A hybrid proximal extragradient self-concordant primal barrier method for monotone variational inequalities

Renato D.C. Monteiro * Mauricio R. Sicre † B. F. Svaiter ‡

June 30, 2013

Abstract

In this paper we present a primal interior-point hybrid proximal extragradient (HPE) method for solving a monotone variational inequality over a closed convex set endowed with a self-concordant barrier and whose underlying map has Lipschitz continuous derivative. In contrast to the method of [7] in which each iteration required an approximate solution of a linearized variational inequality over the original feasible set, the present one only requires solving a Newton linear system of equations. The method performs two types of iterations, namely: those which follow an ever changing path within a certain "proximal interior central surface" and those which correspond to a large-step HPE iteration of the type described in [7]. Due to its first-order nature, the iteration-complexity of the method is shown to be faster than its 0-th order counterparts such as Korpelevich's algorithm and Tseng's modified forward-backward splitting method.

Keywords: hybrid proximal extra-gradiente, interior point methods, monotone variational inequality, complexity.

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

1 Introduction

The very early papers dealing with iteration-complexity analysis of methods for VIs are as follows. Nemirovski [9] studied the complexity of Korpelevichs extragradient method under the assumption that the feasible set is bounded and an upper bound on its diameter is known. Nesterov [12] proposed a new dual extrapolation algorithm for solving VI problems whose termination depends on the guess of a ball centered at the initial iterate.

A broad class of optimization, saddle-point, equilibrium and variational inequality (VI) problems can be posed as the *monotone inclusion* (MI) problem, namely: finding x such that $0 \in T(x)$, where

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by 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. (email: 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, Brasil (email: 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.

T is a maximal monotone point-to-set operator. The proximal point method (PPM), proposed by Martinet [4], and further generalized by Rockafellar [16, 17], is a classical iterative method for solving the MI problem which generates a sequence $\{x_k\}$ according to

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

where $\lambda_k > 0$. It has been used as a framework for the design and analysis of several implementable algorithms. The classical inexact version of the proximal point method allows for the presence of a sequence of summable errors in the above iteration, that is:

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

Convergence results under the above error condition have been established in [17] and have been used in the convergence analysis of other methods that can be recast in the above framework [16].

New inexact versions of the proximal point method with relative error tolerance were proposed by Solodov and Svaiter [19, 20, 21, 22]. Iteration-complexity results for one of these inexact versions of the proximal point method introduced in [19], namely the hybrid proximal extragradient (HPE) method, were established in [5]. Application of this framework in the iteration-complexity analysis of several zero-order (or, in the context of optimization, first-order) methods for solving monotone variational inequalities (VI), and monotone inclusion 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-complexity 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 over a closed convex set X is equivalent to a MI problem whose corresponding operator is obtained by adding the point-to-point map F of the VI to the normal cone operator N_X of its feasible set. The latter paper presented 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 a class of inexact proximal point interior-point methods in which each iteration can be viewed as performing an approximate proximal point iteration to the approximate 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

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

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 parameters (μ, λ, z) are then updated. The resulting method performs two types of iterations which depend on the way these parameters are updated. On the "interior-point" 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 certain fixed z. On the other hand, the second type of 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 here that the complexity of the resulting method is about the order of magnitude as the one of the algorithm presented in [7]. Moreover, while the latter algorithm (approximately) solves linearized VIs subproblems at every iteration, the former method solves a Newton linear system of equations.

It should be noted that prior to this work, [18] presents 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: 1) the algorithm of [18] deals with the special of VI in which $X = \mathbb{R}^n_+ \times \mathbb{R}^m$, and; 2) the algorithm here is a primal one while the one in [18] 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 [15], a Newton-type method for unconstrained convex optimization 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 [13]. It should be mentioned that these methods are specifically designed for (unconstrained) convex optimization and hence do not apply to the problems studied in this paper.

This paper is organized as follows. Section 2 contains three subsections as follows. 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 iteration-complexity. Subsection 2.3 reviews some basic properties of self-concordant functions and barriers. Section 3 describes the problem of interest, namely the monotone VI problem, introduces the proximal interior central surface $\{x(\mu, \lambda, z) \,|\, (\mu, \lambda, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}^n\}$, and describes its main properties. This section contains two subsections as follows. Subsection 3.1 introduces the proximal interior central surface and gives conditions for this surface to approach the solution of the VI problem. Subsection 3.2 introduces a neighborhood of the proximal interior central surface and shows that it has the quadratic convergence property with respect to a Newton step applied to 1. Section 4 contains two subsections as follows. Subsection 4.1 presents an hybrid proximal extragradiente interior-point algorithm (HPE-IP) for solving the monotone VI problem, and performs a preliminary analysis of the algorithm. Subsection 4.2 estimates the iteration-complexity of the HPE-IP algorithm. Section 5 discusses a Phase I procedure for computing the required input for the HPE-IP algorithm and establishes its iteration-complexity. The Appendix gives the proof of some technical results.

1.1 Notation

Throughout this paper, we let $\mathbf{E}$ denote a finite-dimensional real vector space with inner product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. The range and null spaces of a linear operator $A : \mathbf{E} \to \mathbf{E}$ are denoted by $\mathcal{R}(A) := \{Ah \mid h \in \mathbf{E}\}$ and $\mathcal{N}(A) := \{u \in \mathbf{E} \mid Au = 0\}$, 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

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

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

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

The cardinality of a finite set $A \subset \mathbb{N}$ is denoted by $\#A$. For $t > 0$, we let $\log^+(t) := \max\{\log(t), 0\}$. The domain of definition of a one-to-one function F is denoted by $Dom(F)$.

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 [18] together with its corresponding complexity results. The third subsection reviews the definitions and some properties of self-concordant 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

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

Alternatively, one can consider T as a multi-valued 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,

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

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

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

and it is *maximal monotone* if it is monotone and maximal in the family of monotone operators with respect to the partial order of the inclusion.

A well-known example of a maximal monotone operator is the *normal cone operator* $N_X : \mathbf{E} \rightrightarrows \mathbf{E}$ of a closed convex set $X \subset \mathbf{E}$ defined as

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

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

$$T^\varepsilon(x) = \{v \in \mathbf{E} \mid \langle x - \tilde{x}, v - \tilde{v}\rangle \geq -\varepsilon, \quad \forall \tilde{x} \in \mathbf{E},\ \forall \tilde{v} \in T(\tilde{x})\}, \quad \forall\, x \in \mathbf{E}. \tag{2}$$

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 $\varepsilon_1 \le \varepsilon_2$, 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 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 $\varepsilon \ge 0$, it follows from a) and d), that $T^{\varepsilon}(x) \supset T(x)$ 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 $\varepsilon = 0$.

2.2 The underelaxed large-step HPE method

In this subsection, we review an underelaxed version of the large-step HPE method presented in [18] and its corresponding complexity results.

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

$$0 \in T(x). \tag{3}$$

Next we present the underelaxed large-step HPE method for solving (3).

Underelaxed large-step HPE Method:

0) Let $x_0 \in \mathbf{E}$, $c > 0$, $t \in (0, 1]$, and $\sigma \in [0, 1)$ be given and set $k = 1$;

1) if $0 \in T(x^{k-1})$, 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

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

and

$$\lambda_k \|x_k - z_{k-1}\| \ge c > 0;$$

2) choose relaxation parameter $t_k \in [t, 1]$ and define $z_k = z_{k-1} - t_k \lambda_k v_k$ and go to step 1.

end

We now make some remarks about the underelaxed large-step HPE method. The special case in which $t = 1$, and hence $t_k = 1$ for all k, corresponds to the large-step HPE method introduced in [7], which in turn is a particular case of the large-step HPE method for smooth operators presented in [5]. The iteration-complexities of the HPE and the large-step HPE methods were established in [5] and [7], respectively. In the next result we present the abstract pointwise and ergodic complexity estimations of the underelaxed large-step HPE method.

Define, for each k,

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

Proposition 2.2. *If $T^{-1}(0)$ is non-empty and d_0 is the distance of z_0 to $T^{-1}(0)$, then, for every k,*

a) there exists $i_0 \le k$ such that

$$\|v_{i_0}\| \le \frac{d_0^2}{ct(1-\sigma)k}, \qquad \varepsilon_{i_0} \le \frac{\sigma^2 d_0^3}{2ct^{3/2}(1-\sigma^2)^{3/2}k^{3/2}};$$

b) $\bar{v}_k \in T^{[\bar{\varepsilon}_k]}$,

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

Proof. For a proof of this result see Proposition 3.4 of [18]. □

2.3 Basic properties of self-concordant functions and barriers

In this subsection we review some basic properties of self-concordant functions and barriers which will useful in our presentation. A detailed treatment of this topic can be found for example in [11] and [14].

