d_0 is the distance of z_0 to $(F + B)^{-1}(0)$ and L_0, L_g and $\theta(\cdot)$ are as in condition O.2, (83) *and* (84)*, respectively. Then, the following statements hold for every $k \geq 1$:*

(a) the triple $(\beta_k, \phi_k, \mathcal{A}_k)$ is an approximation of F at z_{k-1};

(b) relations (31) *with $\sigma = \sigma_u$,* (32)*,* (33) *and* (34) *hold.*

As a consequence, the re-SQP method is a special instance of the GLS-IPE framework with $\sigma = \sigma_u$ and the sequences $\{\hat{\sigma}_k\}$, $\{\beta_k\}$ and $\{\varepsilon_k\}$ given by (106)*. Moreover, the re-SQP method satisfies condition* (40) *with ψ given by* (107)*.*

Proof. (a) This statement follows from step 2 of the re-SQP method, Lemma 5.5 and the definition of β_k in (106).

(b) Using the definition of $\hat{\sigma}_k$ in (106) and the assumption that $\sigma_\ell < \sigma_u$ (see step 0 of the re-SQP method), we easily see that (31) holds with $\sigma = \sigma_u$. Relation (104) and the definitions of u_k, ε_k and $\hat{\sigma}_k$ in (106) immediately imply (32). Moreover, (33) follows from (105). To show (34), it suffices to show in view of step 3 of the re-SQP method that $F(\tilde{z}_k) + u_k - \mathcal{A}_k(\tilde{z}_k) = v_k$. Indeed, using the inclusion in (32) with $\varepsilon_k = 0$ and the fact that the x-component of $B(\tilde{z}_k)$ is zero by (68), we conclude that the x-component of $F(\tilde{z}_k) + u_k - \mathcal{A}_k(\tilde{z}_k)$ is equal to the x-component of $F(\tilde{z}_k)$, which in turn is equal to p_k in view of the fact that $\tilde{y}_k \in \mathbb{R}^m_+$, (69) and the definition of p_k in step 3 of the re-SQP. Likewise, using the definition of F and u_k in (68) and (106), respectively, and relation (90) with $z = z_{k-1}$ and $z' = \tilde{z}_k$, we easily see that the y-component of $F(\tilde{z}_k) + u_k - \mathcal{A}_k(\tilde{z}_k)$ is equal to q_k.

We have thus shown that the re-SQP method is a special case of the GLS-IPE framework with sequences $\{\hat{\sigma}_k\}$ and $\{\beta_k\}$ given by (106). Hence, it follows by Lemma 3.4(a) that the re-SQP method is a special case of the HPE framework, and hence satisfies the conclusions of Proposition 2.5, and in particular condition (17). Using the latter condition, the Cauchy-Schwarz inequality and inequality (87) with $x' = x_{k-1}$ and $x = x_0$, we conclude that

$$\begin{aligned}
\langle L_g, y_{k-1}^+ \rangle + 3\theta(x_{k-1}) &= \langle L_g, y_{k-1}^+ - y_0^+ \rangle + \langle L_g, y_0^+ \rangle + 3[\theta(x_{k-1}) - \theta(x_0)] + 3\theta(x_0) \\
&\leq \langle L_g, y_0^+ \rangle + 3\theta(x_0) + \|L_g\| \|y_{k-1} - y_0\| + 3\|L_g\| \|x_{k-1} - x_0\| \\
&\leq \langle L_g, y_0^+ \rangle + 3\theta(x_0) + \sqrt{10}\|L_g\| \|z_{k-1} - z_0\| \\
&\leq \langle L_g, y_0^+ \rangle + 3\theta(x_0) + 2\sqrt{10}\|L_g\| d_0 \,.
\end{aligned}$$

The conclusion that the re-SQP method satisfies condition (40) now follows immediately from the above inequality and the definitions of $\phi_z(\cdot)$ and $\psi(\cdot)$ in (89) and (107), respectively. $\square$

