An $\mathcal{O}(1/k^{3/2})$ Hybrid Proximal Extragradient Primal-Dual interior point method for non-linear monotone complementarity problems

Mauricio Romero Sicre* Benar F. Svaiter†

Abstract

We present a mixed Newton-type Hybrid Proximal Extragradient, primal-dual interior point method for solving smooth monotone complementarity problems. Dual variables for the non-negativity constraints are introduced. The ergodic complexity of the method is $\mathcal{O}(1/k^{3/2})$. The methods performs two types of iterations: under relaxes Hybrid Proximal-Extragradient iterations and short-steps primal-dual interior point iterations.

keywords. hybrid proximal extra-gradiente, interio point methods, primal-dual, monotone, complementarity problem, complexity.

AMS subject calssification. 90C60, 90C33, 90C30, 90C51, 47H05, 47J20, 65K10, 65K05, 49M15.

Introduction

The complexity of an algorithm is a theoretical estimation of its computational cost for obtaining a solution, exact or approximated, of the problem it is designed to solve. Although the ultimate test of an algorithm is practical performance, complexity analysis may point to improvements and even new methods. The practical performance may be estimated also by numerical tests. However, performance of an algorithms on numerical tests depends on the test suite, the actual implementation (coding), stopping criterion and parameter tuning; while mathematical estimation of theoretical complexity are verifiable (or proven false) by anyone able to follow the proofs and to find flaws (if any) in them.

Leonid G. Hačijan revolutionary work [8], the first polynomial algorithm for Linear Programming, changed radically this field and opened new roads in Linear Programming and beyond. A second revolutionary work it the one of Karmarkar [9], which introduced interior point methods in linear programing, and presented the first efficient polynomial algorithm for linear programming.

In non-linear programming, instead of exact solutions, algorithms produce approximate solutions, and a raw complexity measure is the number of iterations required for finding such approximate solutions. The *ergodic complexity* of an iterative algorithm estimates the quality (as approximated solution) of mean values of the iterates, computed with suitable weights; the *point-wise complexity* estimate the quality of the iterates,

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

†IMPA, Estrada Dona Castorina 110, CEP 22460-320, Rio de Janeiro, RJ, Brazil (email: benar@impa.br). The work of this author was partially supported by CNPq grants 302962/2011-5, 474944/2010-7, 480101/2008-6, 303583/2008-8, Faperj grants E-26/102.940/2011, E-26/102.821/2008 and PRONEX Optimization

In this work we study the complexity of a new method for solving smooth monotone non-linear complementarity problems. These problems are monotone variational inequalities where the feasible set is the positive orthant and the operator is smooth. In its turn, monotone variational inequality problems are particular instances of the problem of finding a zero of the sum of two monotone operators, where one of them is a normal cone, and the other is point-to point.

Any method for finding a zero of the sum of monotone operators, and in particular any method for solving monotone variational inequalities may be used for solving monotone non-linear complementarity problems. Korpelevič's method [10] solves monotone variational inequality problems for Lipschitz continuous operators, and requires in each iteration two projections onto the feasible set. Tseng's modified forward-backward method [26] finds the zero of a sum of two operators, one of which is Lipschitz, and the other one shall have an easily computable resolvent, which is computed once in each iteration. Douglas-Rachford method [5, 11] finds a zero of a sum of two monotone operator; it requires the computation of the resolvents of each of the operators; and, in practice, it appears as the classical alternating direction method of multipliers [7, 6].

The ergodic and the point-wise complexities of Korpelevič's method are, respectively, $\mathcal{O}(1/k)$ [17] and $\mathcal{O}(1/k^{1/2})$ [13]; likewise, the ergodic and point-wise complexities of Tseng's modified forward-backward method are, respectively, $\mathcal{O}(1/k)$ and $\mathcal{O}(1/k^{1/2})$ [13, 14]; while the ergodic complexity of Douglas-Rachford method is $\mathcal{O}(1/k)$ [16]. All these results can be obtained within the same theoretical framework [13, 14, 16], as fixed step-size implementations of the Hybrid Proximal Extragradiente Methods. Korpelevič's and Tseng's methods are "zero order" methods for solving monotone variational inequality problems for Lipschitz continuous monotone operators, in the sense that derivatives are neither used nor required by these methods.

Complexity estimations of first order methods (which use first derivatives) for smooth monotone variational inequalities were presented in [15]. In that work, in each iteration a "Newton-Josephy block-box" was used to solve, within a relative error tolerance, a prox-regularized variational problem in which the smooth operator was substitute by its linearization at the current iterate. However, the problem of how to implement such "Newton-Josephy black box" was not addressed. Moreover, in that work a binary search was used to compute the step-size, resulting in a multiplicative $\log\log$ term in the complexity estimation.

The present work is a sequel of [15] in two senses. First, the binary line search is replaced by an homotopy method, so eliminating the multiplicative $\log\log$ term in the complexity estimation. Second, for the smooth monotone complementarity problem, using a primal-dual interior-point like technique/method, we eliminate the "Newton-Josephy black-box", and require only a linear solver. The resulting algorithm, for the monotone linear complementarity problem, is a mixture of a relaxed version of a Newton-type HPE method [20, 24, 23] and a primal-dual interior point method.

This work is organized as follows. In section 1 we review some basic properties of maximal monotone and their enlargements. In Section 3 we derive the abstract iteration complexity of an under relaxed version of the Hybrid Proximal Extragradient Method, an further specialize these estimations to an under relaxed, large-step version of the Hybrid Proximal Extragradient method. In Section 4 In Section 7.

1 Maximal Monotone Operators and Their Enlargements

Maximal monotone operators and their enlargements will be used in the design an analysis of the new algorithm proposed in this paper. Here he review some basic properties of these objects

From now on X is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|$. A *point-to-set operator* $T : X \rightrightarrows X$ (or an operator in X) is a relation $T \subset X \times X$ and

$$T(x) = \{v \mid (x, v) \in T\}, \qquad T^{-1}(v) = \{x \mid (x, v) \in T\}, \qquad x, v \in X.$$

The *domain* and *range* of T are

$$D(T) = \{x \mid T(x) \neq \emptyset\}, \qquad R(T) = \{v \mid T^{-1}(v) \neq \emptyset\},$$

respectively. An operator $T : X \rightrightarrows X$ is *point-to-point* if $T(x)$ is a singleton or an empty set for any $x \in X$. We identify a point-to-point operator $T : X \rightrightarrows X$ with the unique function $F : D(T) \to X$ such that $T(x) = \{F(x)\}$ for any $x \in D(T)$.