Given $A \in \mathcal{S}_+^{\mathbf{E}}$ a self-adjoint positive semidefinite symmetric linear operator in $\mathbf{E}$, we consider two types of (semi-) norms induced by A. The first one is the seminorm in $\mathbf{E}$ defined as

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

The second one is defined as

$$\|u\|_A^* := \sup\left\{2\langle u, h\rangle - \langle Ah, h\rangle \mid h \in \mathbf{E}\right\}^{1/2}, \quad \forall u \in \mathbf{E}. \tag{4}$$

Some basic properties of these norms 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 the Lemma A.1 justifies the use of a norm notation for $\|\cdot\|_A^*$.

Definition 1. *A proper closed convex function $h : \mathbf{E} \to (-\infty, \infty]$ is said to be self-concordant (SC) if $Dom(h)$ is open, h is three-times continuously differentiable and*

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

If additionally $\nabla^2 h(x)$ is nonsingular for every $x \in Dom(h)$, then h is said to be a nondegenerate self-concordant function.

Definition 2. *A SC-function h is called a η-SC barrier if, for every $x \in Dom(h)$, $\|\nabla h(x)\|^*_{\nabla^2 h(x)} \le \sqrt{\eta}$.*

Lemma 2.3. *Let h be a closed convex function such that S^*, the set of minimizers of h, is nonempty. Then, for every $\nu > 0$, the function h_ν defined as $h_\nu(x) := h(x) + (\nu/2)\|x - \bar{x}\|^2$, $\forall x \in \mathbf{E}$, has a unique minimizer x_ν^* and it holds*

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

Proof. From the assumptions it follows that S^* is a closed convex set. Let $x^* = P_{S^*}(\bar{x})$. Since h_ν is a strongly convex function, it has unique minimizer x^*_ν over $\mathbf{E}$ and in particular it holds

$$h(x^*_\nu) + \frac{\nu}{2}\|x^*_\nu - \bar{x}\|^2 \leq h(x^*) + \frac{\nu}{2}\|x^* - \bar{x}\|^2.$$

Hence, it follows

$$\begin{aligned} \frac{\nu}{2}\|x^*_\nu - \bar{x}\|^2 &= \left(h(x^*_\nu) + \frac{\nu}{2}\|x^*_\nu - \bar{x}\|^2\right) - h(x^*_\nu) \\ &\leq \left(h(x^*) + \frac{\nu}{2}\|x^* - \bar{x}\|^2\right) - h(x^*) = \frac{\nu}{2}\|x^* - \bar{x}\|^2. \end{aligned}$$

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

$$h(\hat{x}) \leq h(x^*), \quad \|\hat{x} - \bar{x}\| \leq \|x^* - \bar{x}\|.$$

From the first above relation if follows that $\hat{x} \in S^*$, which combined with the second above relation and the definition of x^* implies $\hat{x} = x^*$. Thus, any sequence $\{x_{\nu_k}\}$ such that $\nu_k \to 0$ when $k \to +\infty$ converges to $P_{S^*}(\bar{x})$, which implies the claim of the lemma. □

The next result gives some basic properties of a SC-function. Below we give a prove for a degenerate case.

Proposition 2.4. *If h is a SC-function and $x \in Dom(h)$ is such that $\|\nabla h(x)\|^*_{\nabla^2 h(x)} < 1$, then the following statements hold:*

a) $\nabla h(x) \in \mathcal{R}(\nabla^2 h(x))$ and for every x^+ such that $\nabla h(x) + \nabla^2 h(x)(x^+ - x) = 0$, we have $x^+ \in Dom(h)$ 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 of a) follows from the assumption $\|\nabla h(x)\|^*_{\nabla^2 h(x)} < 1$ and Lemma A.1(b). Moreover, the second inclusion and the inequality of a) follows Theorems 2.1.1 and 2.2.1 of [14], respectively. Also, the first part of b) follows from Theorem 2.2.2 of [14]. We also observe that the inequality in b) has already been established in Theorem 4.1.13 of [11] under the assumption that h is a nondegenerate SC-function (see also [10]). 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

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

where the first and third inequalities follow from Lemma A.1(a). Since by Lemma 2.3 x_ν^* converges to the minimizer x^* of h closest to x as $\nu \to 0$, the result follows. □

The following result is equivalent to Proposition 2.4(a), but it is in a form which is more suitable for our analysis in this paper. For every $x \in Dom(h)$ and $y \in \mathbf{E}$, define

$$L_{h,x}(y) := \nabla h(x) + \nabla^2 h(x)(y - x). \tag{5}$$

Proposition 2.5. *If h is a SC-function and, $x \in Dom(h)$ and $y \in \mathbf{E}$ are points such that $r := \|y - x\|_{\nabla^2 h(x)} < 1$, then $y \in Dom(h)$ and*

$$\|\nabla h(y) - L_{h,x}(y)\|^*_{\nabla^2 h(y)} = \|\nabla h(y) - \nabla h(x) - \nabla^2 h(x)(y - x)\|^*_{\nabla^2 h(y)} \le \left(\frac{r}{1 - r}\right)^2. \tag{6}$$

Proof. First note that the equality in (6) follows immediately from (5). To show the inequality in (6), 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 $\tilde{x} \in Dom(h)$, it follows from (5) that

$$\nabla \phi(\tilde{x}) = \nabla h(\tilde{x}) - \nabla h(x) - \nabla^2 h(x)(y - x), \quad \nabla^2 \phi(\tilde{x}) = \nabla^2 h(\tilde{x}), \tag{7}$$

and hence

$$\|\nabla \phi(x)\|^*_{\nabla^2 \phi(x)} = \|\nabla^2 h(x)(y - x)\|^*_{\nabla^2 h(x)} = \|y - x\|_{\nabla^2 h(x)} < 1, \tag{8}$$

where the last identity follows from Lemma A.1(b). Since ϕ is also a SC-function and (7) implies that

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

it follows from (8) and Proposition 2.4(a) that $y \in Dom(h)$ and

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

and hence that (6) holds in view of (7) and (8). □

The next result for the case in which $a = 0$ (and possibly $Dom(h)$ unbounded) is proved in Proposition 2.3.2(i.2) of [14]. The general case in which $a \in [0, 1)$ is proved in Theorem 4.2.7 of [11] under the assumptions that h is a nondegenerate barrier with bounded domain. Here we extend the general case for SC barriers without assuming neither non degeneracy of the barrier nor boundedness of its domain. The first part of the proof closely follows that of Theorem 4.2.7 of [11] except that it relies on Proposition 2.4(b) instead of the boundedness of $Dom(h)$.

Proposition 2.6. *Let h be a η-SC barrier and let $y \in Dom(f)$ and $q \in \mathbf{E}$ satisfy $\|q - \nabla h(y)\|^*_{\nabla^2 h(y)} \leq a < 1$. Then,*

$$\langle q, u - y \rangle \leq \delta, \quad \forall u \in Dom(h)$$

where

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

As a consequence, $q \in N^\delta_X(y)$ where $X = cl\,(Dom(h))$.

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 $x \in Dom(\phi)$, the proximity assumption of the proposition, Definition 2 and Lemma A.1(e) imply that

$$\|\nabla \phi(y)\|^*_{\nabla^2 h(y)} \leq a, \qquad \|q\|^*_{\nabla^2 h(y)} \leq \sqrt{\eta} + a. \tag{9}$$

Since ϕ is a SC-function and it follows from (9) and Proposition 2.4(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 (9) 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

$$\langle q, u - x^* \rangle = \langle \nabla h(x^*), u - x^* \rangle \leq \eta, \quad \forall\, u \in Dom(h).$$

The first part of the proposition now follows by combining the last two inequalities. The second part follows from the first one and the definitions of $N_X(\cdot)$ and $N^\eta_X(\cdot)$. $\square$

Corollary 2.7. *Let $a \in (0, 1)$, $x \in Dom(h)$ and $y \in \mathbf{E}$ be such that*

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

Then, $y \in Dom(h) \subseteq X$ and $L_h(y; x) \in N^\delta_X(y)$ where δ is as in Proposition 2.6.