Lemma 5.8. *Consider the sequences $\{\lambda_k\}$, $\{\tilde{z}_k\}$ and $\{v_k\}$ generated by the re-SQP method and define the ergodic sequences $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ as in* (19) *and* (20) *with $\varepsilon_k = 0$ for all k. Then, $\varepsilon_k^a \geq 0$ and the triples $(z, v, \varepsilon) = (\tilde{z}_k, v_k, 0)$ and $(z, v, \varepsilon) = (\tilde{z}_k^a, v_k^a, \varepsilon_k^a)$ satisfy* (77) *with $\mathcal{L}(\cdot, \cdot)$ given by* (66)*.*

Proof. Since by Proposition 5.7, the re-SQP method is a special case of the GLS-IPE framework with $\varepsilon_k = 0$ for all k, it follows from Lemma 3.4(a) that $v_k \in (F+B)(\tilde{z}_k)$ for every $k \geq 1$. Clearly, due to (68) and (69), the latter inclusion is equivalent to (82) with $(x,y) = \tilde{z}_k$ and $(p,q) = v_k$. In view of the equivalence of (a) and (b) of Lemma 5.3 with $\varepsilon = 0$, the latter condition is in turn equivalent to the triple $(\tilde{z}_k, v_k, 0)$ satisfying (77). The claim that $\varepsilon_k^a \geq 0$ and $(z, v, \varepsilon) = (\tilde{z}_k^a, v_k^a, \varepsilon_k^a)$ also satisfy (77) now follows immediately from Proposition A.3 with $\Gamma = \mathcal{L}$, $X = \mathbb{R}^n$, $Y = \mathbb{R}^m_+$, $\varepsilon_i = 0$ and $\alpha_i = \lambda_i/\Lambda_k$ for all $i = 1, \ldots, k$. $\square$

As an immediate consequence of Theorem 3.5 and Proposition 5.7, we obtain the following outer convergence rate result for the re-SQP method.

Theorem 5.9. *Assume that* $\max\{\sigma_\ell^{-1}, (1-\sigma_u)^{-1}\} = \mathcal{O}(1)$ *and consider the sequences* $\{\lambda_k\}$, $\{\tilde{z}_k\}$ *and* $\{v_k\}$ *generated by the re-SQP method and the ergodic sequences* $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ *and* $\{\varepsilon_k^a\}$ *defined in Lemma 5.8 with* $\varepsilon_k = 0$ *for all* k. *Let* d_0 *denote the distance of* z_0 *to the solution set of* (26), L_0, L_g *and* $\theta(\cdot)$ *be as in condition O.2,* (83) *and* (84), *respectively, and define*

$$m_0 := \max\left\{L_0, \langle L_g, y_0^+\rangle, \theta(x_0)\right\}. \tag{108}$$

Then, for every $k \geq 1$*:*

(a) (pointwise convergence rate) the triple $(\tilde{z}_k, v_k, 0)$ *satisfies* (77) *and there exists an index* $i \leq k$ *such that*

$$\|v_i\| \leq \mathcal{O}\left(\frac{d_0^2}{k}\max\{m_0, \|L_g\|d_0\}\right); \tag{109}$$

(b) (ergodic convergence rate) the triple $(\tilde{z}_k^a, v_k^a, \varepsilon_k^a)$ *satisfies* (77) *and*

$$\|v_k^a\| \leq \mathcal{O}\left(\frac{d_0^2}{k^{3/2}}\max\{m_0, \|L_g\|d_0\}\right), \quad 0 \leq \varepsilon_k^a \leq \mathcal{O}\left(\frac{d_0^3}{k^{3/2}}\max\{m_0, \|L_g\|d_0\}\right). \tag{110}$$

