SIAM J. OPTIM.
Vol. 25, No. 4, pp. 1965–1996

A HYBRID PROXIMAL EXTRAGRADIENT SELF-CONCORDANT PRIMAL BARRIER METHOD FOR MONOTONE VARIATIONAL INEQUALITIES*

RENATO D. C. MONTEIRO†, MAURICIO R. SICRE‡, AND B. F. SVAITER§

Abstract. This paper presents a hybrid proximal extragradient (HPE) self-concordant primal barrier method for solving a monotone variational inequality over a closed convex set endowed with a self-concordant barrier and with an underlying map that has Lipschitz continuous derivative. In contrast to the iteration of a previous method developed by the first and third authors that has to compute an approximate solution of a linearized variational inequality, the one of the present method solves a simpler Newton system of linear equations. The method performs two types of iterations, namely, those that follow ever changing interior paths and those that correspond to large-step HPE iterations. Due to its first-order nature, the present method is shown to have a better iteration-complexity than its zeroth order counterparts such as Korpelevich's algorithm and Tseng's modified forward-backward splitting method, although its work per iteration is larger than the one for the latter methods.

Key words. self-concordant barriers, hybrid proximal extragradient, interior-point methods, monotone variational inequality, complexity, Newton method

AMS subject classifications. 90C60, 90C30, 90C51, 47H05, 47J20, 65K10, 65K05, 49M15

DOI. 10.1137/130931862

1. Introduction.

Throughout this paper, we denote the set of real numbers by $\mathbb{R}$ and the set of nonnegative and positive real numbers by $\mathbb{R}_+$ and $\mathbb{R}_{++}$, respectively. We use $\mathbb{R}^n$ to denote the set of real n-dimensional column vectors, and $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ to denote the subsets of $\mathbb{R}^n$ consisting of the component-wise nonnegative and positive vectors, respectively. Also, $\mathbf{E}$ denotes a finite-dimensional real inner product space with inner product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively.

Some earlier related works dealing with iteration-complexity analysis of methods for variational inequality (VI) problems are as follows. Nemirovski [10] studied the complexity of Korpelevich's extragradient method under the assumption that the feasible set is bounded and an upper bound on its diameter is known. Nesterov [13] proposed a dual extrapolation algorithm for solving VI problems whose termination criterion depends on the guess of a ball centered at the initial iterate and presumably containing a solution.

A broad class of optimization, saddle-point, equilibrium, and VI problems can be posed as the *monotone inclusion* (MI) problem, namely, finding x such that $0 \in T(x)$, where T is a maximal monotone point-to-set operator. The proximal point method (PPM) proposed by Martinet [4] and further generalized by Rockafellar [18, 19] is a classical iterative method for solving the MI problem. It generates a sequence $\{x_k\}$

*Received by the editors August 5, 2013; accepted for publication (in revised form) March 6, 2015; published electronically October 6, 2015.
http://www.siam.org/journals/siopt/25-4/93186.html

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 (monteiro@isye.gatech.edu). The work of this author was partially supported by NSF grants CCF-0808863, CMMI-0900094, and CMMI-1300221 and ONR grant ONR N00014-11-1-0062.

‡Instituto de Matemática, Universidade Federal da Bahia, BA, Brazil 40170-110 (msicre@ufba.br). The work of this author was partially supported by CNPq and FAPESB grant 022/2009-PPP.

§IMPA, Instituto de Matematica Pura e Aplicada, Rio de Janeiro, RJ, Brazil 22460-320 (benar@impa.br). The work of this author was partially supported by FAPERJ grants E-26/152.512/2006 and E-26/102.821/2008 and CNPq grants 303583/2008-8 and 480101/2008-6.

according to

$$x_k = (\lambda_k T + I)^{-1}(x_{k-1}),$$

where $\{\lambda_k\}$ is a sequence of positive proximal stepsizes. It has been used as a framework for the design and analysis of several implementable algorithms. The classical inexact version of the PPM allows for the presence of a sequence of summable errors in the above iteration according to

$$\|x_k - (\lambda_k T + I)^{-1}(x_{k-1})\| \leq e_k, \qquad \sum_{k=1}^{\infty} e_k < \infty.$$

Convergence results under the above error condition have been established in [18] and have been used in the convergence analysis of other methods that can be recast in the above framework [19].

New inexact versions of the PPM with relative error tolerance were proposed by Solodov and Svaiter [21, 22, 23, 24]. Iteration-complexity results for one of these inexact versions introduced in [21], namely, the hybrid proximal extragradient (HPE) method, were established in [5]. Application of this framework to the iteration-complexity analysis of several zeroth-order (or, in the context of optimization, first-order) methods for solving monotone VI, MI, and saddle-point problems were discussed in [5] and in the subsequent papers [6, 8].

The HPE framework was also used to study the iteration-complexities of first-order (or, in the context of optimization, second-order) methods for solving either a monotone nonlinear equation (see section 7 of [5]) and, more generally, a monotone VI (see [7]). It is well known that a monotone VI determined by a monotone operator F and a closed convex set X is equivalent to the MI problem

$$0 \in T(x) = (F + N_X)(x), \tag{1.1}$$

where

$$N_X(x) := \begin{cases} \emptyset, & x \notin X, \\ \{v \in \mathbf{E} \;:\; \langle v, y - x\rangle \leq 0, \;\forall y \in X\}, & x \in X. \end{cases} \tag{1.2}$$

The paper [7] presents a first-order inexact (Newton-like) version of the PPM which requires at each iteration the approximate solution of a first-order approximation (obtained by linearizing F) of the current proximal point inclusion and uses it to perform an extragradient step as prescribed by the HPE method. Pointwise and ergodic iteration-complexity results are derived for the aforementioned first-order method using general results obtained also in [7] for a large-step variant of the HPE method.

The present paper deals with an inexact proximal point self-concordant (SC) barrier method for solving (1.1) in which each iteration can be viewed as performing an approximate proximal point iteration to the system of nonlinear equations $0 = F(x) + \mu^{-1}\nabla h(x)$, where $\mu > 0$ is a dynamic parameter (converging to ∞) and h is a self-concordant barrier for X. The corresponding proximal equation

$$0 = \lambda[F(x) + \mu^{-1}\nabla h(x)] + x - z, \tag{1.3}$$

whose solution is denoted (in this introduction only) by $x(\mu, \lambda, z)$, then yields a system of nonlinear equations parametrized by μ, the proximal stepsize $\lambda > 0$, and the base point z. At each iteration, an approximate solution for the above proximal system is obtained by performing a Newton step and the triple of parameters (μ, λ, z) is then updated. The resulting method performs two types of iterations which depend on the way (μ, λ, z) is updated. On the path-following iterations, only the parameters μ and λ are updated and the method can be viewed as following a certain curve within the surface $\{x(\mu, \lambda, z) : \mu > 0, \lambda > 0\}$ for a fixed base point z. On the other hand, the other iterations update all three parameters simultaneously and can be viewed as large-step HPE iterations applied to the original inclusion $0 \in F(x) + N_X(x)$. We establish that the complexity of the resulting method is about the same order of magnitude as the one of the method presented in [7]. Moreover, while the method of [7] (approximately) solves a linearized VI subproblem at every iteration, the method presented in this paper solves a Newton system of linear equations with respect to (1.3).

It should be noted that prior to this work, [20] presented an inexact proximal point primal-dual interior-point method based on similar ideas. The main differences between the latter algorithm and the one presented in this paper are that (1) the algorithm of [20] deals with the special class of VIs in which $X = \mathbb{R}^n_+ \times \mathbb{R}^m$ and (2) the algorithm here is a primal one while the one in [20] uses the logarithmic barrier for the latter set X in the context of a primal-dual setting.

There have been other Newton-type methods in the context of degenerate unconstrained convex optimization problems for which complexity results have been derived. In [16], a Newton-type method for unconstrained convex programs based on subproblems with a cubic regularization term is proposed and iteration-complexity results are obtained. An accelerated version of this method is studied in [14]. Also, [9] presents an accelerated inexact proximal point method for (possibly constrained) convex optimization problems based on quadratic regularized subproblems and establishes a better iteration complexity than the one derived in [14]. It should be mentioned that these methods are specifically designed for convex optimization problems and hence do not apply to the monotone VI problems studied in this paper.

This paper is organized as follows. Section 2 contains three subsections. Subsection 2.1 reviews some basic properties of the ε-enlargement of a point-to-set monotone operator. Subsection 2.2 reviews an underelaxed HPE method for finding a zero of a maximal monotone operator and presents its corresponding convergence rates bounds. Subsection 2.3 reviews some basic properties of SC functions and barriers. Section 3 contains two subsections. Subsection 3.1 introduces the proximal interior surface $\{x(\mu, \lambda, z) : (\mu, \lambda, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}^n\}$ and gives conditions for points of this surface to approach the solution of the VI problem. Subsection 3.2 introduces a neighborhood of the point $x(\mu, \lambda, z)$ and shows that it has the quadratic convergence property with respect to a Newton step applied to (1.3). Section 4 also contains two subsections. Subsection 4.1 states the HPE-IP method and derives preliminaries results about the behavior of its two types of iterations. Subsection 4.2 estimates the iteration-complexity of the HPE-IP method. Section 5 discusses a Phase I procedure for computing the required input for the HPE-IP method and establishes its iteration-complexity. The appendix states and proves some technical results.

1.1. Notation. In addition to the notation introduced at the beginning of section 1, we will also use the following notation throughout the paper. The domain of definition of a point-to-point function F is denoted by $DomF$. The effective domain of a function $f : \mathbf{E} \to (-\infty, \infty]$ is denoted as $\operatorname{dom} f := \{x \in \mathbf{E} \, : \, f(x) < +\infty\}$. The range and null spaces of a linear operator $A : \mathbf{E} \to \mathbf{E}$ are denoted by $\operatorname{Range}(A) := \{Ah \, : \, h \in \mathbf{E}\}$ and $\mathcal{N}(A) := \{u \in \mathbf{E} \, : \, Au = 0\}$, respectively. The space of self-adjoint linear operators in $\mathbf{E}$ is denoted by $\mathcal{S}^{\mathbf{E}}$ and the cone of self-adjoint positive semidefinite linear operators in $\mathbf{E}$ by $\mathcal{S}_+^{\mathbf{E}}$, that is,

$$\mathcal{S}^{\mathbf{E}} := \{A : \mathbf{E} \to \mathbf{E} \ : \ A \text{ linear}, \ \langle Ax, x\rangle = \langle x, Ax\rangle \quad \forall x \in \mathbf{E}\},$$
$$\mathcal{S}_+^{\mathbf{E}} := \{A \in \mathcal{S}^{\mathbf{E}} \ : \ \langle Ax, x\rangle \geq 0 \quad \forall x \in \mathbf{E}\}.$$

For $A, B \in \mathcal{S}^{\mathbf{E}}$, we write $A \preceq B$ (or $B \succeq A$) whenever $B - A \in \mathcal{S}_+^{\mathbf{E}}$. The orthogonal projection of a point $x \in \mathbf{E}$ onto a closed convex set $S \subset \mathbf{E}$ is denoted by

$$P_S(x) := \operatorname{argmin}\{\|x - y\| \ : \ y \in S\}.$$

The cardinality of a finite set A is denoted by $\#A$. For $t > 0$, we let $\log^+(t) := \max\{\log(t), 0\}$. For $t \in \mathbb{R}$, $\lceil t \rceil$ stands for the smallest integer grater or equal than t.

2. Technical background.

This section contains three subsections. The first subsection reviews the basic definition and properties of the ε-enlargement of a point-to-set monotone operator. The second one reviews an underelaxed large-step HPE method studied in [20] together with its corresponding convergence rate bounds. The third subsection reviews the definitions and some properties of SC functions and barriers.

2.1. The ε-enlargement of monotone operators.

In this subsection, we give the definition of the ε-enlargement of a monotone operator and review some of its properties.

A point-to-set operator $T : \mathbf{E} \rightrightarrows \mathbf{E}$ is a relation $T \subset \mathbf{E} \times \mathbf{E}$ and

$$T(x) := \{v \in \mathbf{E} \ : \ (x, v) \in T\}.$$

Alternatively, one can consider T as a multivalued function of $\mathbf{E}$ into the family $\wp(\mathbf{E}) = 2^{(\mathbf{E})}$ of subsets of $\mathbf{E}$. Regardless of the approach, it is usual to identify T with its graph

$$\operatorname{Gr}(T) := \{(x, v) \in \mathbf{E} \times \mathbf{E} \ : \ v \in T(x)\}.$$

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

$$\langle v - \tilde{v}, x - \tilde{x}\rangle \geq 0 \quad \forall (x, v), (\tilde{x}, \tilde{v}) \in \operatorname{Gr}(T),$$

and it is *maximal monotone* if it is monotone and maximal in the family of monotone operators with respect to the partial order of inclusion, i.e., $S : E \rightrightarrows E$ monotone and $\operatorname{Gr}(T) \subset \operatorname{Gr}(S)$ imply that $S = T$.

In [1], Burachik, Iusem, and Svaiter introduced the ε-enlargement of maximal monotone operators. Here, we extend this concept to a generic point-to-set operator in $\mathbf{E}$. Given $T : \mathbf{E} \rightrightarrows \mathbf{E}$ and a scalar ε, define $T^\varepsilon : \mathbf{E} \rightrightarrows \mathbf{E}$ as

$$T^\varepsilon(x) := \{v \in \mathbf{E} \ : \ \langle x - \tilde{x}, v - \tilde{v}\rangle \geq -\varepsilon \quad \forall \tilde{x} \in \mathbf{E}, \forall \tilde{v} \in T(\tilde{x})\} \ \ \forall x \in \mathbf{E}. \tag{2.1}$$

We now state a few useful properties of the operator T^ε that will be needed in our presentation.

PROPOSITION 2.1. *Let $T, T' : \mathbf{E} \rightrightarrows \mathbf{E}$. Then,*

(a) *if $\varepsilon_1 \leq \varepsilon_2$, then $T^{\varepsilon_1}(x) \subset T^{\varepsilon_2}(x)$ for every $x \in \mathbf{E}$;*

(b) *$T^\varepsilon(x) + (T')^{\varepsilon'}(x) \subset (T + T')^{\varepsilon + \varepsilon'}(x)$ for every $x \in \mathbf{E}$ and $\varepsilon, \varepsilon' \in \mathbb{R}_+$;*

(c) *T is monotone if and only if $T \subset T^0$;*

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

(e) *if T is maximal monotone, $\{(x_k, v_k, \varepsilon_k)\} \subset \mathbf{E} \times \mathbf{E} \times \mathbb{R}_+$ converges to $(\bar{x}, \bar{v}, \bar{\varepsilon})$, and $v_k \in T^{\varepsilon_k}(x_k)$ for every k, then $\bar{v} \in T^{\bar{\varepsilon}}(\bar{x})$.*

Proof. Statements (a), (b), (c), and (d) follow directly from Definition 2.1 and the definition of (maximal) monotonicity. For a proof of statement (e), see [2]. ∎

We now make two remarks about Proposition 2.1. If T is a monotone operator and $\varepsilon \geq 0$, it follows from (a) and (d) that $T(x) \subset T^{\varepsilon}(x)$ for every $x \in \mathbf{E}$ and hence that T^{ε} is really an enlargement of T. Moreover, if T is maximal monotone, then (e) says that T and T^{ε} coincide when $\varepsilon = 0$.

Finally, if $T = N_X$, where N_X is the normal cone operator defined in (1.2), then its ε-enlargement $(N_X)^{\varepsilon}$ is simply denoted by N_X^{ε}.

2.2. The underelaxed large-step HPE method. This subsection reviews an underelaxed version of the large-step HPE method presented in [20] and its corresponding convergence rate bounds.