Proof. The assumptions and Proposition 2.5 imply that $y \in Dom(h)$ and $\|\nabla h(y) - L_h(y; x)\|_{\nabla^2 h(x)} \leq a$. Hence, it follows from Proposition 2.6 with $q = L_h(y; x)$ that $L_h(y; x) \in N^\delta_X(y)$. $\square$

3 The main problem and preliminary technical results

This section describes the main problem and motivates our approach towards obtaining its solution. It consists of two subsections. The first one describes a proximal interior central surface parametrized by a triple $(\mu, \nu, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$ and gives conditions on these parameters under which the corresponding point $x(\mu, \nu, z)$ on the surface approaches a solution of our problem. The second subsection introduces a neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta)$ of $x(\mu, \nu, z)$ whose size depends on a specified opening $\beta \in (0, 1)$ and shows that this neighborhood enjoys the quadratic convergence property under a certain Newton step iterate in the sense that the resulting Newton iterate will be in $\mathcal{N}_{\mu,\nu,z}(\beta^2)$

Consider the inclusion problem

$$0 \in T(x) = (F + N_X)(x), \tag{10}$$

where $X \subseteq \mathbf{E}$ and $F : Dom(F) \subseteq \mathbf{E} \to \mathbf{E}$ satisfy

C.1) X is a closed convex set endowed with a η-SC barrier h such that $cl\,(Dom(h)) = X$;

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

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

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

where the norm on the left hand side is the operator norm;

C.4) the solution set X^* of problem (10) is non-empty.

We 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]). From Assumption C.3 it follows

$$\|F(y) - F(x) - F'(x)(y - x)\| \leq \frac{L}{2}\|y - x\|^2 \tag{11}$$

for any $x, y \in X$.

3.1 Proximal interior map

In this subsection we introduce a proximal interior map which argument is a triple $(\mu, \nu, z) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$ and give conditions on these parameters under which the corresponding image point $x(\mu, \nu, z)$ approaches a solution of problem (10).

The classical central path for (10) assigns to each $\mu > 0$ the solution x_μ of

$$0 = \mu F(x) + \nabla h(x). \tag{12}$$

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 (10) as μ goes to ∞ (see for example [14]). Interior-point path following methods for solving (10) have been proposed in [14] under the assumption that h satisfies C.1 and F is β-compactible with h for some $\beta \geq 0$ (see Definition 7.3.1 in [14]). It is worth noting that Assumptions C.1-C.3 do not imply that F is β-compactible with h for any $\beta \geq 0$ even when F

is an analytic map. Hence, it is not clear how the interior-point path following methods of [14] can be used to solve (10) 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 (12), our approach is based on the regularized perturbed equation

$$0 = G_{\mu,\nu,z}(x) \tag{13}$$

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

$$G_{\mu,\nu,z}(x) := \mu F(x) + \nabla h(x) + \nu(x - z), \quad \forall x \in Dom(h). \tag{14}$$

As opposed to (12), equation (14) has a (unique) solution regardless of whether the solution set of (10) 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) which guarantee that $x(\mu, \nu, z)$ approaches the solution set of (10).

Proposition 3.1. *If $\{(\mu_k, \nu_k, z_k)\} \subseteq \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$ is a sequence such that $\{z_k\}$ is bounded, $\lim_{k\to\infty} \mu_k/\nu_k = \infty$, and for some $\bar{\nu} > 0$,*

$$\nu_k \geq \bar{\nu} > 0 \quad k = 0, 1, 2, \ldots, \tag{15}$$

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 (10).

Proof. For any $k \in \mathbb{N}$, define

$$x_k := x(\mu_k, \nu_k, z_k), \quad v_k := F(x_k) + \frac{1}{\mu_k}\nabla h(x_k), \quad \lambda_k := \frac{\mu_k}{\nu_k}, \quad \varepsilon_k := \frac{\eta}{\mu_k} \tag{16}$$

and note that the assumptions of the proposition imply that

$$\lim_{k\to\infty} \lambda_k = \infty, \quad \lim_{k\to\infty} \varepsilon_k = 0. \tag{17}$$

Using (16) and the fact that x_k satisfies (13) with $(\mu, \nu, z) = (\mu_k, \nu_k, z_k)$, we conclude that

$$v_k = \frac{z_k - x_k}{\lambda_k}. \tag{18}$$

Also the definition of v_k in (16) implies

$$v_k \in \left(F + N_X^{\varepsilon_k}\right)(x_k) \subset T^{\varepsilon_k}(x_k), \tag{19}$$

where the first inclusion is due to the last claim of Proposition 2.6 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 (10).

Now, let $\hat{x}_k := (I + \lambda_k T)^{-1}(z_k)$. Using the fact that $(I + \lambda_k T)^{-1}$ is non-expansive (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

$$\hat{v}_k := \frac{z_k - \hat{x}_k}{\lambda_k} \in T(\hat{x}_k). \tag{20}$$

The latter conclusion together with (19) and (2) then imply that

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

where the equality is due to (18) and (20). Hence,

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

where the equality follows from (16) and the second inequality follows from (15). 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 (17), (18), and the boundedness of $\{x_k\}$ and $\{z_k\}$, we then conclude that $\lim_{k\to\infty} v_k = 0$. This conclusion together with (17) and Proposition 2.1(e) then imply that any accumulation point x^* of $\{x_k\}$ satisfies $0 \in T(x^*)$. Hence, that the last assertion of the proposition follows. □

3.2 A neighborhood of a proximal interior point.

In this subsection, we introduce a neighborhood, denoted by $\mathcal{N}_{\mu,\nu,z}(\beta)$, of $x(\mu,\nu,z)$ whose size depends on a specified opening $\beta \in (0,1)$. These neighborhoods will play an important role in the algorithm of Section 4. The main result of this subsection shows that a Newton iteration with respect to (13) 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.

We first introduce some preliminary notation and results. For $\nu > 0$, 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 norms simply by $\|\cdot\|_x$ and $\|\cdot\|^*_x$, respectively.

We have the following simple results.

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

$$\|u\|_{\nu,x} = \sqrt{\nu\|u\|^2 + \|u\|_x^2}, \quad \|u\|^*_{\nu,x} \leq \min\left\{\|u\|^*_x, \frac{\|u\|}{\sqrt{\nu}}\right\}, \quad |\langle u, v\rangle| \leq \|u\|_{\nu,x}\|v\|^*_{\nu,x}. \tag{21}$$

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

Lemma 3.3. *For any $x \in Dom(h)$, $\nu', \nu > 0$ and $u \in E$, we have*

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

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

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

□

The following result gives a crucial estimate for the size of the Newton direction of G at x in terms of the size of $G(x)$.

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

Proof. To simplify notation, let $G := G_{\mu,\nu,z}$. Using the definition of G, the fact that $G'(x)d_x + G(x) = 0$ and $F'(x)$ is positive semidefinite, and the definition of the norm $\|\cdot\|_{\nu,x}$, we have

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

where the last inequality follows from (21). 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 $x \in Dom(h)$, $z \in \mathbf{E}$ and $\mu, \nu > 0$ be given. Assume that the Newton direction d_x of $G = G_{\mu,\nu,z}$ at x satisfies $\|d_x\|_x < 1$ and define $x^+ = x + d_x$. Then, $x^+ \in Dom(h)$,*

$$\|\nabla h(x^+) - L_{h,x}(x^+)\|^*_{x^+} \leq \left(\tfrac{\|d_x\|_x}{1-\|d_x\|_x}\right)^2, \tag{22}$$

$$\|\mu F(x^+) + L_{h,x}(x^+) + \nu(x^+ - z)\| \leq \tfrac{\mu L}{2}\|d_x\|^2, \tag{23}$$

and

$$\|G(x^+)\|^*_{\nu,x^+} \leq \max\left\{\frac{\mu L}{2\nu^{3/2}}, \frac{1}{(1-\|d_x\|_x)^2}\right\}\|d_x\|^2_{\nu,x}. \tag{24}$$

Proof. Since $\|x^+ - x\|_x = \|d_x\|_x < 1$, the inclusion $x^+ \in Dom(h)$ and (22) follow from Proposition 2.5 with $y = x^+$.

Direct calculations show 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 (11) proves (23).