A point-to-set operator $T : X \rightrightarrows X$ is *monotone* if

$$\langle x - y, v - u \rangle \geq 0 \qquad \forall (x, v), (y, u) \in T$$

and it is *maximal monotone* if it is monotone and maximal in the family of monotone operators in X. In $f : X \to \mathbb{R}$ is a proper, lower semicontinuous, convex function then the subdifferencial of $\partial f : X \rightrightarrows X$.

$$\partial f(x) = \{v \in X \mid f(y) \geq f(x) + \langle y - x, v \rangle\},$$

is a maximal montotone operator [18]. If $\Omega \subset X$ is a non-empty closed convex set then δ_Ω, the indicator function of $\Omega \subset X$, and $\mathbf{N}_\Omega$, the normal cone operator of Ω, are, respectivelly,

$$\delta_\Omega : X \to \overline{\mathbb{R}} \quad \delta_\Omega(x) = \begin{cases} 0, & x \in \Omega \\ \infty & \text{otherwise,} \end{cases} \qquad \mathbf{N}_\Omega : X \rightrightarrows X \quad \mathbf{N}_\Omega = \partial \delta_\Omega.$$

Since we are assuming Ω to be a non-empty closed convex set, δ_Ω is a proper, lower semicontinuous, convex function and $\mathbf{N}_\Omega$ is maximal monotone.

Recall that the ε-subdifferential [2], of a proper, lower semicontinuous, convex function $f : X \to \overline{\mathbb{R}}$ is the point-to-set operator $\partial_\varepsilon f : X \rightrightarrows X$,

$$\partial_\varepsilon f(x) = \{v \in X \mid f(y) \geq f(x) + \langle y - x, v \rangle\} \quad \varepsilon \geq 0, x \in X.$$

Altough the ε-subdifferential was origianly defined in abstract Banach spaces, it proven to be a very useful construct in optimization and convex analysis. It is natural to inquire wheter a similar construct exists for arbitrary maximal monotone operators. The anser to this question is our next topic.

The ε-enlargement [3, 4] of a maximal monotone $T : X \rightrightarrows X$ is the point-to-set operator $T^{[\varepsilon]} : X \rightrightarrows X$

$$T^{[\varepsilon]}(x) = \{v \in X \mid \langle x - y, v - u \rangle \geq -\varepsilon \qquad \forall y \in X,\ u \in T(y)\}, \qquad x \in X,\ \varepsilon \geq 0. \tag{1}$$

Observe that $T^{[\varepsilon]}$ is defined only for $\varepsilon \geq 0$. Direct use of this definition yields the following basic results (see [4, 25]).

Proposition 1.1. *Let $T : X \rightrightarrows X$, $T' : X \rightrightarrows X$ be maximal monotone.*

1. $T = T^{[\varepsilon=0]}$;

2. *if* $0 \le \varepsilon \le \eta$ *then* $T^{[\varepsilon]}(x) \subset T^{[\eta]}(x)$ *for any* $x \in X$*;*

3. *if* $\varepsilon, \eta \ge 0$ *then* $T^{[\varepsilon]}(x) + (T')^{[\eta]}(x) \subseteq (T + T')^{[\varepsilon+\eta]}(x)$ *for any* $x \in X$*;*

4. *if* $T = \partial f$*, where* f *is a proper closed convex function, then* $\partial_\varepsilon f(x) \subseteq T^{[\varepsilon]}(x)$ *for any* $\varepsilon \ge 0$ *and* $x \in X$*.*

The next propertie of the ε-enlargements of maximal montone operators is used for evaluating the ergodic complexity of the Hybrid Proximal Extragradient Method.

Theorem 1.2 (weak transportation formula [?])**.** *Let* $T : X \rightrightarrows X^*$ *be maximal monotone. Suppose that*

$$v_i \in T^{[\varepsilon_i]}(x_i), \qquad i = 1, \dots, m$$

and that $\alpha_1, \dots, \alpha_m \ge 0$*,* $\sum_{i=1}^m \alpha_i = 1$*. Define*

$$\bar{x} = \sum_{i=1}^m \alpha_i x_i, \qquad \bar{v} = \sum_{i=1}^m \alpha_i v_i, \qquad \bar{\varepsilon} = \sum_{i=1}^m \alpha_i(\varepsilon_i + \langle x_i - \bar{x}, v_i - \bar{v}\rangle).$$

Then $\bar{\varepsilon} \ge 0$ *and* $\bar{v} \in T^{[\bar{\varepsilon}]}(\bar{x})$*.*

2 Minty's path and its σ-neighborhoods

According to Minty's Theorem [12], if $T : X \rightrightarrows X$ is maximal monotone, then for any $\lambda > 0$, the operator $\lambda T + I$ is onto and $(\lambda T + I)^{-1}$, the *resolvent* of T, is a point-to-point operator with domain X. Observe that

$$x = (\lambda T + I)^{-1}(z) \iff \begin{cases} \exists v \in T(x) \\ \lambda v + x - z = 0\,. \end{cases}$$

It is convenient to regard such pair (x, v) as a function on λ.

Definition 2.1. *Let* $z \in X$ *and* $T : X \rightrightarrows X$ *be maximal monotone.*

1. Minty's system *for parameter* $\lambda > 0$*, base-point* z *and operator* T *is the inclusion/equation system*

$$x, v \in X, \qquad v \in T(x), \qquad \lambda v + x - z = 0; \tag{2}$$

2. Minty's path *with base-point* z *for operator* T *is the curve which associate to each* $\lambda > 0$ *the pair* (x_λ, v_λ) *unique solution of* (2)*; equivalently, it is the curve*

$$(0, \infty) \to X \times X, \qquad \lambda \mapsto (x_\lambda, v_\lambda) \qquad x_\lambda = (\lambda T + I)^{-1}(z), \qquad v_\lambda = \lambda^{-1}(z - x_\lambda). \tag{3}$$

Next we state some useful properties of Minty's path, which follows trivially from basic properties of the resolvent map [12, 1].