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

$$0 \in T(x). \tag{2.2}$$

The underelaxed large-step HPE method for solving (2.2) is as follows:

(0) Let $z_0 \in \mathbf{E}$, $c > 0$, $\xi \in (0, 1]$, and $\sigma \in [0, 1)$ be given and set $k = 1$;

(1) if $0 \in T(z_{k-1})$, then stop; else, compute stepsize λ_k and $(x_k, v_k, \varepsilon_k) \in \mathbf{E} \times \mathbf{E} \times \mathbb{R}_+$ such that

$$v_k \in T^{\varepsilon_k}(x_k), \qquad \|\lambda_k v_k + x_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|x_k - z_{k-1}\|^2 \tag{2.3}$$

and

$$\lambda_k \|x_k - z_{k-1}\| \geq c > 0; \tag{2.4}$$

(2) choose a relaxation parameter $\xi_k \in [\xi, 1]$, define $z_k = z_{k-1} - \xi_k \lambda_k v_k$, set $k \leftarrow k + 1$, and go to step 1.

end

We now make a few remarks about the underelaxed large-step HPE method. First, the special case in which $\xi = 1$, and hence $\xi_k = 1$ for all k, corresponds to the large-step HPE method introduced in [7], which in turn is a generalization of a large-step HPE method for smooth operators presented in [5]. Second, the iteration-complexities of the HPE method and its large-step counterpart were established in [5] and [7], respectively. Third, similar to the large-step HPE method of [7], its underelaxed version stated above does not specify how to compute λ_k and $(x_k, v_k, \varepsilon_k)$ satisfying (2.3) and (2.4). The particular choice of λ_k and the algorithm used to compute $(x_k, v_k, \varepsilon_k)$ will depend on the particular instance of the method and the properties of the operator T. Fourth, instances of the underelaxed HPE method are assumed to be able to (either implicitly or explicitly) compute the above quantities (and in particular the two residuals v_k and ε_k which measures the accuracy of x_k as as approximate solution of (2.2)) a posteriori, i.e., using information gathered up to the current iteration. Hence, it is assumed that the sequence of tolerances $\{\varepsilon_k\}$ is computed as the method progresses instead of being specified by the user a priori (e.g., [18] assumes that $\{\varepsilon_k\}$ is a summable sequence given a priori).

The following result presents global pointwise and ergodic convergence rates for the underelaxed large-step HPE method.

PROPOSITION 2.2. *For every $k \geq 1$, define*

$$\Lambda_k := \sum_{i=1}^{k} \xi_i \lambda_i, \qquad \bar{v}_k := \sum_{i=1}^{k} \frac{\xi_i \lambda_i}{\Lambda_k} v_i,$$
$$\bar{x}_k := \sum_{i=1}^{k} \frac{\xi_i \lambda_i}{\Lambda_k} x_i, \quad \bar{\varepsilon}_k := \sum_{i=1}^{k} \frac{\xi_i \lambda_i}{\Lambda_k} [\varepsilon_i + \langle x_i - \bar{x}_k, v_i \rangle].$$

If $T^{-1}(0)$ is nonempty and d_0 denotes the distance of z_0 to $T^{-1}(0)$, then the following statements hold for every $k \geq 1$:

(a) *There exists $i_0 \leq k$ such that*

$$\|v_{i_0}\| \leq \frac{d_0^2}{c\xi(1-\sigma)k}, \qquad \varepsilon_{i_0} \leq \frac{\sigma^2 d_0^3}{2c\xi^{3/2}(1-\sigma^2)^{3/2}k^{3/2}}.$$

(b) $\bar{v}_k \in T^{\bar{\varepsilon}_k}(\bar{x}_k)$,

$$\|\bar{v}_k\| \leq \frac{2d_0^2}{c\xi^{3/2}(1-\sigma^2)^{1/2}k^{3/2}}, \qquad \bar{\varepsilon}_k \leq \frac{2d_0^3}{c\xi^{3/2}(1-\sigma)^2(1-\sigma^2)^{1/2}k^{3/2}}.$$

Proof. For a proof of this result, see Proposition 3.4 of [20]. ∎

2.3. Basic properties of self-concordant functions and barriers. In this subsection, we review some basic properties of SC functions and barriers which will be useful in our presentation. A detailed treatment of this topic can be found for example in [12] and [15].

Given $A \in \mathcal{S}_+^{\mathbf{E}}$, we consider two types of seminorms induced by A. The first one is the seminorm in $\mathbf{E}$ defined as

$$\|u\|_A := \langle Au, u \rangle^{1/2} = \|A^{1/2}u\| \quad \forall u \in \mathbf{E}. \tag{2.5}$$

The second one is defined as

$$\|u\|_A^* := \sup \{2\langle u, h \rangle - \langle Ah, h \rangle \ : \ h \in \mathbf{E}\}^{1/2} \quad \forall u \in \mathbf{E}. \tag{2.6}$$

Some basic properties of these seminorms are presented in Lemma A.1 of the appendix. We observe that $\|\cdot\|_A^*$ is a nonnegative function which may take value $+\infty$, and hence is not a norm on $\mathbf{E}$. However, the third statement of Lemma A.1 justifies the use of a norm notation for $\|\cdot\|_A^*$.

DEFINITION 2.3. *A proper closed convex function $h : \mathbf{E} \to (-\infty, \infty]$ is said to be SC if* $\operatorname{dom} h$ *is open, h is three-times continuously differentiable and*

$$h'''(x)[u, u, u] \leq 2\|u\|_{\nabla^2 h(x)}^{3/2} \quad \forall x \in \operatorname{dom} h, \forall u \in \mathbf{E}.$$

Additionally, if $\nabla^2 h(x)$ is nonsingular for every $x \in$ $\operatorname{dom} h$*, then h is said to be a nondegenerate SC-function.*

DEFINITION 2.4. *For some scalar $\eta \geq 0$, an SC-function h is said to be a η-SC barrier whenever*

$$\|\nabla h(x)\|_{\nabla^2 h(x)}^* \leq \sqrt{\eta} \quad \forall x \in \operatorname{dom} h.$$

It is easy to see that any constant function $h : \mathbf{E} \to \mathbb{R}$ is an η-SC barrier for any $\eta \geq 0$. Also, it is well known that if h is an η-SC barrier which is not constant, then $\eta \geq 1$ (see, for example, Remark 2.3.1 of [15]).

The next result gives some basic properties of an SC-function.

PROPOSITION 2.5. *If h is an SC-function and $x \in \operatorname{dom} h$ is such that $\|\nabla h(x)\|^*_{\nabla^2 h(x)} < 1$, then the following statements hold:*

(a) $\nabla h(x) \in \operatorname{Range}(\nabla^2 h(x))$ *and, for every x^+ such that $\nabla h(x) + \nabla^2 h(x)(x^+ - x) = 0$, we have $x^+ \in \operatorname{dom} h$ and*

$$\|\nabla h(x^+)\|^*_{\nabla^2 h(x^+)} \leq \left(\frac{\|\nabla h(x)\|^*_{\nabla^2 h(x)}}{1 - \|\nabla h(x)\|^*_{\nabla^2 h(x)}} \right)^2.$$

(b) *h has a minimizer x^* over $\mathbf{E}$ satisfying*

$$\|x^* - x\|_{\nabla^2 h(x)} \leq \frac{\|\nabla h(x)\|^*_{\nabla^2 h(x)}}{1 - \|\nabla h(x)\|^*_{\nabla^2 h(x)}}.$$

Proof. The first inclusion in (a) follows from the assumption that $\|\nabla h(x)\|^*_{\nabla^2 h(x)} < 1$ and Lemma A.1(b) with $A = \nabla^2 h(x)$. Moreover, the second inclusion and the inequality in (a) follow from Lemma A.1(b) and Theorems 2.1.1(ii) and 2.2.1 with $s = 1$ of [15]. Also, the first part of (b) follows from Theorem 2.2.2 of [15]. We also observe that the inequality in (b) has already been established in Theorem 4.1.13 of [12] under the assumption that h is a nondegenerate SC-function (see also [11]). We now prove this inequality for the case in which h is a degenerate SC-function. Define the function h_ν as

$$h_\nu(\tilde{x}) = h(\tilde{x}) + \frac{\nu}{2}\|\tilde{x} - x\|^2 \quad \forall \tilde{x} \in \mathbf{E}.$$

Then, $\nabla h_\nu(x) = \nabla h(x)$ and $\nabla^2 h_\nu(x) = \nabla^2 h(x) + \nu I \succeq \nabla^2 h(x)$, and hence

$$\|\nabla h_\nu(x)\|^*_{\nabla^2 h_\nu(x)} = \|\nabla h(x)\|^*_{\nabla^2 h_\nu(x)} \leq \|\nabla h(x)\|^*_{\nabla^2 h(x)} < 1,$$

where the first inequality follows from Lemma A.1(a). In view of the observation made at the beginning of this proof and the fact that h_ν is a nondegenerate SC-function, it follows that h_ν has a unique minimizer x^*_ν in $\mathbf{E}$ satisfying

$$\|x^*_\nu - x\|_{\nabla^2 h(x)} \leq \|x^*_\nu - x\|_{\nabla^2 h_\nu(x)} \leq \frac{\|\nabla h_\nu(x)\|^*_{\nabla^2 h_\nu(x)}}{1 - \|\nabla h_\nu(x)\|^*_{\nabla^2 h_\nu(x)}} \leq \frac{\|\nabla h(x)\|^*_{\nabla^2 h(x)}}{1 - \|\nabla h(x)\|^*_{\nabla^2 h(x)}},$$

where the first and third inequalities follow from Lemma A.1(a). The result now follows by noting that Lemma A.2 implies that x^*_ν converges to the minimizer x^* of h closest to x with respect to $\|\cdot\|$ as $\nu \to 0$. ∎

The following result is equivalent to Proposition 2.5(a), but it is in a form which is more suitable for our analysis in this paper. For every $x \in \operatorname{dom} h$ and $y \in \mathbf{E}$, define

$$L_{h,x}(y) := \nabla h(x) + \nabla^2 h(x)(y - x). \tag{2.7}$$

PROPOSITION 2.6. *If h is an SC-function and $x \in \operatorname{dom} h$ and $y \in \mathbf{E}$ are points such that $r := \|y - x\|_{\nabla^2 h(x)} < 1$, then $y \in \operatorname{dom} h$ and*

$$\|\nabla h(y) - L_{h,x}(y)\|^*_{\nabla^2 h(y)} \leq \left(\frac{r}{1 - r} \right)^2. \tag{2.8}$$

Proof. Let $x \in \operatorname{dom} h$ and $y \in \mathbf{E}$ be points such that $r := \|y - x\|_{\nabla^2 h(x)} < 1$ and define the function

$$\phi(\tilde{x}) = h(\tilde{x}) - \langle L_{h,x}(y), \tilde{x}\rangle \quad \forall \tilde{x} \in \mathbf{E}.$$

Then, for every $\tilde{x} \in \operatorname{dom} h$, it follows from (2.7) that

$$\nabla\phi(\tilde{x}) = \nabla h(\tilde{x}) - \nabla h(x) - \nabla^2 h(x)(y - x), \quad \nabla^2\phi(\tilde{x}) = \nabla^2 h(\tilde{x}), \tag{2.9}$$

and hence

$$\|\nabla\phi(x)\|^*_{\nabla^2\phi(x)} = \|\nabla^2 h(x)(y - x)\|^*_{\nabla^2 h(x)} = \|y - x\|_{\nabla^2 h(x)} < 1, \tag{2.10}$$

where the last equality follows from Lemma A.1(b). Since ϕ is also an SC-function and (2.9) implies that

$$\nabla\phi(x) + \nabla^2\phi(x)(y - x) = 0,$$

it follows from (2.10) and Proposition 2.5(a) that $y \in \operatorname{dom} h$ and

$$\|\nabla\phi(y)\|^*_{\nabla^2\phi(y)} \leq \left(\frac{\|\nabla\phi(x)\|^*_{\nabla^2\phi(x)}}{1 - \|\nabla\phi(x)\|^*_{\nabla^2\phi(x)}}\right)^2$$

and hence that (2.8) holds in view of (2.9) and (2.10). ◻

It is shown in Proposition 2.3.2(i.2) of [15] that $\langle\nabla h(y), u - y\rangle \leq \eta$ for every $y, u \in \operatorname{dom} h$, or, equivalently, $\nabla h(y) \in N^\eta_D(y)$ for every $y \in \operatorname{dom} h$, where $D := cl(\operatorname{dom} h)$. The proposition below extends this result to vectors close to $\nabla h(y)$ with respect to $\|\cdot\|^*_{\nabla^2 h(y)}$. Its proof closely follows that of Theorem 4.2.7 of [12] except that Proposition 2.5(b) is used in order to circumvent the restrictive assumption made in [12] that h is a nondegenerate η-SC barrier with bounded domain.

PROPOSITION 2.7. *Let h be an η-SC barrier and let $y \in \operatorname{dom} h$ and $q \in \mathbf{E}$ satisfy $\|q - \nabla h(y)\|^*_{\nabla^2 h(y)} \leq a < 1$. Then, $\langle q, u - y\rangle \leq \delta$ for every $u \in \operatorname{dom} h$, where*

$$\delta := \eta + \frac{(\sqrt{\eta} + a)a}{1 - a}.$$

As a consequence, $q \in N^\delta_D(y)$, where $D := cl\,(\operatorname{dom} h)$.

Proof. Define the function ϕ as

$$\phi(x) = -\langle q, x\rangle + h(x) \quad \forall x \in \mathbf{E}.$$

Since $\nabla\phi(x) = -q + \nabla h(x)$ and $\nabla^2\phi(x) = \nabla^2 h(x)$ for every $x \in \operatorname{dom}\phi$, the proximity assumption of the proposition, Definition 2.4, and Lemma A.1(e) imply that

$$\|\nabla\phi(y)\|^*_{\nabla^2 h(y)} \leq a, \qquad \|q\|^*_{\nabla^2 h(y)} \leq \sqrt{\eta} + a. \tag{2.11}$$

Since ϕ is an SC-function, it follows from (2.11) and Proposition 2.5(b) with $h = \phi$ and $x = y$ that function ϕ has a minimizer x^* satisfying

$$\|x^* - y\|_{\nabla^2 h(y)} \leq \frac{a}{1 - a}.$$

This conclusion together with (2.11) and Lemma A.1(f) yield

$$\langle q, x^* - y\rangle \leq \|q\|^*_{\nabla^2 h(y)}\|x^* - y\|_{\nabla^2 h(y)} \leq \frac{(\sqrt{\eta} + a)a}{1 - a}.$$

Since $\nabla\phi(x^*) = 0$, or, equivalently $q = \nabla h(x^*)$, it follows from the special case of this proposition with $a = 0$ and $y = x^*$ (see the first remark on the paragraph preceding the proposition) that

$$\langle q, u - x^* \rangle = \langle \nabla h(x^*), u - x^* \rangle \le \eta \quad \forall\, u \in \operatorname{dom} h.$$

The first part of the proposition now follows by combining the last two inequalities. The second part follows from the first one and the definition of D and $N_D^\delta(\cdot)$. ∎

3. The main problem and preliminary technical results.

This section motivates our approach toward obtaining a solution of the main problem, namely, the MI problem (1.1), and establishes some important preliminary technical results related to it. It contains two subsections. The first one introduces a family of proximal interior nonlinear equations $G_{\mu,\nu,z}(x) = 0$ parametrized by $(\mu, \nu, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$ whose unique solution $x(\mu, \nu, z)$ is shown to approach a solution of (1.1) under suitable conditions on the triple of parameters (μ, ν, z). The second subsection introduces a neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta)$ of $x(\mu, \nu, z)$ whose size depends on a specified constant $\beta \in (0, 1)$ and shows that it enjoys a useful quadratic convergence property, i.e., the full Newton step with respect to the system $G_{\mu,\nu,z}(x) = 0$ from any point in $\mathcal{N}_{\mu,\nu,z}(\beta)$ belongs to the smaller neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta^2)$.