Using the definition of G, triangle inequality, the second relation in (21), (22) and (23) it follows

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

To end the proof of (24) use the above inequality and the first relation in (21). □

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 $\theta \in [0,1)$, $\mu, \nu > 0$, $x \in Dom(h)$, and $z \in \mathbf{E}$ be given and define*

$$\gamma_{\mu,\nu}(\theta) = \max\left\{\frac{\mu L}{2\nu^{3/2}}, \frac{1}{(1-\theta)^2}\right\}. \tag{25}$$

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

$$\|G(x)\|^*_{\nu,x} \leq \theta,$$

then $x^+ \in Dom(h)$ and

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

Proof. To simplify the exposition, let us denote $\gamma_{\mu,\nu}(\theta)$ simply by γ. Lemma 3.4, the first relation in (21) and the assumption that $\|G(x)\|^*_{\nu,x} \leq \theta$ imply that

$$\|d_x\|_x \leq \|d_x\|_{\nu,x} \leq \|G(x)\|^*_{\nu,x} \leq \theta < 1.$$

This observation and Proposition 3.5 then imply that $x^+ \in Dom(h)$ and

$$\gamma\|G(x^+)\|^*_{\nu,x^+} \leq \gamma^2\|d_x\|^2_{\nu,x} \leq \left[\gamma\|G(x)\|^*_{\nu,x}\right]^2.$$

□

We now introduce the neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta)$ of a point $x(\mu,\nu,z)$ of the proximal interior central path surface with opening $\beta \geq 0$ which will play an important role in the algorithm of Section 4. Indeed, for a given scalar $\beta \geq 0$, define

$$\mathcal{N}_{\mu,\nu,z}(\beta) := \{x \in Dom(h) : \gamma_{\mu,\nu}\|G_{\mu,\nu,z}(x)\|^*_{\nu,x} \leq \beta\}, \quad \text{where} \quad \gamma_{\mu,\nu} = \max\left\{\frac{\mu L}{2\nu^{3/2}}, 4\right\}. \tag{26}$$

The following simple consequence of Proposition 3.6 shows that a Newton iteration takes a point from the neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta)$ to the smaller neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta^2)$ whenever $\beta \in [0,1)$.

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. Let d_x denote the Newton direction of $G_{\mu,\nu,z}$ at x and set $x^+ = x + d_x$. Then, $x_+ \in \mathcal{N}_{\mu,\nu,z}(\beta^2)$.*

Proof. Note that (25) and the definition of $\gamma_{\mu,\nu}$ in (26) imply that

$$\gamma_{\mu,\nu} = \gamma_{\mu,\nu}(1/2) \geq 4. \tag{27}$$

The assumption that $\beta < 1$ and $x \in \mathcal{N}_{\mu,\nu,z}(\beta)$ then imply that

$$\|G(x)\|^*_{\nu,x} \leq \frac{\beta}{\gamma_{\mu,\nu}} \leq \frac{\beta}{4} \leq \frac{1}{2}. \tag{28}$$

Hence, the result immediately follows from Proposition 3.6 with $\theta = 1/2$. □

The next result shows that after one iteration of Newton's method from a point in $\mathcal{N}_{\mu,\nu,z}(\beta)$ we obtain an inexact solution of the proximal system

$$v \in (F + N_X)(x), \quad \left(\frac{\mu}{\nu}\right) v + x - z = 0.$$

It also provides some bounds which are useful to assess the quality of this inexact solution.

Proposition 3.8. *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. Let d_x denote the Newton direction of $G_{\mu,\nu,z}$ at x and set $x^+ = x + d_x$. Then,*

a) $\|\nabla h(x^+) - L_{h,x}(x^+)\|^*_{x^+} \le \beta^2/\gamma_{\mu,\nu} \le \beta^2/4$;

b) x^+ *together with the triple* $(\lambda, v^+, \varepsilon^+)$ *defined as*

$$\lambda := \frac{\mu}{\nu}, \quad v^+ := F(x^+) + \frac{1}{\mu} L_{h,x}(x^+), \quad \varepsilon^+ := \frac{1}{\mu}\left(\eta + \frac{a_x(a_x + \sqrt{\eta})}{1 - a_x}\right), \tag{29}$$

where

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

satisfy

$$v^+ \in (F + N_X^{\varepsilon^+})(x^+), \quad \|\lambda v^+ + x^+ - z\|^2 + 2\lambda\varepsilon^+ \le \frac{2}{\nu}\left(\sqrt{\eta} + \frac{\beta^2}{2}\right)^2 \tag{30}$$

and

$$\|v^+\| \le \frac{\sqrt{\nu}}{\mu}\left[\frac{\beta^2}{4} + \sqrt{\nu}\|x^+ - z\|\right], \quad \varepsilon^+ \le \frac{1}{\mu}\left[\eta + \frac{\beta^2}{3}\left(\frac{\beta^2}{4} + \sqrt{\eta}\right)\right]. \tag{31}$$

Proof. We first prove a). By (28), Lemma 3.4 and the first relation in (21), we have

$$\|d_x\|_x \le \|d_x\|_{\nu,x} \le \|G(x)\|^*_{\nu,x} \le \frac{\beta}{\gamma_{\mu,\nu}} \le \frac{\beta}{4} \le \frac{1}{2}. \tag{32}$$

The definition of $\gamma_{\mu,\nu}$ and the above inequality imply

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

The last two conclusions together with inequality (22) of Proposition 3.5 then imply that a) holds.

Now we prove b). The above two relations together with the fact that $x^+ = x + d_x$ and Corollary 2.7 with $y = x^+$ then imply that

$$L_{h,x}(x^+) \in N_X^{\delta}(x^+), \tag{33}$$

where

$$\delta := \eta + \frac{a_x(a_x + \sqrt{\eta})}{1 - a_x} \le \eta + \frac{(\beta^2/4)[(\beta^2/4) + \sqrt{\eta}]}{1 - (\beta^2/4)} \le \eta + \frac{\beta^2}{3}[(\beta^2/4) + \sqrt{\eta}], \tag{34}$$

due to the fact that $\beta < 1$. Note that the definition of v^+, relations (33) and (34), and the definitions of ε^+ and $N_X^{\varepsilon_+}(\cdot)$ imply the inclusion in (30). Now, using the definitions of $\gamma_{\mu,\nu}$, x^+, v^+, and λ, inequalities (23), (32), and the first relation in (21), we have

$$\begin{aligned} \|\lambda v^+ + x^+ - z\| &= \frac{1}{\nu}\|\mu F(x^+) + L_{h,x}(x^+) + \nu(x^+ - z)\| \\ &\le \frac{\mu L}{2\nu}\|d_x\|^2 \\ &\le \frac{\mu L}{2\nu^2}\|d_x\|_{\nu,x}^2 \le \frac{\gamma_{\mu,\nu}}{\sqrt{\nu}}\|d_x\|_{\nu,x}^2 \le \frac{\gamma_{\mu,\nu}}{\sqrt{\nu}}\left(\frac{\beta}{\gamma_{\mu,\nu}}\right)^2 \le \frac{\beta^2}{\gamma_{\mu,\nu}\sqrt{\nu}} \le \frac{\beta^2}{4\sqrt{\nu}}, \end{aligned} \tag{35}$$

where the last inequality is due to (27). The latter conclusion, the definition of λ and ε^+, and (34) then imply that

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

To end the proof, observe that the first inequality in (31) follows from (35), triangle inequality, and the definition of λ; and the second inequality follows from (34) and relation $\varepsilon^+ = \delta/\mu$. □

As a consequence of the above proposition, observe that in the case in which the term $\|x^+ - x\|$ remains bounded, taking μ and $\mu/\sqrt{\nu}$ large enough, we are able to obtain approximated solutions of problem (10), with errors bounds below any preestablished threshold. This could be accomplished, for example, by using interior-point techniques along a suitable curve belonging to the proximal interior surface. On the other hand, if term $\|x^+ - x\|$ grows too large, i.e. larger than the last term of the second inequality in (31), then we could take an underelaxed HPE step in variable z, as described in Subsection 2.2, which takes z to a point closer to the solution set of problem (10). This is the main idea behind the algorithm that will be presented in the next section.

4 The HPE interior-point algorithm

In this section we present an underelaxed HPE interior-point algorithm (HPE-IP) for solving the monotone variational inequality problem (10). This algorithm combines two types of iterations: after a Newton step for system (13), the algorithm linearly updates parameters μ and ν and take an underelaxed HPE large-step, as described in Subsection 2.2, in variable z; when this is not possible, the algorithm fixes z and linearly updates parameters μ and ν, which corresponds to an iteration of a path-following algorithm along a certain trajectory within the proximal central surface. This section is divided in two subsections. In the first one, we present the algorithm and analyze the properties of two steps described above. In the second one, we establish the iteration-complexity of the HPE-IP algorithm using the convergence rate bounds of Subsection 2.2 for the HPE steps, and a standard complexity analysis for the path-following steps.

4.1 The algorithm and preliminary results

The following algorithm generates a sequence $\{(x_k, v_k, \varepsilon_k)\}$ such that $v_k \in (F + N^{\varepsilon_k})(x_k)$. Clearly the size of the residual pair (v_k, ε_k) measures the quality of x_k as an approximate solution of (10). The main goal of this section will be to study how fast the residual sequence $\{(v_k, \varepsilon_k)\}$ and a certain associated ergodic sequence converge to zero.