Proposition 2.2. *For any* $z \in X$ *and* $T : X \rightrightarrows X$ *maximal monotone, Minty's curve* $\lambda \mapsto (x_\lambda, v_\lambda)$*, with base-point* z *for operator* T*, is continuous,* $\|x_\lambda - z\|$ *is increasing,* $\|v_\lambda\|$ *is decreasing, and for any* $x^* \in T^{-1}(0)$*,*

$$\|x^* - x_\lambda\|^2 + \|x_\lambda - z\|^2 \le \|x^* - z\|^2.$$

Moreover, if $\Omega = T^{-1}(0) \neq \emptyset$, then

$$\|v_\lambda\| \leq \frac{d(z,\Omega)}{\lambda}, \quad \|x_\lambda - z\| \leq d(z,\Omega)$$

$\lim_{\lambda\to\infty} v_\lambda = 0$ and $\lim_{\lambda\to\infty} x_\lambda = P_\Omega(z)$ where P_Ω stands for the orthogonal projection onto Ω and these are strong limits.

Observe that v as specified in (2) (which is v_λ in (3)) is redundant. However, in *inexact* solutions of (2), where the equality does not hold, v ceases to be redundant and it is useful to bound its norm and its distance to v_λ. Indeed, Computation of x_λ and v_λ is, in general, quite expensive and we will consider inexact solutions of Minty's system (2) in which the inclusion *and* the equality are relaxed. We will relax inclusion in (2) by means of the the ε-enlargement.

From now on in this section $T : X \rightrightarrows X$ is a maximal monotone operator. We will use, in the analysis of inexact solutions of (2), the error measure

$$\Psi_{\lambda,z,T} : X \times X \times \mathbb{R} \to \mathbb{R}_+ \cup \{\infty\},$$
$$\Psi_{\lambda,z,T}(x,v,\varepsilon) = \begin{cases} \sqrt{\|\lambda v + x - z\|^2 + 2\lambda\varepsilon} & \text{if } \varepsilon \geq 0 \text{ and } v \in T^{[\varepsilon]}(x), \\ \infty & \text{otherwise.} \end{cases} \tag{4}$$

This error measure quantifies the distance to Minty's path, and also quantifies how much an ε-extragradient step is closer to the solution set $T^{-1}(0)$. These results where essentially proved in [20, 21, 22].

Proposition 2.3. *Let $\lambda > 0$ and $z \in X$. For any $x, v \in X$ and $\varepsilon \in \mathbb{R}$,*

$$\sqrt{\|\lambda(v - v_\lambda)\|^2 + \|x - x_\lambda\|^2} \leq \Psi_{\lambda,z,T}(x,v,\varepsilon)$$

where (x_λ, v_λ) is as in Definition 2.1; and

$$\|x^* - (z - \lambda v)\|^2 \leq \|x^* - z\|^2 + \Psi_{\lambda,z,T}(x,v,\varepsilon)^2 - \|x - z\|^2 \qquad \forall x^* \in T^{-1}(0).$$

Proof. To first inequality follows from definition (4) and [22, eqs. (11), (12), and Corollary 1]. The second inequality follows from definition (4) and [22, eqs. (11), (12), and Lemma 3]. □

The next corollary, provides a bound for the norms of v and $x - z$ in an inexact solution (x, v, ε) of (2), as functions of λ.

Corollary 2.4. *If $T^{-1}(0)$ is non-empty, d is the distance of $z \in X$ to this set, $\lambda > 0$, and $\Psi_{\lambda,z,T}(x,v,\varepsilon) \leq \rho < \infty$, then*

$$\|x - z\| \leq d + \rho, \qquad \|v\| \leq \frac{d + \rho}{\lambda},$$

and for any $x^* \in T^{-1}(0)$ and $t \in [0,1]$

$$\|x^* - z\|^2 \geq \|x^* - (z - t\lambda v)\|^2 + t(\|x - z\|^2 - \rho^2).$$

Proof. Let $\hat{x}$ be the projection of z in $T^{-1}(0)$ (which is a non-empty closed convex set). In view of Proposition 2.2,

$$\|x_\lambda - z\| \leq \|\hat{x} - z\| \leq d, \qquad \|v_\lambda\| = \frac{\|x_\lambda - z\|}{\lambda} \leq \frac{d}{\lambda}.$$

The two first inequalities of the corollary follows from the above inequalities and the first inequality in Proposition 2.3.

To prove the last inequality of the corollary, first observe that it holds trivially for $t = 0$. Use the last part of Proposition 2.3 to conclude that this inequality also holds for $t = 1$. To end the proof, observe that the its right-hand side is convex in t. □

We define a σ-neighborhood of the proximal path (3) for T at point z and parameter $\lambda > 0$ as the set

$$\mathcal{H}_\sigma(\lambda; z, T) := \left\{ (x, v, \varepsilon) \in X \times X \times [0, \infty) \;\middle|\; \begin{array}{l} v \in T^{[\varepsilon]}(x), \\ \|\lambda v + x - z\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|x - z\|^2 \end{array} \right\}, \tag{5}$$

that is, $\mathcal{H}_\sigma(\lambda; z, T) = \{(x, v, \varepsilon) \mid \Psi_{\lambda,z,T}(x, v, \varepsilon) \leq \sigma^2 \|x - z\|^2\}$.