Our problem of interest in this paper is the MI problem (1.1), where $X \subset \mathbf{E}$ and $F : DomF \subset \mathbf{E} \to \mathbf{E}$ satisfy the following conditions:

(C.1) X is closed convex and is endowed with an η-SC barrier h such that $cl\,(\operatorname{dom} h) = X$.

(C.2) F is monotone and differentiable on $X \subset DomF$.

(C.3) F' is L-Lipschitz continuous on X, i.e.,

$$\|F'(\tilde{x}) - F'(x)\| \le L\|\tilde{x} - x\| \quad \forall x, \tilde{x} \in X,$$

where the norm on the left-hand side is the operator norm.

(C.4) The solution set X^* of problem (1.1) is nonempty.

Observe that assumptions C.1 and C.2 imply that the operator $T = F + N_X$ is maximal monotone (see, for example, Proposition 12.3.6 of [3]). Also, assumption C.3 implies that

$$\|F(y) - F(x) - F'(x)(y - x)\| \le \frac{L}{2}\|y - x\|^2 \quad \forall x, y \in X. \tag{3.1}$$

3.1. Proximal interior map.

This subsection introduces a proximal interior map whose argument is a triple $(\mu, \nu, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$ and gives conditions on these parameters under which the corresponding image point $x(\mu, \nu, z)$ approaches a solution of problem (1.1).

The classical central path for (1.1) assigns to each $\mu > 0$ the solution x_μ of

$$\mu F(x) + \nabla h(x) = 0. \tag{3.2}$$

Under some regularity conditions, it can be shown that the path $\mu > 0 \mapsto x_\mu$ is well-defined and x_μ approaches the solution set of (1.1) as μ goes to ∞ (see, for example, [15]). Interior-point methods for (1.1) which follow this path have been proposed in [15] under the assumption that h satisfies C.1 and F is β-compactible with h for some $\beta \ge 0$ (see Definition 7.3.1 in [15]). It is worth noting that assumptions C.1–C.3 do not imply that F is β-compactible with h for any $\beta \ge 0$ even when F is an analytic map. Hence, it is not clear how the interior-point path-following methods of [15] can be used to solve (1.1) under assumptions C.1–C.3.

This paper pursues a different strategy based on the following two ideas: (i) a parametrized proximal term is added to ∇h, and (ii) path-following steps are combined with proximal extragradient steps. Next, we discuss idea (i). Instead of the perturbed equation (3.2), our approach is based on the regularized perturbed equation

$$G_{\mu,\nu,z}(x) = 0 \tag{3.3}$$

parametrized by $(\mu, \nu, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$, where $G_{\mu,\nu,z} : \operatorname{dom} h \subset \mathbf{E} \to \mathbf{E}$ is the map defined as

$$G_{\mu,\nu,z}(x) := \mu F(x) + \nabla h(x) + \nu(x - z) \quad \forall x \in \operatorname{dom} h. \tag{3.4}$$

As opposed to (3.2), (3.4) has a (unique) solution even when the solution set of (1.1) is empty or unbounded. Throughout this section, we refer to this solution, which we denote by $x(\mu, \nu, z)$, as the *proximal interior point* associated with (μ, ν, z). Moreover, we refer to the map $(\mu, \nu, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E} \mapsto x(\mu, \nu, z)$ as the *proximal interior map*.

The following result describes sufficient conditions on the parameter (μ, ν, z) that guarantee that $x(\mu, \nu, z)$ approaches the solution set of (1.1).

PROPOSITION 3.1. *If* $\{(\mu_k, \nu_k, z_k)\} \subset \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$ *is a sequence such that* $\{z_k\}$ *is bounded and* $\lim_{k\to\infty} \mu_k/\nu_k = \infty$ *and*

$$\nu_k \geq \bar{\nu} > 0 \quad \forall k \geq 0 \tag{3.5}$$

for some $\bar{\nu} > 0$*, then* $\{x(\mu_k, \nu_k, z_k)\}$ *is bounded and every accumulation point of* $\{x(\mu_k, \nu_k, z_k)\}$ *is a solution of* (1.1).

Proof. For any $k \geq 0$, define

$$x_k := x(\mu_k, \nu_k, z_k), \quad v_k := F(x_k) + \frac{1}{\mu_k}\nabla h(x_k), \quad \lambda_k := \frac{\mu_k}{\nu_k}, \quad \varepsilon_k := \frac{\eta}{\mu_k}, \tag{3.6}$$

where η is the SC parameter of h (see Definition 2.4). Note that the assumptions of the proposition imply that

$$\lim_{k\to\infty} \lambda_k = \infty, \quad \lim_{k\to\infty} \varepsilon_k = 0. \tag{3.7}$$

Using (3.6) and the fact that x_k satisfies (3.3) with $(\mu, \nu, z) = (\mu_k, \nu_k, z_k)$, we conclude that

$$v_k = \frac{z_k - x_k}{\lambda_k}. \tag{3.8}$$

Also, the definition of v_k in (3.6) implies

$$v_k \in (F + N_X^{\varepsilon_k})(x_k) \subset T^{\varepsilon_k}(x_k), \tag{3.9}$$

where the first inclusion is due to the last claim of Proposition 2.7 with $y = x_k$, $q = \nabla h(x_k)$, and $a = 0$ and the fact that $(1/\mu_k)N_X^{\eta}(\cdot) = N_X^{\eta/\mu_k}(\cdot)$, and the second inclusion follows from Proposition 2.1 and the definition of T in (1.1). Now, let $\hat{x}_k := (I + \lambda_k T)^{-1}(z_k)$. Using the fact that $(I + \lambda_k T)^{-1}$ is nonexpansive (see, for example, Proposition 12.3.1 of [3]), we easily see that $\{\hat{x}_k\}$ is bounded. Also, the definition of $\hat{x}_k$ implies that

$$\hat{v}_k := \frac{z_k - \hat{x}_k}{\lambda_k} \in T(\hat{x}_k). \tag{3.10}$$

The latter conclusion together with (3.9) and (2.1) then imply that

$$-\varepsilon_k \le \langle \hat{v}_k - v_k, \hat{x}_k - x_k \rangle = -\frac{\|\hat{x}_k - x_k\|^2}{\lambda_k},$$

where the equality is due to (3.8) and (3.10). Hence,

$$\|\hat{x}_k - x_k\| \le \sqrt{\lambda_k \varepsilon_k} = \sqrt{\frac{\eta}{\nu_k}} \le \sqrt{\frac{\eta}{\bar{\nu}}},$$

where the equality follows from (3.6) and the second inequality follows from (3.5). The latter two conclusions then imply that the sequence $\{x_k\}$ is bounded and hence that the first assertion of the proposition holds. In view of (3.7), (3.8), and the boundedness of $\{x_k\}$ and $\{z_k\}$, we then conclude that $\lim_{k\to\infty} v_k = 0$. This conclusion together with (3.7) and Proposition 2.1(e) then imply that any accumulation point x^* of $\{x_k\}$ satisfies $0 \in T(x^*)$. Hence, the last assertion of the proposition follows. ∎

3.2. A neighborhood of a proximal interior point. This subsection introduces a neighborhood, denoted by $\mathcal{N}_{\mu,\nu,z}(\beta)$, of $x(\mu,\nu,z)$ whose size depends on a specified constant $\beta \in (0,1)$. These neighborhoods will play an important role in the method of section 4. The main result of this subsection shows that a Newton iteration with respect to (3.3) from a point x in $\mathcal{N}_{\mu,\nu,z}(\beta)$ yields a point $x^+ \in \mathcal{N}_{\mu,\nu,z}(\beta^2)$, thereby showing that these neighborhoods possess the quadratic convergence property with respect to a Newton iteration. It also shows that the above Newton iteration can be used to generate an inexact solution of a prox inclusion corresponding to (1.1) with explicit error bounds on its residuals.

We first introduce some preliminary notation and results. For $x \in \operatorname{dom} h$ and $\nu > 0$, define the norms

$$\|u\|_{\nu,x} := \|u\|_{\nabla^2 h(x)+\nu I}, \quad \|u\|^*_{\nu,x} := \|u\|^*_{\nabla^2 h(x)+\nu I} \quad \forall u \in \mathbf{E}.$$

When $\nu = 0$, we denote the above (semi)norms simply by $\|\cdot\|_x$ and $\|\cdot\|^*_x$, respectively.

We have the following simple results.

Lemma 3.2. *For any $x \in \operatorname{dom} h$, $\nu > 0$, and $u, v \in \mathbf{E}$, we have*

$$\|u\|_{\nu,x} = \sqrt{\nu\|u\|^2 + \|u\|_x^2}, \quad \|u\|^*_{\nu,x} \le \min\left\{\|u\|^*_x, \frac{\|u\|}{\sqrt{\nu}}\right\}, \quad |\langle u, v\rangle| \le \|u\|_{\nu,x}\|v\|^*_{\nu,x}. \tag{3.11}$$

Proof. The proof of this result follows immediately from the definition of the above norms and Proposition A.1. ∎

Lemma 3.3. *For any $x \in \operatorname{dom} h$, $\nu', \nu > 0$, and $u \in E$, we have*

$$\|u\|_{\nu',x} \le \max\left\{1, \sqrt{\frac{\nu'}{\nu}}\right\}\|u\|_{\nu,x}, \quad \|u\|^*_{\nu',x} \le \max\left\{1, \sqrt{\frac{\nu}{\nu'}}\right\}\|u\|^*_{\nu,x}.$$

Proof. This result follows from the definition of the above norms and the fact that the assumption $\nu < \nu'$ implies that

$$\frac{\nu}{\nu'}(\nabla^2 h(x) + \nu' I) \preceq \nabla^2 h(x) + \nu I \preceq \nabla^2 h(x) + \nu' I. \qquad ∎$$

The following result gives a simple but crucial estimate for the size of the Newton direction of the map $G_{\mu,\nu,z}$ defined in (3.4) at x, i.e., the vector satisfying

$$G'_{\mu,\nu,z}(x)d_x + G_{\mu,\nu,z}(x) = 0, \tag{3.12}$$

in terms of the size of $G_{\mu,\nu,z}(x)$.

LEMMA 3.4. *Let $x \in \operatorname{dom} h$, $z \in \mathbf{E}$, and $\mu, \nu > 0$ be given and let d_x denote the Newton direction of $G_{\mu,\nu,z}$ at x. Then, $\|d_x\|_{\nu,x} \leq \|G_{\mu,\nu,z}(x)\|^*_{\nu,x}$.*

Proof. Using the definition of $G_{\mu,\nu,z}$, relation (3.12), the fact that $F'(x)$ is positive semidefinite, and the definition of the norm $\|\cdot\|_{\nu,x}$, we have

$$\begin{aligned}\|d_x\|^2_{\nu,x} &= \langle d_x, (\nabla^2 h(x) + \nu I)d_x\rangle \leq \langle d_x, (\mu F'(x) + \nabla^2 h(x) + \nu I)d_x\rangle \\ &= \langle d_x, G'_{\mu,\nu,z}(x)d_x\rangle = -\langle d_x, G_{\mu,\nu,z}(x)\rangle \leq \|d_x\|_{\nu,x}\|G_{\mu,\nu,z}(x)\|^*_{\nu,x},\end{aligned}$$

where the last inequality follows from (3.11). The result now trivially follows from the above relation. ∎

The following result provides some important estimates of a Newton iteration with respect to $G_{\mu,\nu,z}$.

PROPOSITION 3.5. *Let $x \in \operatorname{dom} h$, $z \in \mathbf{E}$, and $\mu, \nu > 0$ be given and consider the map $L_{h,x}(\cdot)$ defined in (2.7). Assume that the Newton direction d_x of $G_{\mu,\nu,z}$ at x satisfies $\|d_x\|_x < 1$ and define $x^+ = x + d_x$. Then, $x^+ \in \operatorname{dom} h$,*

$$\|\nabla h(x^+) - L_{h,x}(x^+)\|^*_{x^+} \leq \left(\frac{\|d_x\|_x}{1 - \|d_x\|_x}\right)^2, \tag{3.13}$$

$$\|\mu F(x^+) + L_{h,x}(x^+) + \nu(x^+ - z)\| \leq \frac{\mu L}{2}\|d_x\|^2, \tag{3.14}$$

and

$$\|G_{\mu,\nu,z}(x^+)\|^*_{\nu,x^+} \leq \max\left\{\frac{\mu L}{2\nu^{3/2}}, \frac{1}{(1 - \|d_x\|_x)^2}\right\}\|d_x\|^2_{\nu,x}. \tag{3.15}$$

Proof. Since $\|x^+ - x\|_x = \|d_x\|_x < 1$, the inclusion $x^+ \in \operatorname{dom} h$ and (3.13) follow from Proposition 2.6 with $y = x^+$. Using the definition of $L_{h,x}(\cdot)$ in (2.7) and relation (3.12), it is easy to see that

$$\mu F(x^+) + L_{h,x}(x^+) + \nu(x^+ - z) = \mu\Big(F(x^+) - \big[F(x) + F'(x)(x^+ - x)\big]\Big),$$

which combined with (3.1) and the fact that $x^+ = x + d_x$ yields (3.14). Now, using the definition of $G_{\mu,\nu,z}$, the triangle inequality for norms, the first inequality in (3.11), and relations (3.13) and (3.14), we conclude that

$$\begin{aligned}&\|G_{\mu,\nu,z}(x^+)\|^*_{\nu,x^+} \\ &\quad\leq \|\mu F(x^+) + L_{h,x}(x^+) + \nu(x^+ - z)\|^*_{\nu,x^+} + \|\nabla h(x^+) - L_{h,x}(x^+)\|^*_{\nu,x^+} \\ &\quad\leq \nu^{-1/2}\|\mu F(x^+) + L_{h,x}(x^+) + \nu(x^+ - z)\| + \|\nabla h(x^+) - L_{h,x}(x^+)\|^*_{x^+} \\ &\quad\leq \frac{\mu L}{2\nu^{3/2}}(\nu\|d_x\|^2) + \frac{1}{(1 - \|d_x\|_x)^2}\|d_x\|^2_x,\end{aligned}$$

which together with the equality in (3.11) immediately imply (3.15). ∎

The following result introduces the measure which has the desired quadratic behavior under a Newton step with respect to $G_{\mu,\nu,z}$.

PROPOSITION 3.6. *Let $\theta \in [0,1)$, $\mu,\nu > 0$, $x \in \operatorname{dom} h$, and $z \in \mathbf{E}$ be given and define*

$$\gamma_{\mu,\nu}(\theta) = \max\left\{\frac{\mu L}{2\nu^{3/2}}, \frac{1}{(1-\theta)^2}\right\}. \tag{3.16}$$

Let d_x denote the Newton direction of $G_{\mu,\nu,z}$ at x and define $x^+ = x + d_x$. Then, the condition

$$\|G_{\mu,\nu,z}(x)\|^*_{\nu,x} \le \theta$$

implies that $x^+ \in \operatorname{dom} h$ and

$$\gamma_{\mu,\nu}(\theta)\|G_{\mu,\nu,z}(x^+)\|^*_{\nu,x^+} \le \left[\gamma_{\mu,\nu}(\theta)\|G_{\mu,\nu,z}(x)\|^*_{\nu,x}\right]^2.$$

Proof. To prove the proposition, assume that $\|G_{\mu,\nu,z}(x)\|^*_{\nu,x} \le \theta$. This assumption, Lemma 3.4, and the first relation in (3.11) yield

$$\|d_x\|_x \le \|d_x\|_{\nu,x} \le \|G_{\mu,\nu,z}(x)\|^*_{\nu,x} \le \theta < 1,$$

which together with Proposition 3.5 then imply that $x^+ \in \operatorname{dom} h$ and

$$\gamma_{\mu,\nu}(\theta)\|G_{\mu,\nu,z}(x^+)\|^*_{\nu,x^+} \le \gamma_{\mu,\nu}(\theta)^2\|d_x\|^2_{\nu,x} \le \left[\gamma_{\mu,\nu}(\theta)\|G_{\mu,\nu,z}(x)\|^*_{\nu,x}\right]^2. \qquad \square$$

We now introduce the neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta)$ of the proximal interior point $x(\mu,\nu,z)$ with a scalar $\beta \ge 0$ which will play an important role in the method of section 4. Indeed, for a given scalar $\beta \ge 0$, define the β-neighborhood of $x(\mu,\nu,z)$

$$\mathcal{N}_{\mu,\nu,z}(\beta) := \{x \in \operatorname{dom} h \ : \ \gamma_{\mu,\nu}\|G_{\mu,\nu,z}(x)\|^*_{\nu,x} \le \beta\}, \tag{3.17}$$

where

$$\gamma_{\mu,\nu} = \max\left\{\frac{\mu L}{2\nu^{3/2}}, 4\right\}. \tag{3.18}$$

The main result of this section stated below shows that the Newton iterate $x^+ = x + d_x$ with $x \in \mathcal{N}_{\mu,\nu,z}(\beta)$ belongs to the smaller neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta^2)$ and generates a residual pair (v^+, ε^+) for x^+ (in the strong sense that $v^+ \in (F + N_X^{\varepsilon^+})(x^+)$) with explicit bounds on its size (see (3.22)). It also quantifies the quality of $(x^+, v^+, \varepsilon^+)$ as an approximate solution of the proximal system

$$v \in (F + N_X)(x), \quad \lambda v + x - z = 0,$$

where $\lambda = \mu/\nu$. (See the inequality in (3.21).)