HPE Interior-Point (HPE-IP) Method:

0) Let $\sigma \in [0,1)$, $\beta \in (0,1)$, $\mu_0, \nu_0 > 0$, $z_0 \in \mathbf{E}$ and $x_0 \in Dom(h)$ be such that

$$x_0 \in \mathcal{N}_{\mu_0,\nu_0,z_0}(\beta), \tag{36}$$

define

$$\gamma := \max\left\{\frac{\mu_0 L}{2\nu_0^{3/2}}, 4\right\}, \quad \tau_1 := \frac{\beta(1-\beta)}{\gamma(\sqrt{\eta}+1)}\left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1}, \quad \tau_2 := \frac{\beta(1-\beta)}{3\gamma\sqrt{\eta}}, \tag{37}$$

and set $k = 1$;

1) set $(x, z, \mu, \nu) = (x_{k-1}, z_{k-1}, \mu_{k-1}, \nu_{k-1})$ and $G = G_{\mu,\nu,z}$;

2) compute the Newton direction d_x of G at x and set $x^+ = x + d_x$ and (v^+, ε^+) as in (29);

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

$$\mu^+ = \mu(1+\tau_1)^3, \quad \nu^+ = \nu(1+\tau_1)^2, \quad z^+ = z; \tag{38}$$

3.b) else

$$\mu^+ = \frac{\mu}{(1+\tau_2)^3}, \quad \nu^+ = \frac{\nu}{(1+\tau_2)^2}, \quad z^+ = z - \frac{\tau_2}{1+\tau_2}\left(\frac{\mu}{\nu}\right) v^+; \tag{39}$$

4) let $(x_k, z_k, \mu_k, \nu_k) = (x_+, z_+, \mu_+, \nu_+)$ and $(v_k, \varepsilon_k) = (v_+, \varepsilon_+)$, set $k \leftarrow k+1$, and go to step 1.

end

We now make a few remarks about the HPE-IP method. First, in view of the update rules for μ and ν in step 3.a) or 3.b), we have

$$\left(\frac{\mu_k}{\mu_0}\right)^2 = \left(\frac{\nu_k}{\nu_0}\right)^3, \quad \forall k \ge 0. \tag{40}$$

Second, recalling the definition of $\gamma_{\mu,\nu}$ in (26) and noting the definition of γ in (37), it follows from (40) that

$$\gamma_{\mu_k,\nu_k} = \gamma, \quad \forall k \ge 0.$$

Hence, in view of (26), it follows that the definition of the neighborhood $\mathcal{N}_{\mu_k,\nu_k,z_k}(\beta)$ simplifies to

$$\mathcal{N}_{\mu_k,\nu_k,z_k}(\beta) = \left\{x \in Dom(h) \quad | \quad \|G_{\mu_k,\nu_k,z_k}(x)\|^*_{\nu_k,x} \le \beta/\gamma\right\}.$$

Third, noting that x_0 is chosen so that (36) holds, we will show in the results below that the condition $x_k \in \mathcal{N}_{\mu_k,\nu_k,z_k}(\beta)$ is maintained in every iteration of the HPE-IP method. Fourth, in those iterations in which step 3.a occurs, the computation of x^+ and (μ^+, ν^+) in steps 2 and 3.a correspond to a path-following iteration with respect to the path

$$\left\{x(\mu, \nu, z_k) : \mu = t^3\mu_0, \ \nu = t^2\nu_0, \ \ t > 0\right\}.$$

More specifically, given that $(x, z, \mu, \nu) = (x_k, z_k, \mu_k, \nu_k)$ satisfies $x \in \mathcal{N}_{\mu,\nu,z}(\beta)$, it follows from Proposition 3.7 that the point x^+ computed in step 2 belongs to the smaller neighborhood $\mathcal{N}_{\mu,\nu,z}(\beta^2)$ and the Proposition 4.2 below shows that the parameters μ and ν are increased by the uniform factors $(1+\tau_1)^3$ and $(1+\tau_1)^2$, respectively, in such a way as to keep x^+ in the β-neighborhood of the updated parameters μ^+ and ν^+. Fifth, in those iterations in which step 3.b occurs, Proposition 4.3 below shows that: i) the update rule for z^+ corresponds to a (underelaxed) large-step HPE iteration, in variable z, for the operator $F + N_X$, and; ii) the update of the parameters (μ, ν, z) to (μ^+, ν^+, z^+) keeps x^+ in the β-neighborhood of the updated parameters μ^+, ν^+ and z^+. Sixth, observe that the above algorithm assumes that an initial quadruple $(\mu_0, \nu_0, x_0, z_0) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times Dom(h) \times \mathbf{E}$ satisfying (36) is known. Section 5 describes a phase I procedure, and its corresponding iteration-complexity, which, for a given triple $(\mu_0, \nu_0, z_0) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbf{E}$, finds a point $x_0 \in Dom(h)$ satisfying (36).

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

Proposition 4.1. *Let $z, p \in \mathbf{E}$ and positive scalars $\alpha, \mu, \mu^+, \nu, \nu^+$ be given and define*

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

Then, for every $x \in Dom(h)$, we have

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

Proof. Let be $G^+(x) = G_{\mu^+,\nu^+,z^+}(x)$. By (14), it follows that

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

Using triangle inequality, Lemmas (3.2) and (3.3), it follows that

$$\begin{aligned}\|G^+(x)\|^*_{\nu^+,x} &= \left(\frac{\mu^+}{\mu} + \alpha\frac{\nu^+}{\nu}\right)\|G(x)\|^*_{\nu^+,x} + \left|\frac{\mu^+}{\mu} - 1\right|\|\nabla h(x)\|^*_x + \alpha\frac{\nu^+}{\nu}\|p - \nabla h(x)\|^*_x \\ &\quad + \left|(1-\alpha)\frac{\nu^+}{\nu} - \frac{\mu^+}{\mu}\right|\frac{\nu}{\sqrt{\nu^+}}\|x - z\| \\ &\leq \left(\frac{\mu^+}{\mu} + \alpha\frac{\nu^+}{\nu}\right)\max\left\{1, \sqrt{\frac{\nu}{\nu^+}}\right\}\|G(x)\|^*_{\nu,x} + \left|\frac{\mu^+}{\mu} - 1\right|\sqrt{\eta} \\ &\quad + \alpha\frac{\nu^+}{\nu}\|p - \nabla h(x)\|^*_x + \left|(1-\alpha)\frac{\nu^+}{\nu} - \frac{\mu^+}{\mu}\right|\frac{\nu}{\sqrt{\nu^+}}\|x - z\|,\end{aligned}$$

which clearly implies the conclusion of the proposition. □

The following result analyzes the situation in which step 3.a) occurs.

Proposition 4.2. *For some $\beta \in (0,1)$ and $\mu_0, \nu_0 > 0$, let $\mu, \nu > 0$ and $x, z \in \mathbf{E}$ be such that*

$$x \in \mathcal{N}_{\mu,\nu,z}(\beta), \quad \left(\frac{\mu}{\mu_0}\right)^2 = \left(\frac{\nu}{\nu_0}\right)^3. \tag{41}$$

Let d_x be the Newton direction of $G_{\mu,\nu,z}$ at x and define $x^+ = x + d_x$, (v^+, ε^+) as in (29) and (μ^+, ν^+, z^+) as in (38). If the condition $\sqrt{\nu}\|x^+ - z\| \leq \sqrt{2}(\sqrt{\eta}+1)/\sigma$ holds, then the following statements hold:

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

b) $v^+ \in (F + N_X)^{\varepsilon^+}(x^+)$ *and*

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

Proof. The first inclusion in a) follows from Proposition 3.7. Note that the definition of $\gamma_{\mu,\nu}$ in (26), the second condition in (41), and the first two identities in (38) imply that $\gamma_{\mu^+,\nu^+} = \gamma_{\mu,\nu} = \gamma$. Also, from the definition of τ_1 in (37) and the fact that $\gamma \geq 4$ and $\eta \geq 1$ it follows

$$\tau_1 = \left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1} \frac{\beta(1-\beta)}{\gamma(\sqrt{\eta}+1)} \leq \left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1} \frac{\beta(1-\beta)}{8} \leq \frac{1}{32}\left(4 + \frac{\sqrt{2}}{\sigma}\right)^{-1}. \tag{42}$$

By (38) and Proposition 4.1 with $x = x^+$, $\alpha = 0$ and $p = 0$ it follows

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

where the second last inequality is due to the fact that $\beta^2 \leq 1 \leq \gamma$, the last inequality is due to (42), and the last equality follows from the definition of τ_1 in (37). The second inclusion in a) follows from the above inequality and the fact that $\gamma = \gamma_{\mu^+,\nu^+}$.