Proof. We know from Proposition 5.7 that the re-SQP method is a special instance of the GLS-IPE framework with $\sigma = \sigma_u$ and it satisfies condition (40) with ψ given by (107). Hence, in view of the assumption that $\max\{\sigma_\ell^{-1}, (1-\sigma_u)^{-1}\} = \mathcal{O}(1)$ and Theorem 3.5, we conclude that for every $k \geq 1$, there exists $i \leq k$ such that

$$\|v_i\| \leq \mathcal{O}\left(\frac{d_0}{\sqrt{k}}\,\psi\left(\frac{d_0}{\sqrt{k}}\right)\right),$$

and

$$\|v_k^a\| \leq \mathcal{O}\left(\frac{d_0}{k}\,\psi\left(\frac{d_0}{\sqrt{k}}\right)\right), \quad 0 \leq \varepsilon_k^a \leq \mathcal{O}\left(\frac{d_0^2}{k}\,\psi\left(\frac{d_0}{\sqrt{k}}\right)\right).$$

These bounds together with relations (107) and (108), and Lemma 5.8, now imply the conclusions of the theorem. $\square$

Corollary 5.10. *Assume that* $\max\{\sigma_\ell^{-1}, (1-\sigma_u)^{-1}\} = \mathcal{O}(1)$ *and consider the re-SQP method in which, at every iteration* k, *the Bracketing/Bisection procedure with initial stepsize* $\lambda_k^0 = 1$ *is used to find a stepsize* $\lambda_k > 0$ *satisfying the conditions of step 2. Let* d_0 *denote the distance of* z_0 *to the solution set of* (26) *and let* L_g *and* m_0 *be as in* (83) *and* (108), *respectively. Then, given* $\bar{\rho} > 0$ *and* $\bar{\varepsilon} > 0$, *the re-SQP will find:*

(a) a $(\bar{\rho}, 0)$-saddle point of $\mathcal{L}$ in at most

$$\mathcal{O}\left(\frac{d_0^2}{\bar{\rho}} \max\{m_0, \|L_g\| d_0\}\right) \tag{111}$$

iterations;

(b) a $(\bar{\rho}, \bar{\varepsilon})$-saddle point of $\mathcal{L}$ in at most

$$\mathcal{O}\left(\max\left\{m_0^{2/3}, \|L_g\|^{2/3} d_0^{2/3}\right\} \max\left\{\frac{d_0^{4/3}}{(\bar{\rho})^{2/3}}, \frac{d_0^2}{(\bar{\varepsilon})^{2/3}}\right\}\right) \tag{112}$$

iterations.

Moreover, if in addition $[\log(\sigma_u/\sigma_\ell)]^{-1} = \mathcal{O}(1)$, then each iteration will make at most

$$\mathcal{O}\left(1 + \log^+\left[\log\left(1 + m_0 + d_0 + \|L_g\| + (L_0\bar{\rho})^{-1}\right)\right]\right) \tag{113}$$

black-box calls.

Proof. Statements (a) and (b) of the corollary follow immediately from (a) and (b) of Theorem 5.9 and the concept of $(\bar{\rho}, \bar{\varepsilon})$- saddle point of Definition 5.2.

We will now prove the last assertion of the corollary. First note that, by Proposition 5.7, the re-SQP method is a special instance of the GLS-IPE framework which satisfies (40) with ψ given by (107) and for which $\hat{\sigma}_k = 0$, $\beta_k = 2$ and $\varepsilon_k = 0$ for all $k \geq 1$. By Theorem 4.11, we conclude that at the k-th iteration of the re-SQP method, one of the following two cases occur: i) $\tilde{z}_k^0$ and the vector v_k^0 as in Lemma 4.10 satisfy statement (b) of the latter lemma, i.e., $v_k^0 \in (F+B)(\tilde{z}_k^0)$ and $\|v_k^0\| \leq \bar{\rho}$, or; ii) the Bracketing/Bisection procedure performs a number of black-box calls bounded by (63). In the case i), only one black-box call is performed and the triple $(\tilde{z}_k^0, v_k^0, 0)$ is clearly a $(\bar{\rho}, 0)$-saddle point of $\mathcal{L}$. In the case ii), relations (40), (63), (106) and (107), and the assumption that $[\log(\sigma_u/\sigma_\ell)]^{-1} = \mathcal{O}(1)$, imply that the number of black-box calls is bounded

$$\mathcal{O}\left(1 + \log^+\left[\log\left(\max\{\psi([1 + \psi(d_0)]d_0), \zeta_k^{-1}\}\right)\right]\right) \tag{114}$$

where ζ_k is as in (57). Moreover, (57), (58), (106), the fact that $\lambda_k^0 = 1$ and (89) with $z = z_{k-1}$ imply that

$$\zeta_k = \min\{1, \phi_k(\bar{\rho}/2)\} \geq \min\{1, L_0\bar{\rho}/4\}. \tag{115}$$