PROPOSITION 3.7. *Let $\beta \in [0,1)$, $(\mu,\nu,z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$, and $x \in \mathcal{N}_{\mu,\nu,z}(\beta)$ be given and consider the map $L_{h,x}(\cdot)$ defined in (2.7). Also, let d_x denote the Newton direction of $G_{\mu,\nu,z}$ at x and define $x^+ = x + d_x$. Then, the following statements hold:*

(a) *$x^+ \in \mathcal{N}_{\mu,\nu,z}(\beta^2)$ and*

$$\|\nabla h(x^+) - L_{h,x}(x^+)\|^*_{x^+} \le \frac{\beta^2}{\gamma_{\mu,\nu}}. \tag{3.19}$$

(b) *The point* x^+ *and the triple* $(\lambda^+, v^+, \varepsilon^+)$ *defined as*

$$\lambda^+ := \frac{\mu}{\nu}, \quad v^+ := F(x^+) + \frac{1}{\mu} L_{h,x}(x^+), \quad \varepsilon^+ := \frac{1}{\mu}\left(\eta + \frac{a_x(a_x + \sqrt{\eta})}{1 - a_x}\right), \tag{3.20}$$

where

$$a_x := \left(\frac{\|d_x\|_x}{1 - \|d_x\|_x}\right)^2,$$

satisfy

$$v^+ \in (F + N_X^{\varepsilon^+})(x^+), \quad \|\lambda^+ v^+ + x^+ - z\|^2 + 2\lambda^+\varepsilon^+ \le \frac{2}{\nu}\left(\sqrt{\eta} + \frac{\beta^2}{2}\right)^2 \tag{3.21}$$

and

$$\|v^+\| \le \frac{\sqrt{\nu}}{\mu}\left[\frac{\beta^2}{4} + \sqrt{\nu}\|x^+ - z\|\right], \quad \varepsilon^+ \le \frac{1}{\mu}\left[\eta + \frac{\beta^2}{3}\left(\frac{\beta^2}{4} + \sqrt{\eta}\right)\right]. \tag{3.22}$$

Proof. Noting that the definition of $\gamma_{\mu,\nu}$ in (3.18) implies that $\gamma_{\mu,\nu} \ge 4$ and using Lemma 3.4, the first relation in (3.11), definition (3.17), and the assumptions that $\beta < 1$ and $x \in \mathcal{N}_{\mu,\nu,z}(\beta)$, we conclude that

$$\|d_x\|_x \le \|d_x\|_{\nu,x} \le \|G_{\mu,\nu,z}(x)\|^*_{\nu,x} \le \frac{\beta}{\gamma_{\mu,\nu}} \le \frac{\beta}{4} \le \frac{1}{2} \tag{3.23}$$

and hence that

$$a_x = \left(\frac{\|d_x\|_x}{1 - \|d_x\|_x}\right)^2 \le 4\|d_x\|_x^2 \le \frac{\beta^2}{\gamma_{\mu,\nu}} \le \frac{\beta^2}{4} < 1. \tag{3.24}$$

Now, relation (3.23), Proposition 3.6 with $\theta = 1/2$, definition (3.17), and the fact that $\gamma_{\mu,\nu}(1/2) = \gamma_{\mu,\nu}$ in view of (3.16) and (3.18) imply the inclusion in (a). Moreover, relations (3.23) and (3.24) together with conclusion (3.13) of Proposition 3.5 then imply that (3.19) holds. Also, the conclusion that $a_x < 1$, (3.13) and Proposition 2.7 with $y = x^+$ imply that

$$L_{h,x}(x^+) \in N_X^{\delta}(x^+), \tag{3.25}$$

where

$$\delta := \eta + \frac{a_x(a_x + \sqrt{\eta})}{1 - a_x} \le \eta + \frac{(\beta^2/4)[(\beta^2/4) + \sqrt{\eta}]}{1 - (\beta^2/4)} \le \eta + \frac{\beta^2}{3}\left[\frac{\beta^2}{4} + \sqrt{\eta}\right], \tag{3.26}$$

due to (3.24) and the fact that $\beta < 1$. The above two relations together with the definitions of v^+ and ε^+ in (3.20), relations (3.25) and (3.26), and the definition of N_X^{ε} imply the inclusion in (3.21) and the second inequality in (3.22). Moreover, using the definitions of v^+ and λ^+ in (3.20), definition (3.18), inequalities (3.14) and (3.23), and the first relation in (3.11), we have

$$\begin{aligned} \|\lambda^+ v^+ + x^+ - z\| &= \frac{1}{\nu}\|\mu F(x^+) + L_{h,x}(x^+) + \nu(x^+ - z)\| \\ &\le \frac{\mu L}{2\nu}\|d_x\|^2 \le \frac{\mu L}{2\nu^2}\|d_x\|_{\nu,x}^2 \le \frac{\gamma_{\mu,\nu}}{\sqrt{\nu}}\|d_x\|_{\nu,x}^2 \le \frac{\beta^2}{4\sqrt{\nu}}, \end{aligned} \tag{3.27}$$

which together with the definition of ε^+ in (3.20) and relation (3.26) then imply that

$$\|\lambda^+ v^+ + x^+ - z\|^2 + 2\lambda^+\varepsilon^+ \le \frac{1}{\nu}\left\{\frac{\beta^4}{16} + 2\left[\eta + \frac{\beta^2}{3}\left(\frac{\beta^2}{4} + \sqrt{\eta}\right)\right]\right\} \le \frac{2}{\nu}\left(\sqrt{\eta} + \frac{\beta^2}{2}\right)^2.$$

To end the proof of the proposition, observe that the first inequality in (3.22) follows from (3.27), the triangle inequality, and the definition of λ^+ in (3.20). □

The iteration-complexity results derived in subsection 4.2 for the method presented in subsection 4.1 are based on the following two notions of approximate solutions for (1.1). Given a pair of tolerances $(\bar\rho, \bar\varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, the first notion consists of a triple $(x, v, \epsilon) \in X \times \mathbf{E} \times \mathbb{R}_{++}$ satisfying

$$v \in (F + N_X^{\varepsilon})(x), \quad \|v\| \le \bar\rho, \quad \varepsilon \le \bar\varepsilon, \tag{3.28}$$

while the second one consists of a triple $(x, v, \epsilon) \in X \times \mathbf{E} \times \mathbb{R}_{++}$ satisfying

$$v \in (F + N_X)^{\varepsilon}(x), \quad \|v\| \le \bar\rho, \quad \varepsilon \le \bar\varepsilon. \tag{3.29}$$

Since $(F + N_X^{\varepsilon})(x) \subset (F + N_X)^{\varepsilon}(x)$ for every $x \in \mathbf{E}$, we observe that the first condition implies the second one.

Note that the inclusion in (3.21) shows that the triple $(x^+, v^+, \varepsilon^+)$ generated according to Proposition 3.7 satisfies the inclusion in (3.28). If in addition the quantity

$$\max\left\{\frac{\sqrt{\nu}}{\mu}, \frac{1}{\mu}, \frac{\nu}{\mu}\|x^+ - z\|\right\}$$

is sufficiently small, then the inequalities in (3.28) also hold in view of (3.22).

4. The HPE self-concordant primal barrier method. This section presents an underelaxed HPE SC primal barrier method, referred simply to as the HPE interior-point (HPE-IP) method, for solving the monotone variational inequality problem (1.1). The HPE-IP method is a hybrid algorithm whose iterations can be viewed as either path-following ones or large-step HPE iterations as described in subsection 2.2.

This section is divided in two subsections. The first one states the HPE-IP method and analyzes basic properties of the aforementioned iterations. The second one establishes the iteration-complexity of the HPE-IP method using the convergence rate results of subsection 2.2 and the results of the first subsection.

4.1. The method and preliminary results. This subsection states the HPE-IP method and derives preliminary results about the behavior of its two types of iterations.

We start by stating the HPE-IP method, which is then followed by several remarks whose goal is to motivate and explain the main ideas behind it:

(0) Let $\sigma, \beta \in (0, 1)$ and a quadruple $(x_0, z_0, \mu_0, \nu_0) \in \operatorname{dom} h \times \mathbf{E} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ such that

$$x_0 \in \mathcal{N}_{\mu_0,\nu_0,z_0}(\beta) \tag{4.1}$$

be given, set $k = 1$, and define

$$\gamma_0 := \max\left\{\frac{\mu_0 L}{2\nu_0^{3/2}}, 4\right\}, \quad \tau_1 := \frac{\beta(1-\beta)}{\gamma_0(\sqrt{\eta}+1)}\left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1}, \quad \tau_2 := \frac{\beta(1-\beta)}{3\gamma_0\sqrt{\eta+1}}; \tag{4.2}$$

(1) set $(x, z, \mu, \nu) = (x_{k-1}, z_{k-1}, \mu_{k-1}, \nu_{k-1})$, compute the Newton direction d_x of the map $G_{\mu,\nu,z}$ defined in (3.4) at x, and set $x^+ = x + d_x$;

(2.a) if $\sqrt{\nu}\|x^+ - z\| \le \sqrt{2}(\sqrt{\eta}+1)/\sigma$, then

$$\mu^+ = \mu(1+\tau_1)^3, \quad \nu^+ = \nu(1+\tau_1)^2, \quad z^+ = z; \tag{4.3}$$

(2.b) else

$$\mu^+ = \frac{\mu}{(1+\tau_2)^3}, \quad \nu^+ = \frac{\nu}{(1+\tau_2)^2}, \quad z^+ = z - \frac{\tau_2}{1+\tau_2}\left(\frac{\mu}{\nu}\right)v^+, \tag{4.4}$$

where v^+ is as in (3.20);

(3) set $(x_k, z_k, \mu_k, \nu_k) = (x^+, z^+, \mu^+, \nu^+)$ and $(v_k, \varepsilon_k) = (v^+, \varepsilon^+)$, where ε^+ is as in (3.20);

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

end

We now make a several remarks about the HPE-IP method. First, in view of the update rule for $\{\mu_k\}$ and $\{\nu_k\}$ in step 2.a or 2.b, it follows that $\mu_k/\nu_k^{3/2} = \mu_{k-1}/\nu_{k-1}^{3/2}$ for every $k \ge 1$ and hence that the sequence $\{(\mu_k, \nu_k)\}$ belongs to the one-dimensional curve

$$\mathcal{C}(\mu_0, \nu_0) := \left\{(\mu, \nu) \in \mathbb{R}^2_{++} \ : \ \frac{\mu}{\nu^{3/2}} = \frac{\mu_0}{\nu_0^{3/2}}\right\}. \tag{4.5}$$

Second, the curve $\mathcal{C}(\mu_0, \nu_0)$ can be parametrized by a single parameter $t > 0$ as

$$\mathcal{C}(\mu_0, \nu_0) := \left\{(\mu, \nu) = (t^3\mu_0, t^2\nu_0) : t > 0\right\}. \tag{4.6}$$

Clearly, the parameter t_k corresponding to (μ_k, ν_k) is given by

$$t_k := \frac{\mu_k}{\nu_k}\left(\frac{\mu_0}{\nu_0}\right)^{-1} \quad \forall k \ge 0. \tag{4.7}$$

Third, for every $(\mu, \nu) \in \mathcal{C}(\mu_0, \nu_0)$ and $z \in \mathbf{E}$, the neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta)$ defined in (3.17) simplifies to

$$\mathcal{N}_{\mu,\nu,z}(\beta) = \left\{x \in \operatorname{dom} h \ : \ \|G_{\mu,\nu,z}(x)\|^*_{\nu,x} \le \beta/\gamma_0\right\}$$

due to (4.5) and the definition of $\gamma_{\mu,\nu}$ and γ_0 in (3.18) and (4.2), respectively. Fourth, we have

$$t_k = \begin{cases} (1+\tau_1)t_{k-1} & \text{if step 2.a is performed;} \\ t_{k-1}/(1+\tau_2) & \text{if step 2.b is performed.} \end{cases}$$

Fifth, noting that x_0 is chosen so that (4.1) holds, Proposition 4.4(a) below shows that the condition $x_k \in \mathcal{N}_{\mu_k,\nu_k,z_k}(\beta) \subset \operatorname{dom} h$ is maintained for every $k \ge 1$ and hence that the Newton direction in step 1 of the HPE-IP method is always well-defined at every iteration. Sixth, observe that the above method assumes that an initial quadruple $(x_0, z_0, \mu_0, \nu_0) \in \operatorname{dom} h \times \mathbf{E} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ satisfying (4.1) is known. Section 5 describes a Phase I procedure, and its corresponding iteration-complexity, for finding such a quadruple.

Before starting the analysis of the HPE-IP method, we first give two additional remarks to motivate its two types of iteration depending on which of the steps 2.a or step 2.b is performed. First, consecutive iterations $j \le \cdots \le l$ in which step 2.a occurs can be viewed as following the path defined as

$$\{w(t) = w(t;\mu_0,\nu_0,z) := x(t^3\mu_0, t^2\nu_0, z) \ : \ t>0\} \tag{4.8}$$

from $t = t_{j-1}$ to $t = t_l$, where $z = z_{j-1}$. Indeed, first note that (4.6) implies that $\mathcal{N}_{\mu_k,\nu_k,z}(\beta)$ is a β-neighborhood of the point $w(t_k)$ for every $k = j-1,\ldots,l$. Proposition 3.7(a) shows that $x_{k-1} \in \mathcal{N}_{\mu_{k-1},\nu_{k-1},z}(\beta)$ implies that the next iterate x_k belongs to the smaller neighborhood $\mathcal{N}_{\mu_{k-1},\nu_{k-1},z}(\beta^2)$ of $w(t_{k-1})$. Moreover, Proposition 4.2 below then shows that x_k lies in the larger neighborhood $\mathcal{N}_{\mu_k,\nu_k,z}(\beta)$ of $w(t_k)$ as long as μ_k and ν_k are computed according to (4.3) (or, equivalently, $t_k = (1+\tau_1)t_{k-1}$ under the parametrization (4.7)). Hence, given that x_{k-1} is close to $x(t_{k-1})$ in the sense that $x_{k-1} \in \mathcal{N}_{\mu_{k-1},\nu_{k-1},z}(\beta)$, the Newton iterate x_k with respect to the map $G_{\mu_{k-1},\nu_{k-1},z}$ from x_{k-1} is close to $x(t_k)$ as long as $t_k = (1+\tau_1)t_{k-1}$. Thus, the path (4.8) is closely followed by the above iterations in a manner that resembles other well-known path-following methods (see, for example, [15, 17]). In view of the above discussion, we refer to iterations in which step 2.a occur as path-following iterations.