Statement b) follows from relation (30) of Proposition 3.8(b), the definitions of μ^+ and ν^+ combined with the identity in (41), and the assumption $\sqrt{\nu}\|x^+ - z\| \leq \sqrt{2}(\sqrt{\eta}+1)/\sigma$. □

The following result analyzes the situation in which step 3.b) occurs.

Proposition 4.3. *For some $\beta \in (0,1)$ and $\mu_0, \nu_0 > 0$, let $\mu, \nu > 0$ and $x, z \in \mathbf{E}$ be such that* (41) *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* (29) *and (μ^+, ν^+, z^+) as in* (38). *If the condition $\sqrt{\nu}\|x^+ - z\| > \sqrt{2}(\sqrt{\eta}+1)/\sigma$ holds, then the following statements hold:*

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

b) $v^+ \in (F + N_X^{\varepsilon^+})(x^+)$ and

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

Proof. The first inclusion in a) follows from Proposition 3.7(a). Note that the definition of $\gamma_{\mu,\nu}$ in (26), the second condition in (41), the first two identities in (39) imply that $\gamma_{\mu^+,\nu^+} = \gamma_{\mu,\nu} = \gamma$. Using Proposition 4.1 with $x = x^+$, $p = L_{h,x}(x^+)$ and $\alpha = \tau_2/(1+\tau_2)$ and relation (39), 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

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

where in the second inequality we used a) and Proposition 3.8(a), and in the last inequality relations $1/(1+t) + t/(1+t)^3 \le 1$ and $1 - 1/(1+t)^3 \le 3t$ $\forall t > 0$.

We now prove b). In view of relation (30) of Proposition 3.7 and the assumption that $\sqrt{\nu}\|x^+ - z\| > \sqrt{2}(\sqrt{\eta}+1)/\sigma$, we conclude that the inclusion in b) holds and

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

and

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

where the first equality is due to the definition of λ in (29) and the last equality follows from assumption (41). □

The following result follows immediately from Propositions (4.2) and (4.3).

Proposition 4.4. *The algorithm is well-defined and, 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 3.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 3.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 $\lambda_k = \mu_{k-1}/\nu_{k-1}$*.*

Observe that iterations in which step **3.b** is executed can be regarded as (under)relaxed HPE iterations in the sequence (z_k) with relaxation parameter $h/(1+h)$.

4.2 iteration-complexity analysis of the IP-HPE method

In this subsection we establish the iteration-complexity of the HPE-IP algorithm using the convergence rate bounds of Subsection 2.2 for the HPE steps, and a standard complexity analysis for the path-following steps.

More specifically, given a pair of tolerances $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ we will estimate the number of steps of the HPE-IP algorithm to obtain a triple $(x, v, \epsilon) \in X \times \mathbf{E} \times \mathbb{R}_{++}$ satisfying either the condition

$$v \in (F + N_X^{\bar{\varepsilon}})(x), \quad \|v\| \le \bar{\rho}, \quad \varepsilon \le \bar{\varepsilon}, \tag{43}$$

or the condition

$$v \in (F + N_X)^{\bar{\varepsilon}}(x), \quad \|v\| \le \bar{\rho}, \quad \varepsilon \le \bar{\varepsilon}. \tag{44}$$

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

Define, for all $k \in \mathcal{N}$,

$$A_k = \{i \le k \mid \text{step 3.a is executed at iteration } i\}, \tag{45}$$

$$B_k = \{i \le k \mid \text{step 3.b is executed at iteration } i\}, \tag{46}$$

$$a_k = \#A_k, \quad b_k = \#B_k. \tag{47}$$

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

Define also

$$t = \frac{\tau_2}{1+\tau_2}, \quad \lambda_k = \frac{\mu_{k-1}}{\nu_{k-1}} \quad k = 1, \ldots \tag{48}$$

and, for those k such that $B_k \neq \emptyset$, let

$$\Lambda_k = \sum_{i \in B_k} t\lambda_i, \quad \bar{x}_k = \sum_{i \in B_k} \frac{t\lambda_i}{\Lambda_k} x_i, \quad \bar{v}_k = \sum_{i \in B_k} \frac{t\lambda_i}{\Lambda_k} v_i, \tag{49}$$

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

Lemma 4.5. *The following relations hold for all $k \in N$*

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

Proof. The above identities follow immediately from (45)-(47) and the update formulas (38) and (39). □

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

$$d_0 := \min\{\|z_0 - x^*\| : x^* \in X^*\}, \quad \phi_0 := \frac{L\mu_0}{2\nu_0^{3/2}}, \tag{50}$$

which have the following properties: if the number of 3.b)-type iterations performed by the IP-HPE algorithm ever becomes larger than or equal to the first (resp., second) value, then the algorithm yields a triple $(x, v, \epsilon) \in X \times \mathbf{E} \times \mathbb{R}_{++}$ satisfying (43) (resp., (44)). We observe however that there exists the possibility that the IP-HPE algorithm never performs that many number of 3.b)-type iterations, and in fact computes the desired approximate solution triple due its performing a sufficiently large number of 3.a)-type iterations. The latter situation will be analyzed within the proof of Theorem 4.7.

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*

$$K_1^b = O\left(d_0^2 \sqrt{\eta} \lceil \phi_0 \rceil \max\left\{ \tfrac{L}{\sqrt{\eta}\phi_0\bar{\rho}}, \left(\tfrac{L}{\sqrt{\eta}\phi_0\bar{\varepsilon}} \right)^{2/3} \right\} + 1 \right), \tag{51}$$

$$K_2^b = O\left(d_0^{4/3} \sqrt{\eta} \lceil \phi_0 \rceil \left(\tfrac{L}{\sqrt{\eta}\phi_0} \max\left\{ \tfrac{1}{\bar{\rho}}, \tfrac{d_0}{\bar{\varepsilon}} \right\} \right)^{2/3} + 1 \right), \tag{52}$$

and the following statements hold:

a) if k_0 is an iteration index satisfying $b_{k_0} > K_1^b$, 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 (43);

b) if k_0 is an iteration index satisfying $b_{k_0} > K_2^b$, then the triple of $(x, v, \varepsilon) = (\bar{x}_k, \bar{v}_k, \bar{\varepsilon}_k)$ satisfies (44), for every $k \geq k_0$.

Proof. a) Define

$$K_1^b = \left\lceil \max\left\{ \frac{\sigma}{\bar\rho}\left(\frac{\nu_0^{3/2}}{\sqrt{2}(\sqrt{\eta}+1)\mu_0}\right), \frac{\sigma^2}{\bar\varepsilon^{2/3}}\left(\frac{\nu_0^{3/2}}{2^{3/2}(\sqrt{\eta}+1)\mu_0}\right)^{2/3} \right\} \frac{(1+\tau_2)d_0^2}{(1-\sigma)\tau_2} \right\rceil \tag{53}$$

and observe that (51) holds due to the definitions of τ_2 in (37) and ϕ_0 in (50). Let k_0 be an iteration index satisfying $b_{k_0} \geq K_1^b$. 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 $k \notin B_{k_0}$), we conclude that the finite iteration sequence $\{(z_i, x_i, v_i, \epsilon_i),\ i \in B_{k_0}\}$ can be seen as a sequence generated according to the underelaxed large-step HPE method described in Subsection 2.2 applied to the operator $T = F + N_X$ with

$$c = \frac{\sqrt{2}(\sqrt{\eta}+1)\mu_0}{\sigma\nu_0^{3/2}}, \qquad t = \frac{\tau_2}{1+\tau_2}.$$

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

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

The above two inequalities, the assumption that $b_{k_0} \geq K_1^b$, and the definition of K_1^b easily imply that $(x, v, \varepsilon) = (x_i, v_i, \varepsilon_i)$ satisfies the two inequalities in (43). Moreover, Proposition 4.4(b) implies that this triple satisfies the inclusion in (43). Hence, a) follows.

b) Define

$$K_2^b = \left\lceil \max\left\{ \frac{d_0^{4/3}}{\bar\rho^{2/3}}, \frac{d_0^2}{(1-\sigma)^{4/3}\bar\varepsilon^{2/3}} \right\} \frac{2\sigma^{2/3}\nu_0(1+\tau_2)}{(1-\sigma^2)^{1/3}(\sqrt{\eta}+1)^{2/3}\mu_0^{2/3}\tau_2} \right\rceil \tag{54}$$

and observe that (52) is satisfied in view of the definition of τ_2 in (37) and ϕ_0 in (50). Let k_0 be an iteration index satisfying $b_{k_0} \geq K_2^b$. Using the definition of $\bar v_k$ and $\bar\varepsilon_k$ in (49) and Proposition 2.2(b) with T, c, and t as above, we conclude that for any $k \in \mathbb{N}$ such that $B_k \neq \emptyset$, we have $\bar v_k \in T^{\bar\varepsilon_k}(\bar x_k)$ and

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

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

The following results present iteration-complexity bounds for the HPE-IP algorithm to obtain approximate solutions of the $VIP(F, X)$, within a certain $(\bar\rho, \bar\varepsilon)$-tolerance. 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.