Lemma 2.5. *If $(x, v, \varepsilon) \in \mathcal{H}_\sigma(\lambda; z, T)$ with $\lambda > 0$ and $0 \leq \sigma < 1$, then*

$$(1 - \sigma)\|x - z\| \leq \lambda\|v\| \leq (1 + \sigma)\|x - z\|, \qquad 2\lambda\varepsilon \leq \sigma^2 \|x - z\|^2, \tag{6}$$

$$\|x - x_\lambda\| \leq \sigma \|x - z\|, \tag{7}$$

$$\frac{\|x_\lambda - z\|}{1 + \sigma} \leq \|x - z\| \leq \frac{\|x_\lambda - z\|}{1 - \sigma}, \tag{8}$$

and for any $x^* \in T^{-1}(0)$

$$\|x - z\| \leq \frac{\|x^* - z\|}{1 - \sigma}, \qquad \|x^* - x\| \leq \frac{\|x^* - z\|}{1 - \sigma}. \tag{9}$$

Proof. The inequalities in (6) follow trivially from definition (5). Inequality (7) follows from definition (5) and Proposition 2.3. Direct use of triangle inequality yields

$$\|x - z\| - \|x - x_\lambda\| \leq \|x_\lambda - z\| \leq \|x - z\| + \|x - x_\lambda\|,$$

which, combined with (7) and the assumption $0 \leq \sigma < 1$, proves (8).

To end the proof, take $x^* \in T^{-1}(0)$. The first inequality in (9) follows from the second inequality in (8) and the inequality in Proposition 2.2. Direct use of triangle inequality, inequality in Proposition 2.2, and (7) yields

$$\|x^* - x\| \leq \|x^* - x_\lambda\| + \|x_\lambda - x\| \leq \|x^* - z\| + \sigma\|x - z\|,$$

which, combined with the first inequality in (9), proves the second one. □

The next proposition, proved in [20, 21, 24], shows how points in σ-neighborhoods $\mathcal{H}_\sigma(\lambda; z, T)$ of the proximal path can be used to generate points closer to the solution set than the base point z. This result has also been used in [13, 15].

Lemma 2.6. *Suppose that $(x, v, \varepsilon) \in \mathcal{H}_\sigma(\lambda; z, T)$, $\lambda > 0$, $0 \leq \sigma < 1$ and define*

$$z_t = z - t\lambda v.$$

Then, for any $x^* \in T^{-1}(0)$ and $t \in [0, 1]$,

$$\|x^* - z\|^2 \geq \|x^* - z_t\|^2 + t(1 - \sigma^2)\|x - z\|^2.$$

Proof. The result follows from Corollary 2.4 and definition (5). □

3 Under Relaxed Hybrid Proximal Extragradient

In this section we will present a complexity analysis of an under relaxed version of the hybrid proximal extragradient method. The hybrid proximal extragradient method was introduced in [20, 21], and under/over relaxed version of it where analyzed in a more general framework in [24]. The complexity analysis of the HPE was presented in [13]; here we adapt complexity results of [15] to the under relaxed version of the HPE.

Consider a sequence generated by the (under) relaxed hybrid proximal extragradient (RHPE) method for finding a zero of a maximal monotone operator $T : X \rightrightarrows X$. Start with $z_0 \in X$ and for $k = 1, 2, \ldots$

$$\begin{aligned} &v_k \in T^{[\varepsilon_k]}(x_k), \quad \|\lambda_k v_k + x_k - z_{k-1}\|^2 + 2\lambda_k\varepsilon_k \le \sigma_k^2\|x_k - z_{k-1}\|^2, \quad 0 \le \sigma_k < 1, \quad \lambda_k > 0,\\ &z_k = z_{k-1} - t_k\lambda_k v_k \quad 0 < t_k \le 1. \end{aligned} \tag{10}$$

In each iteration k, λ_k is the step-size, σ_k is a *relative* error tolerance, and t_k is a relaxation factor. Of course we could have also considered over-relaxed versions of the HPE. However, for our aims, the under relaxed version will do.

From now on in this section $T : X \rightrightarrows X$ is an arbitrary maximal monotone operator and (σ_k), (λ_k), (t_k), (z_k), (x_k), (v_k), (ε_k) are sequence generated by the RHPE method. The next proposition summarizes the basic convergence properties of the this method, and it is the key for deriving its "abstract" point-wise complexity.

Proposition 3.1. *For any $x^* \in T^{-1}(0)$,*

$$\begin{aligned} &\|x^* - z_{k-1}\|^2 \ge \|x^* - z_k\|^2 + t_k(1-\sigma_k^2)\|x_k - z_{k-1}\|^2,\\ &\|x^* - z_0\|^2 \ge \|x^* - z_k\|^2 + \sum_{i=1}^{k} t_i(1-\sigma_i^2)\|x_i - z_{i-1}\|^2, \qquad k = 1, 2, \ldots.\\ &\|x^* - z_0\| \ge \|x^* - z_k\|, \end{aligned} \tag{11}$$

Additionally, if $T^{-1}(0)$ is non empty and d is the distance of z_0 to this set, then

$$\sum_{i=1}^{k} t_i(1-\sigma_i^2)\|x_i - z_{i-1}\|^2 \le d^2, \quad k = 1, 2, \ldots. \tag{12}$$

Proof. The first inequality in (11) follows from (10), definition (5), and Lemma 2.6 with $z = z_{k-1}$, $\lambda = \lambda_k$, $\sigma = \sigma_k$, $x = x_k$, $v = v_k$, $\varepsilon = \varepsilon_k$, and $t = t_k$. The second inequality in (11) follows trivially from the first one while the third one follows from the second and the assumptions $t_k \ge 0$, $0 \le \sigma_k < 1$. The second inequality in (11), the assumption that $T^{-1}(0)$ is nonempty, and the definition of d trivially imply (12). □

Proposition 3.2. *For any k,*

$$\|v_k\| \le \frac{1+\sigma_k}{\lambda_k}\|x_k - z_{k-1}\|, \qquad \varepsilon_k \le \frac{\sigma^2}{2\lambda_k}\|x_k - z_{k-1}\|^2. \tag{13}$$

Moreover, if $T^{-1}(0)$ is nonempty and d is the distance of z_0 to this set, then

$$\|z_k - z_0\| \le 2d, \quad \|x_k - z_{k-1}\| \le \frac{d}{1-\sigma_k}, \quad \|x_k - z_0\| \le \left(\frac{1}{1-\sigma_k} + 1\right) d \tag{14}$$

for any k.