Second, Proposition 4.3 below shows that for the iterations in which step 2.b occurs (i) the computation of $(z_k,x_k) = (z^+,x^+)$ corresponds to performing an underelaxed large-step HPE iteration to the operator $F + N_X$ with stepsize $\lambda_k = \mu_k/\nu_k$ and underelaxation factor $\xi_k = \tau_2/(1+\tau_2)$ and (ii) the update of the parameters (μ,ν,z) to (μ^+,ν^+,z^+) keeps x^+ in the β-neighborhood of the updated triple (μ^+,ν^+,z^+). Note that these iterations are updating the path (i.e., the variable z) and are followed by a sequence (if any) of consecutive path-following iterations. We refer to iterations in which step 2.b occurs as large-step HPE iterations.

The following result studies how the function $\|G_{\mu,\nu,z}(x)\|^*_{\nu,x}$ changes in terms of the scalars μ and ν (and a possible update of z).

Proposition 4.1. *Let $x \in \operatorname{dom} h$, $z,p \in \mathbf{E}$, and scalars $\alpha \ge 0$ and $\mu,\mu^+,\nu,\nu^+ > 0$ be given and define*

$$z^+ = z - \alpha\left(\frac{\mu}{\nu}\right)\left(F(x) + \mu^{-1}p\right). \tag{4.9}$$

Then, we have

$$\|G_{\mu^+,\nu^+,z^+}(x)\|^*_{\nu^+,x} \le \left(\frac{\mu^+}{\mu} + \alpha\frac{\nu^+}{\nu}\right)\max\left\{1,\sqrt{\frac{\nu}{\nu^+}}\right\}\|G_{\mu,\nu,z}(x)\|^*_{\nu,x} + \left|\frac{\mu^+}{\mu} - 1\right|\sqrt{\eta}$$
$$+ \alpha\frac{\nu^+}{\nu}\|p - \nabla h(x)\|^*_x + \sqrt{\frac{\nu}{\nu^+}}\left|(1-\alpha)\frac{\nu^+}{\nu} - \frac{\mu^+}{\mu}\right|\sqrt{\nu}\|x - z\|.$$

Proof. Let $x \in \operatorname{dom} h$ be given. By (3.4) and (4.9), we have

$$\begin{aligned}
G_{\mu^+,\nu^+,z^+}(x) &= \mu^+F(x) + \nabla h(x) + \nu^+(x - z^+)\\
&= \mu^+F(x) + \nabla h(x) + \nu^+\left[x - z + \alpha\left(\frac{\mu}{\nu}\right)\left(F(x) + \mu^{-1}p\right)\right]\\
&= \left(\frac{\mu^+}{\mu} + \alpha\frac{\nu^+}{\nu}\right)\mu F(x) + \nabla h(x) + \nu^+(x-z) + \frac{\nu^+}{\nu}\alpha p\\
&= \left(\frac{\mu^+}{\mu} + \alpha\frac{\nu^+}{\nu}\right)G_{\mu,\nu,z}(x) + \left(1 - \frac{\mu^+}{\mu}\right)\nabla h(x) + \frac{\nu^+}{\nu}\alpha(p - \nabla h(x))\\
&\quad + \left[(1-\alpha)\frac{\nu^+}{\nu} - \left(\frac{\mu^+}{\mu}\right)\right]\nu(x - z).
\end{aligned}$$

Using the triangle inequality and Lemma 3.2, we then conclude that

$$\begin{aligned}\|G_{\mu^+,\nu^+,z^+}(x)\|^*_{\nu^+,x} \le {} & \left(\frac{\mu^+}{\mu} + \alpha\frac{\nu^+}{\nu}\right)\|G_{\mu,\nu,z}(x)\|^*_{\nu^+,x} \\ & + \left|\frac{\mu^+}{\mu} - 1\right| \|\nabla h(x)\|^*_x + \alpha\frac{\nu^+}{\nu}\|p - \nabla h(x)\|^*_x \\ & + \left|(1-\alpha)\frac{\nu^+}{\nu} - \frac{\mu^+}{\mu}\right| \frac{\nu}{\sqrt{\nu^+}}\|x - z\|,\end{aligned}$$

which together with Lemma 3.3, Definition 2.4, and the definition of the seminorm $\|\cdot\|^*_x$ preceding Lemma 3.2 imply the conclusion of the proposition. ∎

The following result analyzes an iteration of the HPE-IP method in which a path-following step is performed. (See the second-to-last remark preceding Proposition 4.1.)

PROPOSITION 4.2. *Let $\beta \in (0,1)$ and $\sigma, \mu_0, \nu_0 > 0$ be given and assume that $\mu, \nu > 0$ and $x, z \in \mathbf{E}$ are such that*

$$x \in \mathcal{N}_{\mu,\nu,z}(\beta), \quad \frac{\mu}{\nu^{3/2}} = \frac{\mu_0}{\nu_0^{3/2}}. \tag{4.10}$$

Let d_x be the Newton direction of $G_{\mu,\nu,z}$ at x, and define $x^+ = x + d_x$ and the triple (μ^+, ν^+, z^+) according to (4.3)*. Then, the condition*

$$\sqrt{\nu}\|x^+ - z\| \le \frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma} \tag{4.11}$$

implies the following statements:

(a) *$x^+ \in \mathcal{N}_{\mu,\nu,z}(\beta^2)$ and $x^+ \in \mathcal{N}_{\mu^+,\nu^+,z^+}(\beta)$.*

(b) *If (v^+, ε^+) is as in* (3.20)*, then $v^+ \in (F + N_X^{\varepsilon^+})(x^+)$ and*

$$\begin{aligned}\|v^+\| &\le \frac{(1+\tau_1)^2}{\nu^+}\left(\frac{\nu_0^{3/2}}{\mu_0}\right)\left[\frac{\beta^2}{4} + \frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma}\right], \\ \varepsilon^+ &\le \frac{(1+\tau_1)^3}{(\nu^+)^{3/2}}\left(\frac{\nu_0^{3/2}}{\mu_0}\right)\left[\eta + \frac{\beta^2}{3}\left(\frac{\beta^2}{4} + \sqrt{\eta}\right)\right].\end{aligned}$$

Proof. The first inclusion in (a) follows from Proposition 3.7(a). Note that the definition of τ_1 in (4.2) and the fact that $\gamma_0 \ge 4$ and $\eta \ge 1$ imply that

$$\tau_1 = \left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1}\frac{\beta(1-\beta)}{\gamma_0(\sqrt{\eta}+1)} \le \left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1}\frac{\beta(1-\beta)}{8} \le \frac{1}{32}\left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1}. \tag{4.12}$$

Moreover, the definition of $\gamma_{\mu,\nu}$ in (3.18), the second condition in (4.10), and the first two identities in (4.3) imply that $\gamma_{\mu^+,\nu^+} = \gamma_{\mu,\nu} = \gamma_0$. These two observations, Proposition 4.1 with $x = x^+$, $\alpha = 0$, and $p = 0$, relations (4.3) and (3.17), and the first inclusion in (a) then imply that

$$\begin{aligned}
&\gamma_0 \|G_{\mu^+,\nu^+,z^+}(x^+)\|^*_{\nu^+,x^+} \\
&\quad \le \frac{\mu^+}{\mu}\left[\gamma_0\|G_{\mu,\nu,z}(x^+)\|^*_{\nu,x^+}\right] + \left|\frac{\mu^+}{\mu}-1\right|\gamma_0\sqrt{\eta} \\
&\qquad + \sqrt{\frac{\nu}{\nu^+}}\left|\frac{\mu^+}{\mu}-\frac{\nu^+}{\nu}\right|\gamma_0(\sqrt{\nu}\|x^+-z\|) \\
&\quad \le (1+\tau_1)^3\beta^2 + [(1+\tau_1)^3-1]\gamma_0\sqrt{\eta} \\
&\qquad + \frac{1}{1+\tau_1}[(1+\tau_1)^3-(1+\tau_1)^2]\frac{\gamma_0\sqrt{2}(\sqrt{\eta}+1)}{\sigma} \\
&\quad = \beta^2 + [(1+\tau_1)^3-1](\beta^2+\gamma_0\sqrt{\eta}) + (1+\tau_1)\tau_1\frac{\gamma_0\sqrt{2}(\sqrt{\eta}+1)}{\sigma} \\
&\quad \le \beta^2 + \left[(1+\tau_1)^3 - 1 + (1+\tau_1)\tau_1\frac{\sqrt{2}}{\sigma}\right]\gamma_0(\sqrt{\eta}+1) \\
&\quad = \beta^2 + \tau_1\left[\tau_1^2 + \left(3+\frac{\sqrt{2}}{\sigma}\right)\tau_1 + \left(3+\frac{\sqrt{2}}{\sigma}\right)\right]\gamma_0(\sqrt{\eta}+1) \\
&\quad \le \beta^2 + \tau_1\left[1 + \left(3+\frac{\sqrt{2}}{\sigma}\right)\right]\gamma_0(\sqrt{\eta}+1) = \beta,
\end{aligned}$$

where the second-to-last inequality is due to the fact that $\beta^2 \le 1 \le \gamma_0$, the last inequality is due to (4.12), and the last equality follows from the definition of τ_1 in (4.2). The second inclusion in (a) follows from the above inequality, definition (3.17), and the fact that $\gamma_0 = \gamma_{\mu^+,\nu^+}$.

Statement (b) follows from conclusions (3.21) and (3.22) of Proposition 3.7, the update formula for ν^+ in (4.4), the identity in (4.10), and assumption (4.11). ∎

The following result analyzes an iteration of the HPE-IP method in which a large-step HPE step is performed. (See the last remark preceding Proposition 4.1.)

PROPOSITION 4.3. *Let $\beta \in (0,1)$ and $\sigma, \mu_0, \nu_0 > 0$ be given and assume that $\mu, \nu > 0$ and $x, z \in \mathbf{E}$ are such that* (4.10) *holds. Let d_x be the Newton direction of $G_{\mu,\nu,z}$ at x and define $x^+ = x + d_x$, $(\lambda^+, v^+, \varepsilon^+)$ as in* (3.20) *and (μ^+, ν^+, z^+) as in* (4.4). *Then, the following statements hold:*

(a) $x^+ \in \mathcal{N}_{\mu,\nu,z}(\beta^2)$, $x^+ \in \mathcal{N}_{\mu^+,\nu^+,z^+}(\beta)$, *and* $v^+ \in (F + N_X^{\varepsilon^+})(x^+)$.

(b) *If, in addition, the condition $\sqrt{\nu}\|x^+ - z\| > \sqrt{2}(\sqrt{\eta}+1)/\sigma$ holds, then*

$$\|\lambda^+ v^+ + x^+ - z\|^2 + 2\lambda^+\varepsilon^+ \le \sigma^2\|x^+ - z\|^2,$$
$$\lambda^+\|x^+ - z\| \ge \left(\frac{\mu_0}{\nu_0^{3/2}}\right)\frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma}.$$

Proof. The first inclusion in (a) follows from Proposition 3.7(a). Note that the definition of $\gamma_{\mu,\nu}$ in (3.18), the second condition in (4.10), and the first two identities in (4.4) imply that $\gamma_{\mu^+,\nu^+} = \gamma_{\mu,\nu} = \gamma_0$. Using this observation, Proposition 4.1 with $x = x^+$, $\alpha = \tau_2/(1+\tau_2)$, and $p = L_{h,x}(x^+)$ (see (2.7)), and relation (4.4), and noting that

$$\max\left\{1, \sqrt{\frac{\nu}{\nu^+}}\right\} = 1 + \tau_2, \quad \frac{\mu^+}{\mu} + \alpha\frac{\nu^+}{\nu} = \frac{1}{(1+\tau_2)^2}, \quad (1-\alpha)\frac{\nu^+}{\nu} - \frac{\mu^+}{\mu} = 0,$$

we conclude that

$$
\begin{aligned}
&\gamma_{\mu_+,\nu_+}\|G_{\mu^+,\nu^+,z^+}(x^+)\|^*_{\nu^+,x^+} \\
&\quad\le \frac{1}{1+\tau_2}\left[\gamma_{\mu,\nu}\|G_{\mu,\nu,z}(x^+)\|^*_{\nu,x^+}\right] + \left(1-\frac{1}{(1+\tau_2)^3}\right)\gamma_0\sqrt{\eta} \\
&\qquad + \frac{\tau_2}{(1+\tau_2)^3}\gamma_{\mu,\nu}\|\nabla h(x^+) - L_{h,x}(x^+)\|^*_{x^+} \\
&\quad\le \frac{1}{1+\tau_2}\beta^2 + \left(1-\frac{1}{(1+\tau_2)^3}\right)\gamma_0\sqrt{\eta} + \frac{\tau_2}{(1+\tau_2)^3}\beta^2 \\
&\quad\le \beta^2 + 3\tau_2\gamma_0\sqrt{\eta} \le \beta,
\end{aligned}
$$

where the second inequality follows from (a), definition (3.17), and conclusion (3.19) of Proposition 3.7, the third inequality follows from relations $1/(1+t)+t/(1+t)^3 \le 1$ and $1 - 1/(1+t)^3 \le 3t$ for every $t > 0$, and the last inequality follows from the definition of τ_2 in (4.2). The second inclusion in (a) now immediately follows from the above relation and definition (3.17). Finally, the third inclusion in (a) is exactly the inclusion in (3.21).

In order to prove (b), assume now that $\sqrt{\nu}\|x^+ - z\| > \sqrt{2}(\sqrt{\eta}+1)/\sigma$. Using the inequality in (3.21) and the latter assumption, we conclude that

$$
\sigma^2\|x^+ - z\|^2 > \frac{2(\sqrt{\eta}+1)^2}{\nu} \ge \frac{2}{\nu}\left(\sqrt{\eta} + \frac{\beta^2}{2}\right)^2 \ge \|\lambda^+ v^+ + x^+ - z\|^2 + 2\lambda^+\varepsilon^+.
$$

Also, the definition of λ in (3.20), the above assumption, and (4.10) imply that

$$
\lambda^+\|x^+ - z\| = \left(\frac{\mu}{\nu^{3/2}}\right)\sqrt{\nu}\|x^+ - z\| > \left(\frac{\mu}{\nu^{3/2}}\right)\frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma} = \left(\frac{\mu_0}{\nu_0^{3/2}}\right)\frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma}.
$$

We have thus shown (b). ∎

The following result follows immediately from Propositions 4.2 and 4.3.

PROPOSITION 4.4. *The HPE-IP method is well-defined and the following statements hold for every $k \ge 0$:*

(a) $x_k \in \mathcal{N}_{\mu_k,\nu_k,z_k}(\beta)$ *and* $x_{k+1} \in \mathcal{N}_{\mu_k,\nu_k,z_k}(\beta^2)$.

(b) $v_k \in (F + N_X^{\varepsilon_k})(x_k)$.