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

$$\begin{aligned} i &= \mathcal{O}\left(d_0^2\sqrt{\eta}\lceil\phi_0\rceil\max\left\{\frac{L}{\sqrt{\eta}\phi_0\bar\rho},\left(\frac{L}{\sqrt{\eta}\phi_0\bar\varepsilon}\right)^{2/3}\right\}+1\right) \\ &\quad+\mathcal{O}\left(\sqrt{\eta}\lceil\phi_0\rceil\max\left\{\log^+\left(\frac{L\sqrt{\eta}}{\phi_0\nu_0\bar\rho}\right),\log^+\left(\frac{L\eta}{\phi_0\nu_0^{3/2}\bar\varepsilon}\right)\right\}\right) \end{aligned} \tag{55}$$

such that the triple of $(x,v,\varepsilon)=(x_i,v_i,\varepsilon_i)$ satisfies (43).

Proof. Define the constants

$$\begin{aligned} C_1 &:= (1+\tau_1)^2\left(\frac{\beta^2}{4}+\frac{\sqrt{2}(\sqrt{\eta}+1)}{\sigma}\right)\left(\frac{\nu_0^{3/2}}{\mu_0}\right), \\ C_2 &:= (1+\tau_1)^3\left(\eta+\frac{\beta^2}{3}\left(\frac{\beta^2}{4}+\sqrt{\eta}\right)\right)\left(\frac{\nu_0^{3/2}}{\mu_0}\right). \end{aligned} \tag{56}$$

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 (53). Now, using the fact that $t/(1+t)\le\log(1+t)\le t$ for every $t>-1$, the definitions of τ_1 and τ_2 in (37), we easily see that

$$\frac{\log(1+\tau_1)}{\log(1+\tau_2)}=\mathcal{O}(1),\quad \frac{1}{\log(1+\tau_2)}=\mathcal{O}(\sqrt{\eta}\lceil\phi_0\rceil). \tag{57}$$

This observation together with (37), (51) and (56) then imply that K_1 can be estimated according to the right hand side of (55). 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 (43).

Indeed, to proof 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 (46) and 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

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

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 (43). □

The next result presents the iteration-complexity of the HPE-IP for the ergodic means defined in-(49).

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

$$
\begin{aligned}
K_2 \quad &= \quad \mathcal{O}\left(d_0^{4/3} \eta^{1/6} \lceil \phi_0 \rceil \left(\frac{L}{\phi_0} \max\left\{ \frac{1}{\bar{\rho}}, \frac{d_0}{\bar{\varepsilon}} \right\} \right)^{2/3} + 1 \right) \\
&\quad + \mathcal{O}\left(\sqrt{\eta} \lceil \phi_0 \rceil \max\left\{ \log^+\left(\frac{L\sqrt{\eta}}{\phi_0 \nu_0 \bar{\rho}} \right), \log^+\left(\frac{L\eta}{\phi_0 \nu_0^{3/2} \bar{\varepsilon}} \right) \right\} \right)
\end{aligned}
\tag{58}
$$

such that at least one of the following statements hold:

a) there exists an index $i \leq K_2$ such that the triple $(x_i, v_i, \varepsilon_i)$ satisfies (43)*;*

b) for every index $k \geq K_2$, the triple $(\bar{x}_k, \bar{v}_k, \bar{\varepsilon}_k)$ satisfies (44).

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 K_2^b is defined in (54) and C_1, C_2 are defined in (56). Now, (37), (51), (56) and (57) imply that K_2 satisfies (58).

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: $b_{K_2} \geq K_2^b$ and $b_{K_2} < K_2^b$. If $b_{K_2} \geq K_2^b$ 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). ☐

In view of the complexity bounds of Theorems (4.7) and (4.8), we see that using 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., $\phi_0 \in [1, 4]$) seems to be the best strategy. In the next section, we will present a Phase I procedure which generates an initial quadruple $(x_0, z_0, \mu_0, \nu_0) \in Dom(h) \times \mathbf{E} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ for the HPE-IP algorithm satisfying the condition that

$$
\frac{\phi_0}{\min\{1, L\}} \in [1, 4] . \tag{59}
$$

Note that ϕ_0 is divided by the factor $\min\{1, L\}$. Its goal is to prevent the iteration-complexity of the procedure from growing as L becomes small. Finally, we 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 $(\mu_0, \nu_0, x_0) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times Dom(h)$ such that the quadruple (x_0, z_0, μ_0, ν_0) satisfies conditions (36) and (59), 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.

Phase I Procedure:

0) Let $\beta \in (0,1)$ be given 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
$$\nabla h(\tilde{x}_0) + \tilde{\nu}_0(\tilde{x}_0 - z_0) = 0, \tag{60}$$
and let
$$\tilde{L} := \max\{L, 1\}, \quad \tilde{\mu}_0 := \min\left\{\frac{8\tilde{\nu}_0^{3/2}}{\tilde{L}}, \frac{\beta}{4\|F(\tilde{x}_0)\|^*_{\tilde{x}_0,\tilde{\nu}_0}}\right\}, \quad \tilde{\phi}_0 := \frac{L\tilde{\mu}_0}{2\tilde{\nu}_0^{3/2}}, \tag{61}$$
$$\hat{t} := \frac{\beta(1-\beta)}{4\sqrt{\eta}};$$
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
$$x^+ := x + d_x, \quad \mu^+ := \mu(1-\hat{t}\,), \quad \nu^+ := \nu(1-\hat{t}\,), \quad \phi^+ := \frac{\mu^+ L}{2\left(\nu^+\right)^{3/2}}; \tag{62}$$

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, (61) implies that $\tilde{\phi}_0/\min\{L,1\} \leq 4$ but most likely will be very small. Second, (62) implies that
$$\tilde{\phi}_k = \frac{\tilde{\phi}_{k-1}}{(1-\hat{t}\,)^{1/2}}. \tag{63}$$
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 (50).

Proposition 5.1. *The following statements hold about 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
$$\mathcal{O}\left(\sqrt{\eta}\log^+\left(\frac{8\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,\tilde{x}_0}}{\max\{L,1\}}\right) + 1\right), \tag{64}$$
iterations with a triple $(\mu_0, \nu_0, x_0) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times Dom(h)$ satisfying (36), (59) and the estimate
$$\log\frac{\tilde{\nu}_0}{\nu_0} = \mathcal{O}\left(\log^+\left(\frac{\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,\tilde{x}_0}}{\max\{L,1\}}\right) + 1\right). \tag{65}$$

Proof. We prove a) by induction on k. We first prove that a) holds for $k = 1$. Indeed, it follows from (26) and (61) that

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

which together with (14) and (60) then imply that

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

and hence that $\tilde{x}_0 \in \mathcal{N}_{\tilde{\mu}_0,\tilde{\nu}_0,z_0}(\beta)$ in view of (26). We have thus shown that a) holds for $k = 1$.