The result now follows by combining (114) and (115) and using the definitions of m_0 and ψ in (108) and (107), respectively. □

5.2 Primal-dual regularized extragradient SQCQP (re-SQCQP) method

In this subsection, we present the second instance of the GLS-IPE framework in which $\mathcal{A}_k = \mathcal{A}_{z_{k-1}}$ is chosen according to Lemma 5.11 below. The results presented here are all analogues of the ones presented in the previous subsection and, as a consequence, we will skip some details of their proofs whenever we feel that clarity and rigor is not sacrificed.

As we have seen in the remark below Proposition 5.6, the map $\mathcal{A}_z$ for the re-SQP method of Subsection 5.1 can be viewed as the "F map" of the QP approximation (103) of problem (64). On the other hand, the $\mathcal{A}_z$ approximation used in the second instance of the GLS-IPE framework studied in this subsection is the "F map" of the convex quadratic constrained QP approximation

$$\min\{q_x^0(x') : q_x(x') \le 0\} \tag{116}$$

of problem (64), where

$$q_x(x') = (q_x^1(x'), \dots, q_x^m(x')) \quad \forall x' \in \mathbb{R}^n \tag{117}$$

and

$$q_x^i(x') := f_i(x) + \langle \nabla f_i(x), x' - x\rangle + \frac{1}{2}\langle x' - x, \nabla^2 f_i(x)(x' - x)\rangle \quad \forall x' \in \mathbb{R}^n,\ \forall i = 0, \dots, m\,. \tag{118}$$

The following result explicitly introduces this approximation and states its fundamental properties which are essentially analogues of the ones stated in Lemma 5.5.

Lemma 5.11. *Let Ω be as in (71) and let $z = (x, y) \in \mathbb{E}$ given. Define $\beta_z := 2$ and $\mathcal{A}_z : \mathbb{E} \to \mathbb{E}$ by*

$$\mathcal{A}_z(z') := \begin{bmatrix} \nabla q_x^0(x') + \nabla q_x(x')y' \\ -q_x(x') \end{bmatrix} \quad \forall z' = (x', y') \in \mathbb{E} \tag{119}$$

and the map $\phi_z : \mathbb{R}_+ \to \mathbb{R}_+$ as

$$\phi_z(t) := \left(\frac{L_0 + \langle L_g, y^+\rangle}{2}\right) t + \frac{2\|L_g\|}{3} t^2 \quad \forall t \ge 0 \tag{120}$$

where L_0 and L_g are defined in condition O.2 and (83)*, respectively. Then, the triple $(\beta_z, \phi_z, \mathcal{A}_z)$ is an approximation of F at z and*

$$\mathcal{A}_z(z') = \begin{bmatrix} \nabla_x \mathcal{L}(x, y') + \nabla^2_{xx}\mathcal{L}(x, y')(x' - x) \\ -q_x(x') \end{bmatrix} \quad \forall z' = (x', y') \in \Omega. \tag{121}$$

Proof. First note that (121) follows straightforwardly from (119), (118), (117) and (66).

It remains to show that the triple $(\beta_z, \phi_z, \mathcal{A}_z)$ is an approximation of F at z, or equivalently, it satisfies Definition 3.2. Indeed, using (120), it is easy to verify that ϕ_z satisfies conditions (a)-(c) of Definition 2.4 and condition (a) of Definition 3.1 with $p = 1$ and $\beta = 2$, and hence that $\phi_z \in \Psi(2) = \Psi(\beta_z)$ where the latter equality is due to the definition of β_z. It is easy to see that $\mathcal{A}_z$ in (119) is the "F map" of the minimization problem (116). This observation together with the remark (i.e., the fifth one) following (70) and the fact that the functions q_x^i, $i = 0, \dots, m$, are convex quadratic imply that $\mathcal{A}_z$ is monotone and continuous on Ω. We have thus shown that Definition 3.2(a) holds.