(c) *If step* 2.a *occurs at iteration k, then* $z_k = z_{k-1}$,

$$
\|v_k\| \le \frac{(1+\tau_1)^2}{\nu_k}\left(\frac{\nu_0^{3/2}}{\mu_0}\right)\left[\frac{\beta^2}{4} + \frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma}\right],
$$

$$
\varepsilon_k \le \frac{(1+\tau_1)^3}{\nu_k^{3/2}}\left(\frac{\nu_0^{3/2}}{\mu_0}\right)\left[\eta + \frac{\beta^2}{3}\left(\frac{\beta^2}{4} + \sqrt{\eta}\right)\right].
$$

(d) *If step* 2.b *occurs at iteration k, then*

$$
\|\lambda_k v_k + x_k - z_{k-1}\|^2 + 2\lambda_k\varepsilon_k \le \sigma^2\|x_k - z_{k-1}\|^2,
$$

$$
\lambda_k\|x_k - z_{k-1}\| \ge \frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma}\left(\frac{\mu_0}{\nu_0^{3/2}}\right),
$$

$$
z_k = z_{k-1} - \frac{\tau_2}{1+\tau_2}\lambda_k v_k,
$$

where

$$\lambda_k := \frac{\mu_{k-1}}{\nu_{k-1}} \quad \forall k \geq 1. \tag{4.13}$$

Observe that statements (b) and (d) of Proposition 4.4 imply that iterations in which step 2.b is performed can be regarded as underelaxed large-step HPE iterations with respect to the monotone inclusion (1.1) with relaxation parameter $\tau_2/(1+\tau_2)$.

4.2. Iteration-complexity analysis of the HPE-IP method. This subsection establishes the iteration-complexity of the HPE-IP method using the convergence rate results of subsection 2.2 and Proposition 4.4.

For every $k \geq 1$, let

$$A_k := \{i \leq k \ : \ \text{step 2.a is executed at iteration } i\}, \quad a_k := \#A_k,$$
$$B_k := \{i \leq k \ : \ \text{step 2.b is executed at iteration } i\}, \quad b_k := \#B_k,$$

where the notation $\#A_k$ and $\#B_k$ stand for the number of elements of A_k and B_k, respectively.

LEMMA 4.5. *The following relations hold for every $k \geq 1$:*

$$k = a_k + b_k, \quad \mu_k = \mu_0 \frac{(1+\tau_1)^{3a_k}}{(1+\tau_2)^{3b_k}}, \quad \nu_k = \nu_0 \frac{(1+\tau_1)^{2a_k}}{(1+\tau_2)^{2b_k}}.$$

Proof. The above identities follow immediately from the above definitions of A_k, B_k, a_k, and b_k and the update formulas (4.3) and (4.4). □

The next result describes two threshold values, expressed in terms of the Lipschitz constant L, the SC parameter η, the tolerance pair $(\bar{\rho}, \bar{\varepsilon})$, and the quantities

$$d_0 := \min\{\|z_0 - x^*\| \ : \ x^* \in X^*\}, \quad \phi_0 := \frac{L\mu_0}{2\nu_0^{3/2}}, \tag{4.14}$$

which have the following properties: if the number b_k of large-step HPE iterations performed by the HPE-IP method ever becomes larger than or equal to the first (resp., second) threshold value, then the method yields a triple $(x, v, \epsilon) \in X \times \mathbf{E} \times \mathbb{R}_{++}$ satisfying (3.28) (resp., (3.29)). We observe, however, that there exists the possibility that the HPE-IP method never performs that many large-step HPE iterations, and instead computes the desired approximate solution triple due it performing a sufficiently large number of path-following iterations. The latter situation will be analyzed within the proof of Theorem 4.7.

For those k such that $B_k \neq \emptyset$, define

$$\Lambda_k := \sum_{i \in B_k} \lambda_i, \quad \bar{x}_k := \sum_{i \in B_k} \frac{\lambda_i}{\Lambda_k} x_i, \tag{4.15}$$

$$\bar{v}_k := \sum_{i \in B_k} \frac{\lambda_i}{\Lambda_k} v_i, \quad \bar{\varepsilon}_k := \sum_{i \in B_k} \frac{\lambda_i}{\Lambda_k} [\varepsilon_i + \langle x_i - \bar{x}_k, v_i - \bar{v}_k \rangle], \tag{4.16}$$

where λ_i is defined in (4.13).

LEMMA 4.6. *Let a pair $(\bar\rho,\bar\varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ be given. Then, there exist indices K_1^b and K_2^b such that*

(4.17)
$$K_1^b = \mathcal{O}\left(\frac{d_0^{4/3}\lceil\phi_0\rceil L^{2/3}(\eta+1)^{1/6}}{\phi_0^{2/3}} \max\left\{\left(\frac{Ld_0^2}{\phi_0(\sqrt{\eta}+1)}\right)^{1/3}\frac{1}{\bar\rho}, \left(\frac{d_0}{\bar\varepsilon}\right)^{2/3}\right\}+1\right)^{1},$$
$$K_2^b = \mathcal{O}\left(\frac{d_0^{4/3}\lceil\phi_0\rceil L^{2/3}(\eta+1)^{1/6}}{\phi_0^{2/3}} \max\left\{\frac{1}{\bar\rho}, \frac{d_0}{\bar\varepsilon}\right\}^{2/3}+1\right)$$

and the following statements hold:

(a) *If k_0 is an iteration index satisfying $b_{k_0} \ge K_1^b$, then there is an index $i \in B_{k_0}$ such that the triple of $(x,v,\varepsilon) = (x_i, v_i, \varepsilon_i)$ satisfies* (3.28).

(b) *If k_0 is an iteration index satisfying $b_{k_0} \ge K_2^b$, then the triple of $(x,v,\varepsilon) = (\bar x_k, \bar v_k, \bar\varepsilon_k)$ satisfies* (3.29) *for every $k \ge k_0$.*

Proof. (a) Define

(4.18)
$$K_1^b := \left\lceil \frac{d_0^2(1+\tau_2)}{(1-\sigma)\tau_2}\max\left\{\frac{1}{\bar\rho}\left(\frac{\sigma\nu_0^{3/2}}{\sqrt{2}(\sqrt{\eta}+1)\mu_0}\right), \frac{1}{\bar\varepsilon^{2/3}}\left(\frac{\sigma^3\nu_0^{3/2}}{2^{3/2}(\sqrt{\eta}+1)\mu_0}\right)^{2/3}\right\}\right\rceil$$

and observe that (4.17) holds due to the definitions of τ_2 in (4.2) and ϕ_0 in (4.14). Let k_0 be an iteration index satisfying $b_{k_0} \ge K_1^b$. In view of (b) and (d) of Proposition 4.4 and the fact that $z_k = z_{k-1}$ whenever $k \in A_{k_0}$ (and hence $k \notin B_{k_0}$), the (finite) set of iterates $\{(z_i, x_i, v_i, \epsilon_i),\ i \in B_{k_0}\}$ satisfies the rules of the underelaxed large-step HPE method described in subsection 2.2 with

$$T = F + N_X, \quad c = \frac{\sqrt{2}(\sqrt{\eta}+1)\mu_0}{\sigma\nu_0^{3/2}}, \quad \xi_k = \xi = \frac{\tau_2}{1+\tau_2} \quad \forall k \ge 1.$$

Hence, from Proposition 2.2(a) with T, c, and ξ as above and $k = b_{k_0}$, it follows that there exists an index $i \in B_{k_0}$ such that

$$\|v_i\| \le \frac{\sigma(1+\tau_2)\nu_0^{3/2}}{\sqrt{2}(1-\sigma)\tau_2(\sqrt{\eta}+1)\mu_0}\frac{d_0^2}{b_{k_0}}, \qquad \varepsilon_i \le \frac{\sigma^3(1+\tau_2)^{3/2}\nu_0^{3/2}}{2^{3/2}(1-\sigma)^{3/2}\tau_2^{3/2}(\sqrt{\eta}+1)\mu_0}\ \frac{d_0^3}{b_{k_0}^{3/2}}.$$

The above two inequalities, the assumption that $b_{k_0} \ge K_1^b$, and the definition of K_1^b easily imply that $(x,v,\varepsilon) = (x_i, v_i, \varepsilon_i)$ satisfies the two inequalities in (3.28). Moreover, Proposition 4.4(b) implies that this triple satisfies the inclusion in (3.28). Hence, (a) follows.

(b) Define

$$K_2^b := \left\lceil \frac{d_0^{4/3}2\sigma^{2/3}\nu_0(1+\tau_2)}{(1-\sigma^2)^{1/3}\mu_0^{2/3}(\sqrt{\eta}+1)^{2/3}\tau_2}\max\left\{\frac{1}{\bar\rho}, \frac{1}{\bar\varepsilon}\frac{d_0}{(1-\sigma)^2}\right\}^{2/3}\right\rceil \tag{4.19}$$

and observe that (4.18) is satisfied in view of the definition of τ_2 in (4.2) and ϕ_0 in (4.14). Let k_0 be an iteration index satisfying $b_{k_0} \ge K_2^b$. Using the definition of $\bar v_k$ and $\bar\varepsilon_k$ in (4.16), the fact that $\xi_i = \xi$ for every $i \ge 1$, and Proposition 2.2(b) with

[1]Notation $f = \mathcal{O}(t)$ means that there exists a constant $C > 0$ such that $|f(t)| \le Ct\ \forall\, t > 0$.

T, c, and ξ as above, we conclude that for any $k \geq 1$ such that $B_k \neq \emptyset$, we have $\bar{v}_k \in T^{\bar{\varepsilon}_k}(\bar{x}_k)$ and

$$\|\bar{v}_k\| \leq \frac{2^{3/2}\sigma\nu_0^{3/2}(1+\tau_2)^{3/2}}{(1-\sigma^2)^{1/2}(\sqrt{\eta}+1)\mu_0\tau_2^{3/2}}\,\frac{d_0^2}{b_k^{3/2}},$$
$$\bar{\varepsilon}_k \leq \frac{2^{3/2}\sigma\nu_0^{3/2}(1+\tau_2)^{3/2}}{(1-\sigma)^2(1-\sigma^2)^{1/2}(\sqrt{\eta}+1)\mu_0\tau_2^{3/2}}\,\frac{d_0^3}{b_k^{3/2}}.$$

Statement (b) now follows from the last observation, the fact that $k \geq k_0$ implies $b_k \geq b_{k_0}$, the assumption that $b_{k_0} \geq K_2^b$, and the definition of K_2^b. ∎

The following results present iteration-complexity bounds for the HPE-IP method to obtain approximate solutions of (1.1) satisfying either (3.28) or (3.29). For simplicity, we ignore the dependence of these bounds on the parameter σ and other universal constants and express them only in terms of L, d_0, η, the initialization parameters μ_0 and ν_0, and the tolerances $\bar{\rho}$ and $\bar{\varepsilon}$.

The first result gives the pointwise iteration-complexity of the HPE-IP method.

THEOREM 4.7. *Let a pair $(\bar{\rho},\bar{\varepsilon}) \in \mathbb{R}_{++}\times\mathbb{R}_{++}$ be given. Then, there exists an index*

$$i = \mathcal{O}\left(\frac{d_0^2\lceil\phi_0\rceil L^{2/3}(\eta+1)^{1/6}}{\phi_0^{2/3}}\max\left\{\left(\frac{L}{\phi_0(\eta+1)}\right)^{1/6}\frac{1}{\bar{\rho}},\,\frac{1}{\bar{\varepsilon}^{2/3}}\right\}+1\right) + \mathcal{O}\left(\lceil\phi_0\rceil\sqrt{\eta+1}\max\left\{\log^+\left(\frac{L(\eta+1)}{\phi_0\nu_0\bar{\rho}}\right),\log^+\left(\frac{L(\eta+1)}{\phi_0\nu_0\bar{\varepsilon}}\right)\right\}\right) \tag{4.20}$$

such that the triple of $(x,v,\varepsilon) = (x_i,v_i,\varepsilon_i)$ satisfies (3.28).

Proof. Define the constants

$$C_1 := (1+\tau_1)^2\left(\frac{\beta^2}{4}+\frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma}\right)\left(\frac{\nu_0^{3/2}}{\mu_0}\right), \tag{4.21}$$
$$C_2 := (1+\tau_1)^3\left(\eta+\frac{\beta^2}{3}\left(\frac{\beta^2}{4}+\sqrt{\eta}\right)\right)\left(\frac{\nu_0^{3/2}}{\mu_0}\right)$$

and

$$K_1 := \left\lceil K_1^b\left(1+\frac{\log(1+\tau_1)}{\log(1+\tau_2)}\right) + \frac{1}{\log(1+\tau_2)}\max\left\{\frac{1}{2}\log^+\left(\frac{C_1}{\nu_0\bar{\rho}}\right),\,\frac{1}{3}\log^+\left(\frac{C_2}{\nu_0^{3/2}\bar{\varepsilon}}\right)\right\}\right\rceil,$$

where K_1^b is defined in (4.18). Now, using the fact that $t/(1+t) \leq \log(1+t) \leq t$ for every $t > -1$ and the definitions of τ_1 and τ_2 in (4.2), we easily see that

$$\frac{\log(1+\tau_1)}{\log(1+\tau_2)} = \mathcal{O}(1), \quad \frac{1}{\log(1+\tau_2)} = \mathcal{O}(\lceil\phi_0\rceil\sqrt{\eta+1}). \tag{4.22}$$

This observation together with (4.2), (4.17), and (4.21) then imply that K_1 can be estimated according to the right-hand side of (4.20). To end the proof, it suffices to show the existence of an index $i \le K_1$ such that the triple of $(x, v, \varepsilon) = (x_i, v_i, \varepsilon_i)$ satisfies (3.28).

Indeed, to prove the latter claim, we consider the following two cases: $b_{K_1} \ge K_1^b$ and $b_{K_1} < K_1^b$. If $b_{K_1} \ge K_1^b$, then the claim follows immediately from Lemma 4.6(a) with $k_0 = K_1$. Consider now the case in which $b_{K_1} < K_1^b$. Since $K_1^b < K_1$, we have that $b_{K_1} < K_1$. This together with the first identity in Lemma 4.5 then imply that $a_{K_1} > 0$ and hence that $A_{K_1} \neq \emptyset$. Let i be the largest index in A_{K_1}. Observe that $i \le K_1$ and we clearly have

$$b_i \le b_{K_1} < K_1^b, \quad a_i = a_{K_1} = K_1 - b_{K_1} > K_1 - K_1^b,$$

where the last equality is due to the first identity in Lemma 4.5. These inequalities together with Proposition 4.4, Lemma 4.5, and the definition of K_1 can now be easily seen to imply that

$$\|v_i\| \le \frac{C_1}{\nu_i} = \frac{(1+\tau_2)^{2b_i} C_1}{\nu_0 (1+\tau_1)^{2a_i}} \le \frac{(1+\tau_2)^{2K_1^b} C_1}{\nu_0 (1+\tau_1)^{2(K_1 - K_1^b)}} \le \bar{\rho}$$

and

$$\varepsilon_i \le \frac{C_2}{\nu_i^{3/2}} = \frac{(1+\tau_2)^{3b_i} C_2}{\nu_0^{3/2} (1+\tau_1)^{3a_i}} \le \frac{(1+\tau_2)^{3K_1^b} C_2}{\nu_0^{3/2} (1+\tau_1)^{3(K_1 - K_1^b)}} \le \bar{\varepsilon}.$$

The last conclusion together with Proposition 4.4(b) then imply that $(x, v, \varepsilon) = (x_i, v_i, \varepsilon_i)$ satisfies (3.28). ∎

The next result presents the iteration-complexity of the HPE-IP method for the sequences of ergodic means defined in (4.15)-(4.16).

THEOREM 4.8. *Let a pair* $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ *be given. Then, there exists an index*

$$K_2 = \mathcal{O}\left(\frac{d_0^{4/3} \lceil \phi_0 \rceil L^{2/3} (\eta+1)^{1/6}}{\phi_0^{2/3}} \max\left\{ \frac{1}{\bar{\rho}}, \frac{d_0}{\bar{\varepsilon}} \right\}^{2/3} + 1 \right) + \mathcal{O}\left(\lceil \phi_0 \rceil \sqrt{\eta+1} \max\left\{ \log^+\left(\frac{L(\eta+1)}{\phi_0 \nu_0 \bar{\rho}} \right), \log^+\left(\frac{L(\eta+1)}{\phi_0 \nu_0 \bar{\varepsilon}} \right) \right\} \right) \tag{4.23}$$

such that at least one of the following statements hold:

(a) *There exists an index* $i \le K_2$ *such that the triple* $(x_i, v_i, \varepsilon_i)$ *satisfies* (3.28).