Proof. Take $k \in \mathbb{N}$. Using (10) and Lemma 2.5 with $z = z_{k-1}$, $\lambda = \lambda_k$, $x = x_k$, $v = v_k$, $\varepsilon = \varepsilon_k$, and $\sigma = \sigma_k$ we conclude that the two inequalities in (13) hold, and that for any $x^* \in T^{-1}(0)$

$$\|x_k - z_{k-1}\| \leq \frac{\|x^* - z_{k-1}\|}{1-\sigma_k}, \qquad \|x_k - z_0\| \leq \|x^* - x_k\| + \|x^* - z_0\| \leq \frac{\|x^* - z_{k-1}\|}{1-\sigma_k} + \|x^* - z_0\|,$$

where the second inequality follows from triangle inequality. Since we are assuming that $T^{-1}(0)$ is nonempty, we can take x^* as the projection of z_0 in this set. From Proposition 3.1 it follows that $\|x^* - z_{k-1}\| \leq \|x^* - z_0\|$. This inequality trivially implies the first inequality in (14) and, combined with the above inequalities, also implies the second and the third ones. □

Next we analyze the ergodic means associated with the RHPE method. Define, for each k,

$$\Lambda_k = \sum_{i=1}^k t_i\lambda_i, \qquad \bar{v}_k = \sum_{i=1}^k \frac{t_i\lambda_i}{\Lambda_k} v_i, \qquad \bar{x}_k = \sum_{i=1}^k \frac{t_i\lambda_i}{\Lambda_k} x_i, \qquad \bar{\varepsilon}_k = \sum_{i=1}^k \frac{t_i\lambda_i}{\Lambda_k}\left[\varepsilon_i + \langle x_i - \bar{x}_k, v_i\rangle\right]. \quad (15)$$

Proposition 3.3. *For any k*

$$\bar{v}_k \in T^{[\bar{\varepsilon}_k]}(\bar{x}_k), \qquad \bar{v}_k = \frac{1}{\Lambda_k}(z_0 - z_k), \qquad \bar{\varepsilon}_k \leq \frac{\|\bar{x}_k - z_0\|^2}{2\Lambda_k}. \quad (16)$$

Proof. The first relation in (16) follows from definitions (15), the inclusions in (10) and Theorem 1.2. The second relation in (16) follows from the definitions of Λ_k and $\bar{v}_k$ in (15) and the update rule for z_k in (10).

To estimate $\bar{\varepsilon}_k$, define

$$\Gamma_k : X \to \mathbb{R}, \;\; \Gamma_k(z) = \sum_{i=1}^k t_i\lambda_i[\langle z - x_i, v_i\rangle - \varepsilon_i], \quad \beta_k = \min \Gamma_k(z) + \frac{1}{2}\|z - z_0\|^2 \qquad k = 1, 2, \ldots$$

with $\Gamma_0 \equiv 0$ and $\beta_0 = 0$. Direct use of (10) shows that

$$\nabla\Gamma_k = z_0 - z_k, \quad z_k = \arg\min \Gamma_k(z) + \frac{1}{2}\|z - z_0\|^2, \quad \beta_k \geq \beta_{k-1} + t_k(1 - \sigma_k^2)\|x_k - z_{k-1}\|^2.$$

In particular, $\beta_k \geq 0$ for all k. Since Γ_k is affine and $\Gamma_k(\bar{x}_k) = -\Lambda_k\bar{\varepsilon}_k$,

$$\begin{aligned}\beta_k &= \frac{1}{2}\|z_k - z_0\|^2 + \langle z_k - \bar{x}_k, \nabla\Gamma_k\rangle - \Lambda_k\bar{\varepsilon}_k \\ &= \frac{1}{2}\|z_k - z_0\|^2 + \langle z_k - \bar{x}_k, z_0 - z_k\rangle - \Lambda_k\bar{\varepsilon}_k = \frac{1}{2}\|\bar{x}_k - z_0\|^2 - \frac{1}{2}\|z_k - \bar{x}_k\|^2 - \Lambda_k\bar{\varepsilon}_k\,.\end{aligned}$$

Therefore,

$$\Lambda_k\bar{\varepsilon}_k \leq \frac{1}{2}\|\bar{x}_k - z_0\|^2 - \beta_k \leq \frac{1}{2}\|\bar{x}_k - z_0\|^2$$

which concludes the proof. □

Next we derive the abstract complexity estimation of a "large step" relaxed HPE method, similar to the one defined and analyzed in [15]. Hee "large step" means that $\lambda_k\|x_k - z_{k-1}\|$ is bounded away from 0.

Proposition 3.4. *If $T^{-1}(0)$ is non-empty, d is the distance of z_0 to this set, and*

$$t_k \geq t > 0, \quad \sigma_k \leq \sigma < 1, \quad \lambda_k \|x_k - z_{k-1}\| \geq c > 0 \qquad k = 1, 2, \dots$$

then, for all k

$$\sum_{i=1}^{k} \|x_i - z_{i-1}\|^2 \leq \frac{d^2}{t(1-\sigma^2)} \tag{17}$$

and

1. *there exists $i_0 \leq k$ such that*

$$\|v_{i_0}\| \leq \frac{1}{ct} \frac{d^2}{(1-\sigma)k}, \qquad \varepsilon_{i_0} \leq \frac{1}{ct\sqrt{t}} \frac{\sigma^2 d^3}{2\sqrt{1-\sigma^2}^3} \frac{1}{k^{3/2}},$$

2. *$\bar{v}_k \in T^{[\bar{\varepsilon}_k]}$,*

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

Proof. Let $k \geq 1$. Inequality (17) follows the assumptions of the proposition and the last part of Proposition 3.1

To prove item 1, observe that direct use of (17) shows that there exists $i_0 \leq k$ such that

$$\|x_{i_0} - z_{i_0-1}\|^2 \leq \frac{d^2}{t(1-\sigma^2)k}.$$

Using the first part of Proposition 3.2 with $k = i_0$ and the assumptions of the proposition we conclude that

$$\|v_{i_0}\| \leq (1+\sigma)\frac{\|x_{i_0} - z_{i_0-1}\|}{\lambda_{i_0}} = (1+\sigma)\frac{\|x_{i_0} - z_{i_0-1}\|^2}{\lambda_{i_0}\|x_{i_0} - z_{i_0} - 1\|} \leq (1+\sigma)\frac{\|x_{i_0} - z_{i_0-1}\|^2}{c}$$

and

$$\varepsilon_{i_0} \leq \sigma^2 \frac{\|x_{i_0} - z_{i_0-1}\|^2}{2\lambda_{i_0}} = \sigma^2 \frac{\|x_{i_0} - z_{i_0-1}\|^3}{2\lambda_{i_0}\|x_{i_0} - z_{i_0-1}\|} \leq \sigma^2 \frac{\|x_{i_0} - z_{i_0-1}\|^3}{2c}.$$

The bounds on $\|v_{i_0}\|$ and ε_{i_0} follows trivially form the above inequalities.