We will now show that Definition 3.2(b) holds, i.e., that (30) is satisfied. Let $z' = (x', y') \in \Omega$ be given. Letting $b_1 \in \mathbb{R}^n$ and $b_2 \in \mathbb{R}^m$ be defined as

$$b_1 = \nabla_x \mathcal{L}(x', y') - \nabla_x \mathcal{L}(x, y') - \nabla^2_{xx}\mathcal{L}(x, y')(x' - x), \tag{122}$$

$$b_2 = -(g(x') - q_x(x')), \tag{123}$$

it follows from (121) and (69) that $F(z') - \mathcal{A}_z(z') = [b_1\ b_2]^T$, and hence that

$$\|F(z') - \mathcal{A}_z(z')\| \le \|b_1\| + \|b_2\|. \tag{124}$$

Using relations (122), (85) and the fact that $\max\{\|y'-y^+\|, \|x'-x\|\} \le \|z'-z\|$ we obtain

$$\begin{aligned}\|b_1\| &\le \left(\frac{L_0 + \langle L_g, y'\rangle}{2}\right)\|z'-z\|^2 \\ &\le \left(\frac{L_0 + \langle L_g, y^+\rangle}{2}\right)\|z'-z\|^2 + \frac{\|L_g\|}{2}\|z'-z\|^3. \end{aligned} \tag{125}$$

Moreover, due to condition O.2, (118) and Lemma A.2, we have for every $i = 1, \dots, m$,

$$|f_i(x') - q_x^i(x')| \le \frac{L_i}{6}\|x'-x\|^3, \tag{126}$$

which, combined with the fact that $g = (f_1, \dots, f_m)^T$, relations (83), (117) and (123), yields

$$\|b_2\| \le \frac{\|L_g\|}{6}\|z'-z\|^3\,. \tag{127}$$

The conclusion of (b) now follows immediately from (124), (125), (127) and (120). ☐

The following result is the analogue of Proposition 5.6.

Proposition 5.12. *Let $z = (x, y) \in \mathbb{E}$ and $\lambda > 0$ be given and consider the map $\mathcal{A}_z$ as in* (119) *and the functions q_x^0 and q_x as in* (118) *and* (117)*, respectively.*

Then the following conditions on a given pair $(\tilde{x}, \tilde{y}) \in \mathbb{E}$ are equivalent:

(a) $(\tilde{x}, \tilde{y}) = (I + \lambda(\mathcal{A}_z + B))^{-1}(x, y)$;

(b) $\tilde{x}$ is the unique solution of

$$\min_{x' \in \mathbb{R}^n} q_x^0(x') + \frac{1}{2\lambda}\left[\left\|(y + \lambda\, q_x(x'))^+\right\|^2 - \|y\|^2\right] + \frac{1}{2\lambda}\|x'-x\|^2 \tag{128}$$

and $\tilde{y} = (y + \lambda q_x(\tilde{x}))^+$;

(c) $(\tilde{x}, \tilde{y})$ is the unique solution of

$$\begin{aligned} \min_{(x',y') \in \mathbb{E}} \; & q_x^0(x') + \frac{1}{2\lambda}\left[\|y'\|^2 + \|x'-x\|^2\right] \\ \text{s.t.} \; & q_x(x') - \frac{1}{\lambda}(y'-y) \le 0, \end{aligned} \tag{129}$$

in which case $\tilde{y}$ is the unique Lagrange multiplier of (129).

Proof. It follows from Proposition 5.1 with $f_0 = q_x^0$ and $g = q_x$ and the definition of $\mathcal{A}_z$ in (119) that (b) and (c) are equivalent and that they are both equivalent to (a) of Proposition 5.1 with $F = \mathcal{A}_z$. ☐

We now state the second instance of the GLS-IPE framework based on the approximation $\mathcal{A}_k$ of Lemma 5.11.