(b) *For every index* $k \ge K_2$, *the triple* $(\bar{x}_k, \bar{v}_k, \bar{\varepsilon}_k)$ *satisfies* (3.29).

Proof. Define

$$K_2 := \left\lceil K_2^b \left(1 + \frac{\log(1+\tau_1)}{\log(1+\tau_2)} \right) + \frac{1}{\log(1+\tau_2)} \max\left\{ \frac{1}{2} \log^+\left(\frac{C_1}{\nu_0 \bar{\rho}} \right), \frac{1}{3} \log^+\left(\frac{C_2}{\nu_0^{3/2} \bar{\varepsilon}} \right) \right\} \right\rceil,$$

where C_1 and C_2 are defined in (4.21) and K_2^b is defined in (4.19). Now, (4.2), (4.17), (4.21), and (4.22) imply that K_2 satisfies (4.23).

It remains to show that either (a) or (b) holds. Indeed, as in the proof of Theorem 4.7, we consider the following two cases: $b_{K_2} \ge K_2^b$ and $b_{K_2} < K_2^b$. If $b_{K_2} \ge K_2^b$, then (b) holds in view of Lemma 4.6(b) with $k_0 = K_2$. If, on the other hand, $b_{K_2} < K_2^b$, it can be shown using similar arguments as in the proof of Theorem 4.7 that the largest index i in $A_{K_2} \neq \emptyset$ satisfies (a). □

The complexity bounds of Theorems 4.7 and 4.8 suggest that the choice of an initial pair (μ_0, ν_0) such that $\phi_0 := L\mu_0/2\nu_0^{3/2}$ is close to 1 (e.g., $\phi_0 \in [1,4]$) is a good strategy. The next section presents a Phase I procedure which generates an initial quadruple $(x_0, z_0, \mu_0, \nu_0) \in \operatorname{dom} h \times \mathbf{E} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ for the HPE-IP method satisfying the condition that

$$\frac{\phi_0}{\min\{1, L\}} \in [1,4]\,. \tag{4.24}$$

Note that ϕ_0 is divided by the factor $\min\{1, L\}$. Its goal is to prevent the iteration-complexity of the Phase I procedure from growing as L becomes small. Finally, observe that the choice of the interval $[1,4]$ is completely arbitrary and that the procedure can be easily modify for other choices of this interval.

5. A Phase I procedure.

In this section, we discuss a Phase I procedure which, given a pair $(\tilde{\nu}_0, z_0) \in \mathbb{R}_{++} \times \mathbf{E}$, finds a triple $(\mu_0, \nu_0, x_0) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \operatorname{dom} h$ such that the quadruple (x_0, z_0, μ_0, ν_0) satisfies conditions (4.1) and (4.24), and we also establish its iteration-complexity in terms of its input $(z_0, \tilde{\nu}_0)$. As a result, we will derive the iteration-complexity of the overall method consisting of first applying the Phase I procedure and then the HPE-IP method.

We start by describing the Phase I procedure:

(0) Let $\beta \in (0,1)$ and $(\tilde{\nu}_0, z_0) \in \mathbb{R}_{++} \times \mathbf{E}$ be given, define the point $\tilde{x}_0 = \tilde{x}_0(\tilde{\nu}_0, z_0)$ as the unique solution of

$$\nabla h(\tilde{x}_0) + \tilde{\nu}_0(\tilde{x}_0 - z_0) = 0, \tag{5.1}$$

define

$$\tilde{L} := \max\{L, 1\}, \quad \tilde{\mu}_0 := \min\left\{\frac{8\tilde{\nu}_0^{3/2}}{\tilde{L}}, \frac{\beta}{4\|F(\tilde{x}_0)\|^*_{\tilde{x}_0,\tilde{\nu}_0}}\right\}, \tag{5.2}$$

$$\tilde{\phi}_0 := \frac{L\tilde{\mu}_0}{2\tilde{\nu}_0^{3/2}}, \quad \hat{t} := \frac{\beta(1-\beta)}{4\sqrt{\eta+1}}, \tag{5.3}$$

and set $k = 1$;

(1) if $\tilde{\phi}_{k-1}/\min\{L,1\} \geq 1$, then stop and output $(x_0, \mu_0, \nu_0) := (\tilde{x}_{k-1}, \tilde{\mu}_{k-1}, \tilde{\nu}_{k-1})$;

(2) else, set $(x, \mu, \nu, \phi) = (\tilde{x}_{k-1}, \tilde{\mu}_{k-1}, \tilde{\nu}_{k-1}, \tilde{\phi}_{k-1})$ and compute the Newton direction d_x of G_{μ,ν,z_0} at x and set

$$x^+ := x + d_x, \quad \mu^+ := \mu(1-\hat{t}\,), \quad \nu^+ := \nu(1-\hat{t}\,), \quad \phi^+ := \frac{\mu^+ L}{2\,(\nu^+)^{3/2}}; \tag{5.4}$$

(3) let $(\tilde{x}_k, \tilde{\mu}_k, \tilde{\nu}_k, \tilde{\phi}_k) = (x^+, \mu^+, \nu^+, \phi^+)$, set $k \leftarrow k+1$, and go to step 1.

end

We now make a few remarks about the above procedure. First, (5.2) and the first relation in (5.3) imply that $\tilde{\phi}_0/\min\{L,1\} \leq 4$ but most likely will be very small. Second, (5.4) implies that

$$\tilde{\phi}_k = \frac{\tilde{\phi}_{k-1}}{(1-\hat{t}\,)^{1/2}}. \tag{5.5}$$

Hence, after a finite number of iterations of the Phase I procedure, its stopping criterion will eventually be satisfied. The following result establishes the iteration-complexity of the Phase I procedure and shows that its output (x_0, μ_0, ν_0) satisfies $\phi_0/\min\{L,1\} \in [1,4]$, where ϕ_0 is defined in (4.14).

PROPOSITION 5.1. *The following statements hold for the Phase I procedure:*

(a) *For every iteration index k, we have $\tilde{x}_{k-1} \in \mathcal{N}_{\tilde{\mu}_{k-1},\tilde{\nu}_{k-1},z_0}(\beta)$ and $\tilde{\phi}_{k-1}/\min\{1,L\} \leq 4$.*

(b) *The procedure terminates in at most*

$$\mathcal{O}\left(\sqrt{\eta+1}\log^+\left(\frac{8\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,\tilde{x}_0}}{\max\{L,1\}}\right)+1\right) \tag{5.6}$$

iterations with a triple $(\mu_0,\nu_0,x_0) \in \mathbb{R}_{++}\times\mathbb{R}_{++}\times \operatorname{dom} h$ satisfying (4.1), (4.24)*, and the estimate*

$$\log\frac{\tilde{\nu}_0}{\nu_0} = \mathcal{O}\left(\log^+\left(\frac{\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,\tilde{x}_0}}{\max\{L,1\}}\right)+1\right). \tag{5.7}$$

Proof. We prove (a) by induction on k. We first prove that (a) holds for $k=1$. Indeed, it follows from (5.2), the first relation in (5.3), and (3.18) that

$$\tilde{\phi}_0 := \frac{L\tilde{\mu}_0}{2\tilde{\nu}_0^{3/2}} \leq \frac{4L}{\tilde{L}} \leq 4\min\{1,L\} \leq 4, \quad \gamma_{\tilde{\mu}_0,\tilde{\nu}_0} := \max\left\{\frac{L\tilde{\mu}_0}{2\tilde{\nu}_0^{3/2}}, 4\right\} = 4,$$

which together with (3.4) and (5.1) then imply that

$$\gamma_{\tilde{\mu}_0,\tilde{\nu}_0}\|G_{\tilde{\mu}_0,\tilde{\nu}_0,z_0}(\tilde{x}_0)\|^*_{\tilde{x}_0,\tilde{\nu}_0} = 4\tilde{\mu}_0\|F(\tilde{x}_0)\|^*_{\tilde{x}_0,\tilde{\nu}_0} \leq \beta$$

and hence that $\tilde{x}_0 \in \mathcal{N}_{\tilde{\mu}_0,\tilde{\nu}_0,z_0}(\beta)$ in view of (3.17). We have thus shown that (a) holds for $k=1$.

Assume now that statement (a) holds for the kth iteration and hence that $\tilde{x}_{k-1} \in \mathcal{N}_{\tilde{\mu}_{k-1},\tilde{\nu}_{k-1},z_0}(\beta)$ and $\tilde{\phi}_{k-1} \leq 4\min\{1,L\}$. We will now show that statement (a) also holds for the $(k+1)$-st iteration. Let $(x,\mu,\nu,\phi) = (\tilde{x}_{k-1},\tilde{\mu}_{k-1},\tilde{\nu}_{k-1},\tilde{\phi}_{k-1})$ and define (μ^+,ν^+,x^+,ϕ^+) as in (5.4). Observe that Proposition 3.7(a) with $z=z_0$ implies that $x^+ \in \mathcal{N}_{\mu,\nu,z_0}(\beta^2)$. Since $\phi \leq 4\min\{1,L\}$ and the procedure did not stop at the kth iteration (see step 1 of the procedure), we must have $\phi = L\mu/(2\nu^{3/2}) < \min\{L,1\} \leq 1$, and hence $\gamma_{\mu,\nu} = 4$ in view of (3.18). Using the fact that $\phi^+ = \phi/(1-\hat{t})^{1/2}$ and $\hat{t} \leq 1/4$ due to the definition of $\hat{t}$ in step 0 of the procedure, we conclude that $\phi^+ \leq 4\min\{L,1\} \leq 4$ and hence that $\gamma_{\mu^+,\nu^+} \leq 4$ in view of (3.18). Now, using Proposition 4.1 with $x=x^+$, $z=z_0$, $p=0$, and $\alpha=0$ and the definition of $\hat{t}$ in (5.3), we conclude that

$$\begin{aligned}\gamma_{\mu^+,\nu^+}\|G_{\mu^+,\nu^+,z_0}(x^+)\|^*_{\nu^+,x^+} &\leq 4\left((1-\hat{t})\|G_{\mu,\nu,z_0}(x^+)\|^*_{\nu,x^+} + \hat{t}\sqrt{\eta}\right)\\ &\leq \gamma_{\mu,\nu}\|G_{\mu,\nu,z_0}(x^+)\|^*_{\nu,x^+} + 4\hat{t}\sqrt{\eta} \leq \beta^2 + 4\hat{t}\sqrt{\eta} \ \leq \ \beta\end{aligned}$$

and hence that $x^+ \in \mathcal{N}_{\mu^+,\nu^+,z_0}(\beta)$. Hence, due to the definition of $(\tilde{x}_k,\tilde{\mu}_k,\tilde{\nu}_k)$ in step 3 of the procedure, we conclude that (a) holds for $k+1$. We have thus proved (a).

To prove (b), let K denote the last iteration of the Phase I procedure, i.e., the first iteration index for which $\tilde{\phi}_{K-1} \geq \min\{1,L\}$. Note that, since $\tilde{\phi}_{K-1}/\min\{1,L\} \leq 4$

by statement (a), it follows that the output of the procedure satisfies (4.24). We will now show that

$$K \le 2 + \left(\frac{8\sqrt{\eta+1}}{\beta(1-\beta)}\right)\log^+\left(\frac{\min\{L,1\}}{\tilde\phi_0}\right). \tag{5.8}$$

Assume without any loss of generality that $K \ge 2$. In view of the inequality in (a) with $k = K-1$ and the fact that the procedure did not stop at iteration $K-1$, we conclude that

$$\min\{L,1\} > \tilde\phi_{K-2} = \frac{\tilde\phi_0}{(1-\hat t)^{(K-2)/2}} \ge \tilde\phi_0, \tag{5.9}$$

where the equality is due to (5.5). Taking logarithms on both sides and using the inequality $\log(1-\hat t) \le -\hat t$, we then conclude that

$$K \le 2 + \frac{2}{\hat t}\log\left(\frac{\min\{L,1\}}{\tilde\phi_0}\right) \le 2 + \frac{2}{\hat t}\log^+\left(\frac{\min\{L,1\}}{\tilde\phi_0}\right), \tag{5.10}$$

which clearly implies (5.8) due to the definition of $\hat t$ in (5.3). Now, using (5.9), (5.2), and the first relation in (5.3), we easily see that

$$\tilde\phi_0 = \min\left\{4\min\{L,1\}\,,\ \frac{L\beta}{8\tilde\nu_0^{3/2}\|F(\tilde x_0)\|^*_{\tilde x_0,\tilde\nu_0}}\right\} = \frac{L\beta}{8\tilde\nu_0^{3/2}\|F(\tilde x_0)\|^*_{\tilde x_0,\tilde\nu_0}}, \tag{5.11}$$

which together with (5.8) can be easily seen to imply (5.6). To complete the proof, it remains to show that (5.7) holds. First note that the update rule and the definition of K implies that $\nu_0 = \tilde\nu_0(1-\hat t)^{K-1}$ and hence that

$$\begin{aligned}\log\frac{\tilde\nu_0}{\nu_0} &= (K-1)\log\left(\frac{1}{1-\hat t}\right) \le (K-1)\frac{\hat t}{1-\hat t}\\ &\le \frac{4(K-1)\hat t}{3} \le \frac{4\hat t}{3} + \frac{8}{3}\log^+\left(\frac{\min\{L,1\}}{\tilde\phi_0}\right),\end{aligned}$$

where the last two inequalities follow from the fact that $\hat t \le 1/4$ and (5.10). Relation (5.7) now follows from the last conclusion, relation (5.11), and the definition of $\hat t$ in (5.3). □

The following result gives the overall complexity of the combined method in which the Phase I procedure is started from an input pair $(\tilde\nu_0, z_0) \in \mathbb{R}_{++} \times \mathbf{E}$ and is followed by the HPE-IP method started from (x_0, z_0, μ_0, ν_0), where (x_0, μ_0, ν_0) is the output of the Phase I procedure. Recall from (4.14) that d_0 denotes the distance of z_0 to the solution set of problem (1.1).

THEOREM 5.2. *Let a pair $(\bar\rho, \bar\varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ be given. Consider the combined method in which the Phase* I *procedure is started from an input pair $(\tilde\nu_0, z_0) \in \mathbb{R}_{++} \times \mathbf{E}$ and is followed by the HPE-IP method started from (x_0, z_0, μ_0, ν_0), where (x_0, μ_0, ν_0) is the output of the Phase* I *procedure. Then, the method computes the following:*

(i) *a triple* (x, v, ε) *satisfying* (3.28) *in at most*

(5.12)
$$\mathcal{O}\left(\sqrt{\eta+1}\log^+\left(\frac{8\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,\tilde{x}_0}}{\max\{L,1\}}\right)+\sqrt{\eta+1}\right)$$
$$+\ \mathcal{O}\left(d_0^2\max\{L,1\}^{2/3}(\eta+1)^{1/6}\max\left\{\left(\frac{\max\{L,1\}}{\sqrt{\eta+1}}\right)^{1/3}\frac{1}{\bar{\rho}},\ \frac{1}{\bar{\varepsilon}^{2/3}}\right\}+1\right)$$
$$+\ \mathcal{O}\left(\sqrt{\eta+1}\max\left\{\log^+\left(\frac{\max\{L,1\}(\eta+1)}{\tilde{\nu}_0\bar{\rho}}\right),\log^+\left(\frac{\max\{L,1\}(\eta+1)}{\tilde{\nu}_0\bar{\varepsilon}}\right)\right\}\right),$$

iterations, where $\tilde{x}_0$ *is the point determined by* (5.1)*;*

(ii) *a triple* (x, v, ε) *satisfying either* (3.28) *or* (3.29) *in at most*

(5.13)
$$\mathcal{O}\left(\sqrt{\eta+1}\log^+\left(\frac{8\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,\tilde{x}_0}}{\max\{L,1\}}\right)+\sqrt{\eta+1}\right)$$
$$+\ \mathcal{O}\left(d_0^{4/3}\max\{L,1\}^{2/3}(\eta+1)^{1/6}\max\left\{\frac{1}{\bar{\rho}},\ \frac{d_0}{\bar{\varepsilon}}\right\}^{2/3}+1\right)$$
$$+\ \mathcal{O}\left(\sqrt{\eta+1}\max\left\{\log^+\left(\frac{\max\{L,1\}(\eta+1)}{\tilde{\nu}_0\bar{\rho}}\right),\log^+\left(\frac{\max\{L,1\}(\eta+1)}{\tilde{\nu}_0\bar{\varepsilon}}\right)\right\}\right)$$

iterations.