To prove item 2, first use the definition of Λ_k and the assumptions of the proposition to conclude that

$$\Lambda_k = \sum_{i=1}^{k} t_i \lambda_i = \sum_{i=1}^{k} t_i \frac{\lambda_i \|x_i - z_{i-1}\|}{\|x_i - z_{i-1}\|} \geq ct \sum_{i=1}^{k} \frac{1}{\|x_i - z_{i-1}\|}.$$

From the above inequality, (17) and Lemma A.1 it follows that

$$\Lambda_k \geq ct\sqrt{\frac{t(1-\sigma^2)k^3}{d^2}}.$$

To end the proof, use Proposition 3.3 and the second part of Proposition 3.2 to conclude that $\bar{v}_k \in T^{[\bar{\varepsilon}_k]}(\bar{x}_k)$,

$$\|\bar{v}_k\| \leq \frac{2d}{\Lambda_k}, \qquad \bar{\varepsilon}_k \leq \frac{\|\bar{x}_k - x_0\|^2}{2\Lambda_k}.$$

From the assumptions of the proposition and the last inequality in Proposition 3.2 it follows that $\|x_i - z_0\| \leq 2d/(1-\sigma)$. Since $\bar{x}_k$ is an ergodic mean and the norm square is convex,

$$\|\bar{x}_k - z_0\| \leq \frac{2d}{1-\sigma}.$$

The bounds on $\|\bar{v}_k\|$ and $\bar{\varepsilon}_k$ follows from directly from the above inequalities. □

Proposition 3.5. *If $T^{-1}(0)$ is non-empty, d is the distance of z_0 to this set, and*

$$t_k\lambda_k\|x_k - z_{k-1}\| \geq a > 0, \quad \sigma_k \leq \sigma < 1, \quad \lambda_k\|x_k - z_{k-1}\| \geq c > 0 \qquad k = 1, 2, \ldots$$

then, for all k

$$\sum_{i=1}^{k} \frac{\|x_i - z_{i-1}\|}{\lambda_i} \leq \frac{d^2}{a(1-\sigma^2)} \tag{18}$$

and there exists $i_0 \leq k$ such that

$$\|v_{i_0}\| \leq \frac{1}{a(1-\sigma)}\frac{d^2}{k}, \qquad \varepsilon_{i_0} \leq \frac{\sigma^2}{2a(1-\sigma^2)(1-\sigma)}\frac{d^3}{k}.$$

Proof. Let $k \geq 1$. It follows from the assumptions of the Proposition that

$$t_i\|x_i - z_{i-1}\|^2 = (t_i\lambda_i\|x_i - z_{i-1}\|)\frac{\|x_i - z_{i-1}\|}{\lambda_i} \geq a\frac{\|x_i - z_{i-1}\|}{\lambda_i} \qquad i = 1, 2, \ldots,$$

which, combined with the last part of Proposition 3.1 and the assumptions on σ_i's yields (18).

To prove the last part of the proposition, observe that direct use of (18) shows that there exists $i_0 \leq k$ such that

$$\frac{\|x_{i_0} - z_{i_0-1}\|}{\lambda_{i_0}} \leq \frac{d^2}{a(1-\sigma^2)k}.$$

Using Proposition 3.2 with $k = i_0$ and the assumptions on σ_i's of the proposition we conclude that

$$\|v_{i_0}\| \leq (1+\sigma)\frac{\|x_{i_0} - z_{i_0-1}\|}{\lambda_{i_0}}, \quad \varepsilon_{i_0} \leq \sigma^2\frac{\|x_{i_0} - z_{i_0-1}\|^2}{2\lambda_{i_0}}, \quad \|x_{i_0} - z_{i_0-1}\| \leq \frac{d}{1-\sigma}.$$

The bounds on $\|v_{i_0}\|$ and ε_{i_0} follows trivially form the above inequalities. □

4 The smooth monotone complementarity problem and the associated proximal primal-dual system

In this section, first we formally define the smooth monotone complementarity problem; second we define a proximal primal-dual system associated with this problem; third we analyze Newton method for this system; and fourth we define a region where Newton method is quadratically convergent. From now on, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ stands for the set of non-negative and strictly positive real numbers, respectively.

In this work we consider the smooth, monotone, (mixed) complementarity problem, which is to find (x, y) such that

$$(x, y) \in \mathbb{R}^N \times \mathbb{R}^M_+, \qquad \langle (x', y') - (x, y), F(x, y)\rangle \geq 0 \ \ \forall (x', y') \in \mathbb{R}^N \times \mathbb{R}^M_+. \tag{19}$$

where

$$F : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}^N \times \mathbb{R}^M, \qquad F(x, y) = (F_1(x, y), F_2(x, y)) \tag{20}$$

satisfies the following assumptions

a.1) F is monotone in $\mathbb{R}^N \times \mathbb{R}^M_+$;
a.2) F is differentiable and, in $\mathbb{R}^N \times \mathbb{R}^M_+$, DF is L-Lipschitz continuous.

Observe that if $N = 0$, we retrieve the usual complementarity problem. The case $N > 0$ includes, in particular, a classical reformulation of the problem of minimizing a smooth convex function under linear equality *and* inequality constraints.