Primal-dual regularized extragradient SQCQP (re-SQCQP) method:

(0) Choose $z_0 = (x_0, y_0) \in \mathbb{E}$, $0 < \sigma_\ell < \sigma_u < 1$, and set $k \leftarrow 1$;

(1) if $z_{k-1} = (x_{k-1}, y_{k-1})$ is a solution of (67), or equivalently $0 \in (F + B)(z_{k-1})$ where F and B are defined in (68), then STOP;

(2) otherwise, compute $\lambda_k > 0$ such that the unique solution of the quadratically constrained quadratic programming (129) with $z = z_{k-1}$, or equivalently the pair $\tilde{z}_k = (\tilde{x}_k, \tilde{y}_k)$ given by

$$\tilde{z}_k = (I + \lambda_k(\mathcal{A}_{z_{k-1}} + B))^{-1}(z_{k-1}), \tag{130}$$

satisfies

$$\sigma_\ell \leq \lambda_k \phi_{z_{k-1}}(\|\tilde{z}_k - z_{k-1}\|) \leq \sigma_u, \tag{131}$$

where $\mathcal{A}_z$ and ϕ_z are defined in (119) and (120), respectively;

(3) set $z_k = z_{k-1} - \lambda_k v_k$ where

$$v_k = \begin{pmatrix} p_k \\ q_k \end{pmatrix} := \begin{pmatrix} \nabla_x \mathcal{L}(\tilde{x}_k, \tilde{y}_k) \\ \lambda_k^{-1}(y_{k-1} - \tilde{y}_k) - \left[g(\tilde{x}_k) - q_{x_{k-1}}(\tilde{x}_k)\right] \end{pmatrix};$$

(4) let $k \leftarrow k + 1$ and go to step 1.

end

The following three results are the analogues of Proposition 5.7, Theorem 5.9 and Corollary 5.15, respectively. Since their proofs are quite similar to the ones of the latter results, we skip most of their details.

Proposition 5.13. *Consider the sequences $\{\lambda_k\}$, $\{z_k\}$, $\{\tilde{z}_k\}$, $\{\phi_k = \phi_{z_{k-1}}\}$ and $\{\mathcal{A}_k = \mathcal{A}_{z_{k-1}}\}$ generated by the re-SQCQP method with F and B given by* (68)*, and define $\hat{\sigma}_k$, β_k, ε_k and u_k as in* (106) *and the function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ as*

$$\psi(t) = \left(\frac{L_0 + \langle L_g, y_0^+ \rangle + 2\|L_g\| d_0}{2}\right) t + \frac{2\|L_g\|}{3} t^2, \tag{132}$$

where d_0 is the distance of z_0 to $(F + B)^{-1}(0)$, and L_0 and L_g are as in condition O.2 and (83)*, respectively. Then, the following statements hold for every $k \geq 1$:*

(a) the triple $(\beta_k, \phi_k, \mathcal{A}_k)$ is an approximation of F at z_{k-1};

(b) relations (31) *with $\sigma = \sigma_u$,* (32), (33) *and* (34) *hold.*

As a consequence, the re-SQCQP method is a special instance of the GLS-IPE framework with sequences $\{\hat{\sigma}_k\}$, $\{\beta_k\}$ and $\{\varepsilon_k\}$ given by (106)*. Moreover, the re-SQCQP method satisfies condition* (40) *with ψ given by* (132).

Proof. The proof follows the same outline as in Proposition 5.7. Using Lemma 5.11, (130), (131) and (119) instead of Lemma 5.5, (104), (105) and (90), respectively, we conclude that (a) and (b) hold and that the re-SQCQP method is a special instance of the GLS-IPE framework with sequences $\{\hat{\sigma}_k\}$, $\{\beta_k\}$ and $\{\varepsilon_k\}$ given by (106). By the same reasoning as in Proposition 5.7 we have that condition (17) holds. Using the latter condition and the triangle inequality, we conclude that

$$\begin{aligned}\langle L_g, y_{k-1}^+\rangle &= \langle L_g, y_0^+\rangle + \langle L_g, y_{k-1}^+ - y_0^+\rangle \\ &\leq \langle L_g, y_0^+\rangle + \|L_g\|\|y_{k-1} - y_0\| \\ &\leq \langle L_g, y_0^+\rangle + 2\|L_g\|d_0.\end{aligned}$$

The conclusion that the re-SQCQP method satisfies condition (40) now follows immediately from the above inequality and the definitions of $\phi_z(\cdot)$ and $\psi(\cdot)$ in (120) and (132), respectively.