Proof. First we prove (i). It is shown in Proposition 5.1 and Theorem 4.7 that the number of iterations performed by the Phase I procedure is bounded by (5.6) and that the number of iterations necessary for the HPE-IP method to compute a triple (x, v, ε) satisfying (3.28) is bounded by (4.20). The estimate (5.12) now follows from these two observations, relation (4.24), and estimate (5.7). (Note that the latter relation is needed due to the fact that (4.20) is expressed in terms of ν_0 instead of $\tilde{\nu}_0$.)

Using a similar argument with Theorem 4.8 replacing Theorem 4.7, we conclude that (ii) also holds. ☐

We now discuss the complexity bounds of Theorem 5.2 in light of the ones obtained in [5, 7]. For the sake of brevity, we focus our discussion on the ergodic complexity bounds. Recall that [5, 7] present first-order inexact (Newton-like) versions of the PPM which require at each iteration the approximate solution of a first-order approximation (obtained by linearizing F) of the current proximal point equation/inclusion and use it to perform an extragradient step as prescribed by the large-step HPE method of section 2.2. Moreover, the ergodic complexity derived in [7] is

$$\left\{\log\log\left[Ld_0+(L\bar{\rho})^{-1}+\bar{\varepsilon}^{-1}\right]\right\}\mathcal{C},$$

where

$$\mathcal{C} := \mathcal{O}\left(d_0^{4/3}\max\{L,1\}^{2/3}\max\left\{\frac{1}{\bar{\rho}},\ \frac{d_0}{\bar{\varepsilon}}\right\}^{2/3}\right).$$

On the other hand, for small values of the tolerances $\bar{\rho}$ and $\bar{\varepsilon}$, the dominant term in (5.13) is the second one, i.e., $(\eta+1)^{1/6}\mathcal{C}$. Hence, the ergodic complexity bound of

Theorem 5.2 differs from the ergodic one of [7] by an

$$\mathcal{O}\left(\frac{(\eta+1)^{1/6}}{\log\log\left[Ld_0+(L\bar{\rho})^{-1}+\bar{\varepsilon}^{-1}\right]}\right)$$

factor. Note that for the case in which $X=\mathbb{R}^n$ and hence $\eta=0$, the ergodic complexity bound of Theorem 5.2 is better than the one obtained in [7]. Moreover, while the method of [7] (approximately) solves a linearized VI subproblem at every iteration, the method presented in this paper solves a Newton system of linear equations with respect to (1.3), and hence its iterations are cheaper than the ones of the method of [7].

The complexity bounds of Theorem 5.2 depend on $\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,z_0}$, where $(\tilde{\nu}_0,z_0)$ is the input for the Phase I procedure and $\tilde{x}_0$ is determined by (5.1). The next results provides a bound on $\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,z_0}$ in terms of the quantities $\tilde{\nu}_0$, $\|F(z_0^P)\|$, and $\|F'(z_0^P)\|$, where z_0^P is the projection of z_0 onto X. This bound clearly implies alternative complexity bounds for the combined Phase I/HPE-IP method.

PROPOSITION 5.3. *Let $\beta\in(0,1)$, $(\tilde{\nu}_0,z_0)\in\times\mathbb{R}_{++}\times\mathbf{E}$ be given and let $\tilde{x}_0\in\operatorname{dom}h$ be as in the statement of Phase* I *procedure. Then,*

$$\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,\tilde{x}_0}\le\frac{L\eta}{2\tilde{\nu}_0^{3/2}}+\frac{\|F(z_0^P)\|}{\tilde{\nu}_0^{1/2}}+\frac{\sqrt{\eta}\|F'(z_0^P)\|}{\tilde{\nu}_0},$$

where $z_0^P:=P_X(z_0)$.

Proof. It follows from (5.1) and Proposition 2.7 with $y=\tilde{x}_0$, $q=\nabla h(\tilde{x}_0)$, and $a=0$ that

$$\langle z_0-\tilde{x}_0,z_0^P-\tilde{x}_0\rangle=\frac{\langle\nabla h(\tilde{x}_0),z_0^P-\tilde{x}_0\rangle}{\tilde{\nu}_0}\le\frac{\eta}{\tilde{\nu}_0}.$$

Also, using a well-known property of the projection onto a closed convex set, we have

$$\langle z_0-\tilde{x}_0,z_0^P-\tilde{x}_0\rangle=\langle z_0-z_0^P,z_0^P-\tilde{x}_0\rangle+\langle z_0^P-\tilde{x}_0,z_0^P-\tilde{x}_0\rangle\ge\|z_0^P-\tilde{x}_0\|^2.$$

Hence, from the above two conclusions, we conclude that

$$\|z_0^P-\tilde{x}_0\|\le\sqrt{\frac{\eta}{\tilde{\nu}_0}}.$$

The result now follows from the above relation, the fact that (3.1) and the triangle inequality for norms imply that

$$\begin{aligned}\|F(\tilde{x}_0)\|&\le\|F(\tilde{x}_0)-F(z_0^P)-F'(z_0^P)(\tilde{x}_0-z_0^P)\|+\|F(z_0^P)+F'(z_0^P)(\tilde{x}_0-z_0^P)\|\\&\le\frac{L}{2}\|z_0^P-\tilde{x}_0\|^2+\|F(z_0^P)\|+\|F'(z_0^P)\|\|\tilde{x}_0-z_0^P\|,\end{aligned}$$

and the second relation in (3.11) of Lemma 3.2. ∎

Appendix A. Technical lemmas.

LEMMA A.1. *The following statements hold:*

(a) *If $A-B\in\mathcal{S}_+^{\mathbf{E}}$, then $\|\cdot\|_A\ge\|\cdot\|_B$ and $\|\cdot\|_A^*\le\|\cdot\|_B^*$.*

(b) $\operatorname{dom}\|\cdot\|_A^*=\operatorname{Range}(A)$ *and, for every $u\in\mathbf{E}$ and $h\in\mathbf{E}$ such that $Ah=u$, there holds*

$$\|u\|_A^*=\sqrt{\langle u,h\rangle}=\sqrt{\langle h,Ah\rangle}=\|h\|_A. \qquad \text{(A.1)}$$

(c) *If A is nonsingular, then* $\operatorname{dom}\|\cdot\|_A^* = \mathbf{E}$ *and* $\|u\|_A^* = \|u\|_{A^{-1}}$ *for every* $u \in \mathbb{R}^n$.

(d) *If* $\{A_k\} \subset \mathcal{S}_+^{\mathbf{E}}$ *is a sequence converging to A and such that the matrix* $A_k - A \in \mathcal{S}_+^{\mathbf{E}}$ *for every* $k \geq 1$, *then*

$$\lim_{k\to+\infty} \|u\|_{A_k}^* = \|u\|_A^* \quad \forall u \in \mathbf{E}.$$

(e) *The function* $\|\cdot\|_A^*$ *restricted to* $\operatorname{Range}(A)$ *is a norm.*

(f) *For every* $u \in \operatorname{Range}(A)$ *and* $v \in \mathbf{E}$, *we have* $\langle u, v\rangle \leq \|u\|_A^* \|v\|_A$.

Proof. (a) The proof of this statement follows directly from the definitions of the seminorms in (2.5) and (2.6).

(b) Assume first that $u \in \operatorname{Range}(A)$, i.e., $u = Ah \in \operatorname{Range}(A)$ for some $h \in \mathbf{E}$. Since h satisfies the first-order optimality condition of the maximization problem (2.6) and the objective function of this problem is concave, we conclude that h is an optimal solution of (2.6). Hence, the optimal value $\|u\|_A^*$ of (2.6) is finite and (A.1) holds. Assume now that $u \notin \operatorname{Range}(A)$ and consider the decomposition $u = u_0 + u_r$, where $u_0 \in \mathcal{N}(A)$ and $u_r \in \operatorname{Range}(A)$. Clearly, $\langle u_0, u_r\rangle = 0$ and $u_0 \neq 0$. In view of the definition of $\|u\|_A^*$ in (2.6), for every $t \in \mathbb{R}$, the vector $h_t := tu_0$ satisfies

$$\left(\|u\|_A^*\right)^2 \geq 2\langle u, h_t\rangle - \langle Ah_t, h_t\rangle = 2t\|u_0\|^2,$$

where the equality follows from the definition of h_t and fact that $u_0 \in \mathcal{N}(A)$, $\langle u_0, u_r\rangle = 0$ and $u = u_0 + u_r$. Letting $t \uparrow \infty$ in the above inequality and noting that $u_0 \neq 0$, we then conclude that $\|u\|_A^* = \infty$. We have thus shown that (b) holds.

(c) This statement follows directly from (b) and (2.5).

(d) For every $h \in \mathbf{E}$, define the function $p_h : \mathbf{E} \times \mathcal{S}^{\mathbf{E}} :\to R$

$$p_h(u, A) = 2\langle u, h\rangle - \langle Ah, h\rangle.$$

Since p_h is a linear function for every $h \in \mathbf{E}$, the function $p : \mathbf{E} \times \mathcal{S}^{\mathbf{E}} :\to (-\infty, +\infty]$ defined as

$$p(u, A) = \sup\{p_h(u, A) \;:\; h \in \mathbf{E}\}$$

is lower semicontinuous. This implies that

$$\|u\|_A^* = p(u, A) \leq \liminf_{k\to+\infty} p(u, A_k) = \liminf_{k\to+\infty} \|u\|_{A_k}^* \quad \forall u \in \mathbf{E}.$$

The assumption that $A_k - A \in \mathcal{S}_+^{\mathbf{E}}$ and statement (c) imply that $\|u\|_{A_k}^* \leq \|u\|_A^*$ for every $k \geq 1$ and $u \in \mathbf{E}$ and hence that $\limsup_{k\to+\infty} \|u\|_{A_k}^* \leq \|u\|_A^*$ for every $u \in \mathbf{E}$. We have thus shown that (d) holds.

(e) Choosing $A_k = A + (1/k)I$ for every $k \geq 1$, it follows from (d) that the function $\|\cdot\|_A^*$ is the pointwise limit of the norms $\|\cdot\|_{A_k}^*$, and hence it is easily seen to be a seminorm on its domain $\operatorname{Range}(A)$. Now, let $u \in \operatorname{Range}(A)$ be such that $\|u\|_A^* = 0$. Also, let $h_u \in \mathbf{E}$ be such that $Ah_u = u$. Then, it follows by (b) that $\|A^{1/2}h_u\| = \|u\|_A^* = 0$. This implies that $A^{1/2}h_u = 0$ and hence that $u = Ah_u = 0$. Thus, (e) follows.

(f) Let $u \in \operatorname{Range}(A)$ and $v \in \mathbf{E}$ be given. Assume first that $\|v\|_A = 0$. Clearly, this implies that $v \in \mathcal{N}(A)$. Since the subspaces $\mathcal{N}(A)$ and $\operatorname{Range}(A)$ are orthogonal, we conclude that $\langle u, v\rangle = 0$ and hence that (f) holds in this case. Assume now that $\|v\|_A > 0$ and define

$$\tilde{h} := \frac{\langle u, v\rangle}{\|v\|_A^2} v.$$

Since the objective function of (2.6) evaluated at $\tilde{h}$ is equal to $[\langle u, v\rangle]/\|v\|_A]^2$ and the optimal value $(\|u\|_A^*)^2$ of (2.6) exceeds this value, we conclude that (f) holds for this case too. ∎

LEMMA A.2. *Let h be a closed convex function such that the set of minimizers S^* of h is nonempty. Then, for every $\nu > 0$ and $\bar{x} \in \mathbf{E}$, the function h_ν defined as $h_\nu(x) := h(x) + (\nu/2)\|x - \bar{x}\|^2$ for every $x \in \mathbf{E}$ has a unique minimizer x_ν^* and*

$$\lim_{\nu \to 0} x_\nu^* = P_{S^*}(\bar{x}).$$

Proof. The assumptions clearly imply that S^* is a closed convex set. Since h_ν is a strongly convex function, it has a unique minimizer x_ν^* over $\mathbf{E}$ and in particular

$$h(x_\nu^*) + \frac{\nu}{2}\|x_\nu^* - \bar{x}\|^2 \le h(x^*) + \frac{\nu}{2}\|x^* - \bar{x}\|^2,$$

where $x^* := P_{S^*}(\bar{x})$. Since $x^* \in S^*$, and hence x^* is a minimizer of h, we have $h(x^*) \le h(x_\nu^*)$, which together with the above relation then imply that

$$\|x_\nu^* - \bar{x}\| \le \|x^* - \bar{x}\|, \qquad \lim_{\nu \to 0^+} h(x_\nu^*) = h(x^*).$$

Thus, it follows that the set $\{x_\nu^* \,:\, \nu > 0\}$ is bounded and hence that every accumulation point $\hat{x}$ of any sequence $\{x_{\nu_k}\}$ such that $\nu_k \to 0$ as $k \to +\infty$ satisfies

$$\|\hat{x} - \bar{x}\| \le \|x^* - \bar{x}\|, \quad \hat{x} \in S^*,$$

and hence that $\hat{x} = x^*$ due to the definition of x^*. We have thus shown that $\lim_{\nu \to 0} x_\nu^* = P_{S^*}(\bar{x})$. ∎

REFERENCES

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

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

[3] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. II, Springer-Verlag, New York, 2003.

[4] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), pp. 154–158.

[5] R. D. C. Monteiro and B. F. Svaiter, *On the complexity of the hybrid proximal projection method for the iterates and the ergodic mean*, SIAM J. Optim., 20 (2010), pp. 2755–2787.

[6] 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 J. Optim., 21 (2011), pp. 1688–1720.

[7] 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 (2012), pp. 914–935.

[8] R. D. C. Monteiro and B. F. Svaiter, *Iteration-complexity of block-decomposition algorithms and the alternating minimization augmented lagrangian method*, SIAM J. Optim., 23 (2013), pp. 475–507.

[9] R. D. C. Monteiro and B. F. Svaiter *An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods*, SIAM J. Optim., 23 (2013), pp. 1092–1125.

[10] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim., 15 (2005), pp. 229–251.

[11] A. S. Nemirovski and M. J. Todd, *Interior-point methods for optimization*, Acta Numer., 17 (2008), pp. 191–234.

[12] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Appl. Optim., Kluwer Academic, Boston, 2004.

[13] Y. Nesterov, *Dual extrapolation and its applications to solving variational inequalities and related problems*, Math. Program., 109 (2007), pp. 319–344.

[14] Y. Nesterov, *Accelerating the cubic regularization of Newton's method on convex problems*, Math. Program., 112 (2008), pp. 159–181.

[15] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Stud. Appl. Math. 13, SIAM, Philadelphia, 1994.

[16] Y. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Math. Program., 108 (2006), pp. 177–205.

[17] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, MPS/SIAM Ser. Optim., SIAM, Philadelphia, 2001.

[18] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

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

[20] M. R. Sicre and B. F. Svaiter, *An $\mathcal{O}(1/k^{3/2})$ hybrid proximal extragradient primal-dual interior point method for non-linear monotone complementarity problems*, preprint, Associação Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro.

[21] 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 (1999), pp. 323–345.

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

[23] 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 (2000), pp. 214–230.

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