In view of assumptions a.1) and a.2), the point-to-set operator $F + \mathbf{N}_{\mathbb{R}^N \times \mathbb{R}^M_+} : \mathbb{R}^N \times \mathbb{R}^M \rightrightarrows \mathbb{R}^N \times \mathbb{R}^M$ is maximal monotone. Problem (19) is equivalent to

$$0 \in T(x, y), \qquad T = F + \mathbf{N}_{\mathbb{R}^N \times \mathbb{R}^M_+} \tag{21}$$

which is equivalent to the problem of finding $(x, y) \in \mathbb{R}^N \times \mathbb{R}^M$, $s \in \mathbb{R}^M$ such that

$$y, s \geq 0, \quad F_1(x, y) = 0, \; F_2(x, y) - s = 0, \quad \langle y, s \rangle = 0,$$

where z is an auxiliary variable. We will consider a proximal primal-dual system, which combines a proximal regularization of (21) with a primal dual system for the above system:

$$\begin{aligned} (x, y, s) \in \mathbb{R}^N \times \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}, \qquad \mu \left(F(x, y) - \begin{bmatrix} 0 \\ s \end{bmatrix} \right) + \nu \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} z^x \\ z^y \end{bmatrix} \right) &= 0, \\ \mu Y s - e &= 0, \end{aligned} \tag{22}$$

where $\mu, \nu > 0$ and $(z^x, z^y) \in \mathbb{R}^N \times \mathbb{R}^M$ are parameters and $e = \{1\}^M$. Existence and uniqueness of solutions of the above system, as well as the behavior of the solution as $\mu, \nu \to \infty$ were studied in [19]. The next result will be used in the analysis of exact and approximate solutions of this system.

Lemma 4.1. *If* $(x, y, s) \in \mathbb{R}^N \times \mathbb{R}^M_+ \times \mathbb{R}^M_+$ *and* $\varepsilon \geq \langle y, s \rangle$ *then*

$$-s \in \mathbf{N}^{[\varepsilon]}_{\mathbb{R}^M_+}(y), \quad -(0, s) \in \mathbf{N}^{[\varepsilon]}_{\mathbb{R}^N \times \mathbb{R}^M_+}(x, y)$$

and $F(x, y) - (0, s) \in (F + \mathbf{N}^{[\varepsilon]}_{\mathbb{R}^N \times \mathbb{R}^M_+})(x, y) \subset T^{[\varepsilon]}(x, y)$.

Proof. To prove the first inclusion, use definition (1) and observe that for any $y' \in \mathbb{R}^M_+$ and $s' \in \mathbf{N}_{\mathbb{R}^M_+}(y')$,

$$\langle y' - y, s' - s \rangle = \langle y' - y, s' \rangle + \langle y', s \rangle - \langle y, s \rangle \geq \langle y', s \rangle - \langle y, s \rangle \geq -\langle y, s \rangle$$

where the first inequality follows from the inclusions $s' \in \mathbf{N}_{\mathbb{R}^M_+}(y')$, $y \in \mathbb{R}^M_+$, and the second inequality from the inclusions $y', s \in \mathbb{R}^M_+$. The second inclusion is proved using the same reasoning, while the last inclusion follows from the second one and Proposition 1.1, item 3. □

Theorem 4.2. *For any $\mu, \nu > 0$ and $z = (z^x, z^y) \in \mathbb{R}^N \times \mathbb{R}^M$ system (22) has a unique solution*

$$(x, y, s) = (x(z, \mu, \nu), y(z, \mu, \nu), s(z, \mu, \nu)).$$

Moreover, if the solution set of (21) is non-empty, $(z_k)_{k\in\mathbb{N}}$ is bounded,

$$\mu_k \to \infty,\ \nu_k \to \infty,\ \mu_k/\nu_k \to \infty \qquad \text{as } k \to \infty$$

and

$$x_k = x(z_k, \mu_k, \nu_k), \qquad y_k = y(z_k, \mu_k, \nu_k), \quad s_k = s(z_k, \mu_k, \nu_k)$$

then (x_k, y_k) is bounded and all accumulation point of the sequence $((x_k, z_k))_{k\in\mathbb{N}}$ are solutions of (21).

Proof. Define

$$g_{z,\nu}(x, y) = \begin{cases} -\sum \log y_i + \dfrac{\nu}{2}\|(x, y) - z\|^2, & y > 0 \\ \infty & \text{otherwise.} \end{cases}$$

This function is proper, closed, strongly convex and differentiable in its effective domain $\mathbb{R}^N \times \mathbb{R}^M_{++}$. Therefore, $\mu F + \partial g_{z,\nu}$ is maximal monotone and as a unique zero. Since

$$(x, y, s) \text{ solves (22)} \iff 0 \in \mu F(x, y) + \partial g_{z,\nu}(x, y),\ s = \mu^{-1} y^{-1},$$

system (22) has a unique solution.

To prove the second part of the theorem, let (x_k, y_k, s_k) be as in (22) for $z = z_k$, $\mu = \mu_k$ and $\nu = \nu_k$; let d_k be the distance of z_k to the solution set. Define

$$\lambda_k = \mu_k/\nu_k, \quad \tilde{v}_k = F(x_k, y_k) - (0, s_k), \quad \tilde{\varepsilon}_k = n/\mu_k,$$

and let $T : \mathbb{R}^N \times \mathbb{R}^M \rightrightarrows \mathbb{R}^N \times \mathbb{R}^M$ be as in (21). Then $\tilde{v}_k \in T^{[\tilde{\varepsilon}_k]}(x_k, y_k)$ and $\lambda_k \tilde{v}_k + (x_k, y_k) - z_k = 0$. Therefore, defining

$$p_k = (\lambda_k T + I)^{-1}(z_k), \qquad v_k = \lambda_k^{-1}(z - p_k)$$

and using Proposition 2.3 we conclude that

$$\|(x_k, y_k) - p_k\| \le \sqrt{2n/\nu_k}, \qquad \|\tilde{v}_k - v_k\| \le \frac{\nu_k}{\mu_k}\sqrt{2n/\nu_k}$$

To end the proof, use the assumptions on (μ_k) and (ν_k), the above inequality and Proposition 2.2. □