□

Theorem 5.14. *Assume that* $\max\{\sigma_\ell^{-1}, (1-\sigma_u)^{-1}\} = \mathcal{O}(1)$ *and consider the sequences* $\{\lambda_k\}$*,* $\{\tilde{z}_k\}$ *and* $\{v_k\}$ *generated by the re-SQCQP method and the ergodic sequences* $\{\tilde{z}_k^a\}$*,* $\{v_k^a\}$ *and* $\{\varepsilon_k^a\}$ *defined in Lemma 5.8 with* $\varepsilon_k = 0$ *for all* k*. Let* d_0 *denote the distance of* z_0 *to the solution set of* (26)*,* L_0 *and* L_g *be as in condition O.2 and* (83)*, respectively, and define*

$$\widetilde{m}_0 := \max\left\{L_0, \langle L_g, y_0^+\rangle\right\}. \tag{133}$$

Then, all the conclusions of Theorem 5.9 hold with m_0 *replaced by* $\widetilde{m}_0$*.*

Proof. The proof is similar to that of Theorem 5.9 except that it uses $\widetilde{m}_0$, Proposition 5.13, (132) and (133) in place of m_0, Proposition 5.7, (107) and (108), respectively. □

Corollary 5.15. *Assume that* $\max\{\sigma_\ell^{-1}, (1-\sigma_u)^{-1}\} = \mathcal{O}(1)$ *and consider the re-SQCQP method in which, at every iteration* k*, the Bracketing/Bisection procedure with initial stepsize* $\lambda_k^0 = 1$ *is used to find a stepsize* $\lambda_k > 0$ *satisfying the conditions of step 2. Let* d_0 *and* L_g *be as in Theorem 5.14 and* $\widetilde{m}_0$ *as in* (133)*. Then, all the conclusions of Corollary 5.10 hold with* m_0 *replaced by* $\widetilde{m}_0$*.*

Proof. The proof is similar to that of Corollary 5.10 except that it uses Theorem 5.14, Proposition 5.13, $\widetilde{m}_0$ and (132) in place of Theorem 5.9, Proposition 5.7, m_0 and (107), respectively. □

A Appendix

The proofs of the following two lemmas can be found in [6, Lemmas 4.1.12 and 4.1.14].

Lemma A.1. *Let* $H : \mathbb{R}^n \to \mathbb{R}^n$ *be continuously differentiable such that there exists a nonnegative constant* L *satisfying*

$$\|H'(\tilde{x}) - H'(x)\| \leq L\|\tilde{x} - x\| \quad \forall x, \tilde{x} \in \mathbb{R}^n.$$

Then,

$$\|H(\tilde{x}) - H(x) - H'(x)(\tilde{x} - x)\| \leq \frac{L}{2}\|\tilde{x} - x\|^2 \tag{134}$$

for all $x, \tilde{x} \in \mathbb{R}^n$*.*

Lemma A.2. *For any twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ with L-Lipschitz continuous Hessian we have*

$$|f(\tilde{x}) - f(x) - \langle \nabla f(x), \tilde{x} - x \rangle - \frac{1}{2}\langle \tilde{x} - x, \nabla^2 f(x)(\tilde{x} - x)\rangle| \leq \frac{L}{6}\|\tilde{x} - x\|^3 \tag{135}$$

for all $x, \tilde{x} \in \mathbb{R}^n$.