Assume now statement a) holds for the k-th iteration, and hence that $\tilde{x}_{k-1} \in \mathcal{N}_{\tilde{\mu}_{k-1},\tilde{\nu}_{k-1},z_0}(\beta)$ and $\tilde{\phi}_{k-1} \le 4\min\{1, L\}$. 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 $(\mu^+, \nu^+, x^+, \phi^+)$ as in (62). Observe that Proposition 3.7 with $z = z_0$ implies that $x^+ \in \mathcal{N}_{\mu,\nu,z_0}(\beta^2)$. Since $\phi \le 4\min\{1, L\}$ and the procedure did not stop at the k-th iteration (see step 1 of the procedure), we must have $\phi = L\mu/(2\nu^{3/2}) < \min\{L, 1\} \le 1$, and hence $\gamma_{\mu,\nu} = 4$ in view of (26). Using the fact that $\phi^+ = \phi/(1 - \hat{t})^{1/2}$ and $\hat{t} \le 1/4$ due to the definition of $\hat{t}$ in step 0 of the procedure, we conclude that $\phi_+ \le 4\min\{L, 1\} \le 4$, and hence that $\gamma_{\mu^+,\nu^+} \le 4$ in view of (26). Now, using Proposition 4.1 with $x = x^+$, $z = z_0$, $p = 0$ and $\alpha = 0$, and the definition of $\hat{t}$, we conclude that

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

and hence that $x^+ \in \mathcal{N}_{\mu^+,\nu^+,z_0}(\beta)$. Thus, 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 $\tilde{\phi}_{K-1} \ge \min\{1, L\}$. Note that, since $\tilde{\phi}_{K-1}/\min\{1, L\} \le 4$ by statement a), it follows that the output of the procedure satisfies (59). We will now show that

$$K \le 2 + \left(\frac{8\sqrt{\eta}}{\beta(1 - \beta)}\right)\log^+\left(\frac{\min\{L, 1\}}{\tilde{\phi}_0}\right). \tag{66}$$

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

$$\min\{L, 1\} > \tilde{\phi}_{K-2} = \frac{\tilde{\phi}_0}{(1 - \hat{t})^{(K-2)/2}} \ge \tilde{\phi}_0, \tag{67}$$

where the equality is due to (63). Taking logarithms on both sides and using the inequality $\log(1-\hat{t}) \le -\hat{t}$, we then conclude that

$$K \le 2 + \frac{2}{\hat{t}}\log\left(\frac{\min\{L, 1\}}{\tilde{\phi}_0}\right) \le 2 + \frac{2}{\hat{t}}\log^+\left(\frac{\min\{L, 1\}}{\tilde{\phi}_0}\right), \tag{68}$$

which clearly implies (66) due to the definition of $\hat{t}$ in (61). Now, using (67) and the first three relations in (61), we easily see that

$$\tilde{\phi}_0 = \min\left\{4\min\{L, 1\}, \frac{L\beta}{8\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{x}_0,\tilde{\nu}_0}}\right\} = \frac{L\beta}{8\tilde{\nu}_0^{3/2}\|F(\tilde{x}_0)\|^*_{\tilde{x}_0,\tilde{\nu}_0}}, \tag{69}$$

which together with (66) can be easily seen to imply (64). To complete the proof, it remains to show that (65) 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

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

where the last two inequalities follow from the fact that $\hat t \le 1/4$ and (68). Inequality (65) now follows from the last conclusion, relation (69), and the definition of $\hat t$ in (61). □

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.

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:*

i) a triple (x, v, ε) satisfying (43) in at most

$$\begin{aligned}&\mathcal{O}\left(\sqrt{\eta}\log^+\left(\frac{8\tilde\nu_0^{3/2}\|F(\tilde x_0)\|^*_{\tilde\nu_0,\tilde x_0}}{\max\{L,1\}}\right) + \sqrt{\eta}\right) \\ &+ \mathcal{O}\left(d_0^2\max\left\{\frac{\max\{L,1\}}{\bar\rho}, \frac{\eta^{1/6}\max\{L,1\}^{2/3}}{\bar\varepsilon^{2/3}}\right\}\right) \\ &+ \mathcal{O}\left(\sqrt{\eta}\max\left\{\log^+\left(\frac{\max\{L,1\}\sqrt{\eta}}{\tilde\nu_0\bar\rho}\right), \log^+\left(\frac{\max\{L,1\}\eta}{\tilde\nu_0^{3/2}\bar\varepsilon}\right)\right\}\right)\end{aligned} \tag{70}$$

iterations;

ii) a triple (x, v, ε) satisfying either (43) or (44) in at most

$$\begin{aligned}&\mathcal{O}\left(\sqrt{\eta}\log^+\left(\frac{8\tilde\nu_0^{3/2}\|F(\tilde x_0)\|^*_{\tilde\nu_0,\tilde x_0}}{\max\{L,1\}}\right) + \sqrt{\eta}\right) \\ &+ \mathcal{O}\left(d_0^{4/3}\eta^{1/6}\left(\max\{L,1\}\max\left\{\frac{1}{\bar\rho}, \frac{d_0}{\bar\varepsilon}\right\}\right)^{2/3}\right) \\ &+ \mathcal{O}\left(\sqrt{\eta}\max\left\{\log^+\left(\frac{\max\{L,1\}\sqrt{\eta}}{\tilde\nu_0\bar\rho}\right), \log^+\left(\frac{\max\{L,1\}\eta}{\tilde\nu_0^{3/2}\bar\varepsilon}\right)\right\}\right)\end{aligned}$$

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 (64) and that the number of iterations necessary for the HPE-IP method to compute a triple (x, v, ε) satisfying (43) is bounded by (55). The estimate (70) now follows from these two observations, relation (59) and estimate (65). (Note that the latter relation is needed due to the fact that (55) 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. □

The complexity estimates above depend on $\|F(\tilde{x}_0)\|^*_{\tilde{\nu}_0,z_0}$, where we recall that $\tilde{x}_0$ is a function of the input $(\tilde{\nu}_0, z_0)$. The next results shows how to bound this quantity in terms of this input.

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

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

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

Proof. It follows from (60) and Proposition 2.6 with $y = \tilde{x}_0$, $q = \nabla h(\tilde{x}_0)$ and $a = 0$ that

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

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

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

Hence, from the above two conclusions, we conclude that

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

The result now follows from the above relation, the fact that (11) and the triangle inequality for norms imply that

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

and the second relation in (21) of Lemma 3.2. □

A Technical lemmas

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

a) if $A - B \in \mathcal{S}^{\mathbf{E}}_+$, then $\|\cdot\|_A \geq \|\cdot\|_B$ and $\|\cdot\|^*_A \leq \|\cdot\|^*_B$;

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

$$\|u\|^*_A = \sqrt{\langle u, h\rangle} = \sqrt{\langle h, Ah\rangle} = \|h\|_A;$$

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

d) if $\{A_k\} \subseteq \mathcal{S}^{\mathbf{E}}_+$ is a sequence converging to A and such that the matrix $A_k - A \in \mathcal{S}^{\mathbf{E}}_+$ for every $k \in \mathcal{N}$, then

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

e) the function $\|\cdot\|_A^*$ *restricted to* $\mathcal{R}(A)$ *is a norm;*

f) for every $u \in \mathcal{R}(A)$ *and* $v \in \mathbf{E}$*, we have* $\langle u, v\rangle \le \|u\|_A^* \|v\|_A$.

Proof. (Proof of Lemma A.1.) a) The proof of this statement follows directly from definitions (2.3) and (4).

b) Assume first that $u \in \mathcal{R}(A)$, i.e. $u = Ah_u \in \mathcal{R}(A)$ for some $h_u \in \mathbf{E}$. Since h_u satisfies the first-order optimality condition of the maximization problem (4) and the objective function of this problem is concave, we conclude that h_u is an optimal solution of (4), and hence its optimal value $\|u\|_A^* < +\infty$. Assume now that $u \notin \mathcal{R}(A)$ and consider the decomposition $u = u_0 + u_r$ where $u_0 \in \mathcal{N}(A)$ and $u_r \in \mathcal{R}(A)$. Clearly, $\langle u_0, u_r\rangle = 0$ and $u_0 \neq 0$. In view of the definition of $\|u\|_A^*$ in (4), for every $t \in \Re$, the vector $h_t := tu_0$ satisfies

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

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 a) holds.

c) This statement follows directly from b) and (2.3).

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

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

is lower semi-continuous. This implies that

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

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

e) Choosing $A_k = A + (1/k)I$ for every $k \in \mathcal{N}$, 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 semi-norm on its domain $\mathcal{R}(A)$. Now, let $u \in \mathcal{R}(A)$ 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 $u \in \mathcal{R}(A)$ 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 $\mathcal{R}(A)$ 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 (4) evaluated at $\tilde{h}$ is equal to $[\langle u, v\rangle]/\|v\|_A]^2$ and the optimal value $(\|u\|_A^*)^2$ of (4) exceeds this value, we conclude that f) holds for this case too. □

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, Volume 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] 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.

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

[11] Y. Nesterov, *Introductory lectures on convex optimization*, vol. 87 of Applied Optimization, Kluwer Academic Publishers, Boston, MA, 2004. A basic course.

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

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

[14] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, vol. 13 of SIAM Studies in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

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

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

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

[18] 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 / Brasil 22460-320, March 2013.

[19] 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.

[20] ——, *A hybrid projection-proximal point algorithm*, J. Convex Anal., 6 (1999), pp. 59–70.

[21] ——, *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.

[22] ——, *A unified framework for some inexact proximal point algorithms*, Numer. Funct. Anal. Optim., 22 (2001), pp. 1013–1035.