Proposition A.3. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be given convex sets and $\Gamma : X \times Y \to \mathbb{R}$ be a function such that, for each $(x, y) \in X \times Y$, the function $\Gamma(\cdot, y) - \Gamma(x, \cdot) : X \times Y \to \mathbb{R}$ is convex. Suppose that, for $i = 1, \ldots, k$, $(\tilde{x}_i, \tilde{y}_i) \in X \times Y$ and $(p_i, q_i) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfy*

$$(p_i, q_i) \in \partial_{\varepsilon_i} \left(\Gamma(\cdot, \tilde{y}_i) - \Gamma(\tilde{x}_i, \cdot)\right)(\tilde{x}_i, \tilde{y}_i) \, .$$

Let $\alpha_1, \cdots, \alpha_k \geq 0$ be such that $\sum_{i=1}^k \alpha_i = 1$, and define

$$(\tilde{x}_k^a, \tilde{y}_k^a) = \sum_{i=1}^k \alpha_i(\tilde{x}_i, \tilde{y}_i), \quad (p_k^a, q_k^a) = \sum_{i=1}^k \alpha_i(p_i, q_i),$$

$$\varepsilon_k^a = \sum_{i=1}^k \alpha_i \left[\varepsilon_i + \langle \tilde{x}_i - \tilde{x}_k^a, p_i \rangle + \langle \tilde{y}_i - \tilde{y}_k^a, q_i \rangle\right] \, .$$

Then, $\varepsilon_k^a \geq 0$ and

$$(p_k^a, q_k^a) \in \partial_{\varepsilon_k^a} \left(\Gamma(\cdot, \tilde{y}_k^a) - \Gamma(\tilde{x}_k^a, \cdot)\right)(\tilde{x}_k^a, \tilde{y}_k^a) \, .$$

Proof. See [12, Proposition 5.1] . □

References

[1] A. Auslender. An extended sequential quadratically constrained quadratic programming algorithm for nonlinear, semidefinite, and second-order cone programming. *J. Optim. Theory Appl.*, 156(2):183–212, 2013.

[2] A. Auslender. A very simple SQCQP method for a class of smooth convex constrained minimization problems with nice convergence results. *Math. Program.*, 142(1-2, Ser. A):349–369, 2013.

[3] A. Brøndsted and R. T. Rockafellar. On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.*, 16:605–611, 1965.

[4] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, 5(2):159–180, 1997.

[5] R. S. Burachik and B. F. Svaiter. ϵ-enlargements of maximal monotone operators in Banach spaces. *Set-Valued Anal.*, 7(2):117–132, 1999.

[6] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.

[7] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems, Volume I.* Springer-Verlag, New York, 2003.

[8] D. Fernandez and M. V. Solodov. Stabilized sequential quadratic programming: A survey. Manuscript, IMPA, Rio de Janeiro, RJ 22460-320, Brazil, Sept 2013.

[9] M. Fukushima, Z.-Q. Luo, and P. Tseng. A sequential quadratically constrained quadratic programming method for differentiable convex minimization. *SIAM J. Optim.*, 13(4):1098–1119 (electronic), 2003.

[10] W. W. Hager. Stabilized sequential quadratic programming. *Comput. Optim. Appl.*, 12(1-3):253–273, 1999. Computational optimization—a tribute to Olvi Mangasarian, Part I.

[11] G. M. Korpelevič. An extragradient method for finding saddle points and for other problems. *Èkonom. i Mat. Metody*, 12(4):747–756, 1976.

[12] R. D. C. Monteiro and B. F. Svaiter. Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle point and convex optimization problems. *SIAM Journal on Optimization*, 21:1688–1720, 2010.

[13] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20:2755–2787, 2010.

[14] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM J. Optim.*, 22(3):914–935, 2012.

[15] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.

[16] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.

[17] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[18] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

[19] M. V. Solodov. On the sequential quadratically constrained quadratic programming methods. *Math. Oper. Res.*, 29(1):64–79, 2004.

[20] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[21] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[22] M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.*, 25(2):214–230, 2000.

[23] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.

[24] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446 (electronic), 2000.

[25] S. J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.*, 11(3):253–275, 1998.