To apply Newton's method to (22), define for $\mu, \nu > 0$, $(z^x, z^y) \in \mathbb{R}^N \times \mathbb{R}^M$,

$$\begin{aligned} &H_{(z^x,z^y),\mu,\nu} : \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^M, \\ &H_{(z^x,z^y),\mu,\nu}(x, y, s) = \begin{bmatrix} \mu F_1(x, y) & +\nu(x - z^x) \\ \mu(F_2(x, y) - s) & +\nu(y - z^y) \\ \mu Y s - e & \end{bmatrix}. \end{aligned} \tag{23}$$

Observe that (22) is equivalent to $y, s > 0$ and $H_{z,\mu,\nu}(x, y, s) = 0$ with $z = (z^x, z^y)$. Next we analyze a generic (under) relaxed Newton step for solving, in the variables (x, y, s), this non-linear system. Direct use of the above definition yields

$$DH_{z,\mu,\nu}(x, y, s) = \begin{bmatrix} \mu DF(x, y) + \nu I & \begin{matrix} 0 \\ -\mu I \end{matrix} \\ 0 \quad \mu S & \mu Y \end{bmatrix}$$

Newton step for (22) at (x, y, s) is $d = (d^x, d^y, d^s)$ solution of $DH_{z,\mu,\nu}(x, y, s)d = -H_{z,\mu,\nu}(x, y, s)$, that is,

$$\begin{bmatrix} \mu DF(x,y) + \nu I & & 0 \\ & & -\mu I \\ 0 & \mu S & \mu Y \end{bmatrix} \begin{bmatrix} d^x \\ d^y \\ d^s \end{bmatrix} = - \begin{bmatrix} \mu F_1(x,y) & +\nu(x - z^x) \\ \mu(F_2(x,y) - s) & +\nu(y - z^y) \\ \mu Y s - e & \end{bmatrix} \tag{24}$$

First we prove that Newton step is well defined at strictly feasible points and analyze the outcome of a Newton iteration.

Lemma 4.3. *Take $\mu, \nu > 0$, $z \in \mathbb{R}^N \times \mathbb{R}^M$, $(x, y, s) \in \mathbb{R}^N \times \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}$.*

1. *There exists a unique $d = (d^x, d^y, d^s)$ as in (24).*
2. *If (x_+, y_+, s_+) and (r^x_+, r^y_+, r^s_+) are the Newton iterate at point (x, y, s) and the corresponding residuals, respectively,*

$$(x_+, y_+, s_+) = (x, y, s) + (d^x, d^y, d^s), \quad (r^x_+, r^y_+, r^s_+) = H_{z,\mu,\nu}(x_+, y_+, s_+),$$

then

$$\|(r^x_+, r^y_+)\| \leq \frac{\mu L}{2}\|(d^x, d^y)\|^2, \quad \|r^s_+\| = \mu\|D^y d^s\|.$$

3. *If, additionally, $\|\mu Y s - e\| \leq \theta < 1$, and*

$$\frac{\|\mu(F(x,y) - (0,s)) + \nu((x,y) - z)\|^2}{2\nu} + \frac{\|\mu Y s - e\|^2}{2(1-\theta)} < 1$$

then $y_+, s_+ \in \mathbb{R}^M_{++}$,

$$\frac{\nu\|(d^x, d^y)\|^2}{2} + \mu\|D^y d^s\| \leq \frac{\|\mu(F(x,y) - (0,s)) + \nu((x,y) - z)\|^2}{2\nu} + \frac{\|\mu Y s - e\|^2}{2(1-\theta)} .$$

and

$$\frac{\|(r^x_+, r^y_+)\|}{\sqrt{2\nu}} + \frac{\|r^s_+\|}{\sqrt{2(1-\theta)}} \leq \gamma \left(\frac{\|\mu(F(x,y) - (0,s)) + \nu((x,y) - z)\|^2}{2\nu} + \frac{\|\mu Y s - e\|^2}{2(1-\theta)} \right),$$

where $\gamma = \max\left\{\mu L/\sqrt{2\nu^3}, 1/\sqrt{2(1-\theta)}\right\}$.

Proof. Since F is monotone in $\mathbb{R}^N \times \mathbb{R}^M_+$, $DF(x, y)$ is positive semidefinite. Let $(r^x, r^y, r^s) = H_{z,\mu,\nu}(x, y, s)$, that is,

$$r^x = \mu F_1(x,y) - \nu(x - z^x), \;\; r^y = \mu(F_2(x,y) - s) + \nu(y - z^y), \;\; r^s = \mu Y s - e.$$

Using Lemma B.1 with

$$A = \mu DF(x,y), \quad (b^x, b^y, b^s) = -(r^x, r^y, r^s)$$

we conclude that there exists a unique $d = (d^x, d^y, d^s)$ solution of (24), which proves item 1. Item 2 follows from the definition of d, as Newton step for (22) at (x, y, s), and assumption a.2).

To prove item 3, suppose that its assumptions hold. Then

$$\frac{\|(r^x, r^y)\|^2}{4\nu} + \frac{\|r^s\|^2}{2(1 - \|\mu Y s - e\|)} \leq \frac{\|(r^x, r^y)\|^2}{2\nu} + \frac{\|r^s\|^2}{2(1-\theta)} < 1.$$

Using Lemma B.1 (item 3 and 2) with A and (b^x, b^y, b^s) as above we conclude that $y_+, s_+ \in \mathbb{R}^M$ and the penultimate inequality in item 3 holds. Using item 2 and the definition of c we have

$$\begin{aligned}\frac{\|(r^x_+, r^y_+)\|}{\sqrt{2\nu}} + \frac{\|r^s_+\|}{\sqrt{2(1-\theta)}} &\leq \frac{\mu L}{2\sqrt{2\nu}}\|(d^x, d^y)\|^2 + \frac{\mu\|D^y d^s\|}{\sqrt{2(1-\theta)}} \\ &= \left(\frac{\mu L}{\sqrt{2\nu^3}}\right)\frac{\nu\|(d^x, d^y)\|^2}{2} + \left(\frac{1}{\sqrt{2(1-\theta)}}\right)\mu\|D^y d^s\|\end{aligned}$$

The last inequality on item 3 follows from the penultimate one, the above inequality and the definition of γ. □

In view of Lemma 4.3(item 3), it is convenient to consider the error measure Φ_θ of the proximal primal-dual system (22) and the associated neighborhoos $\mathcal{N}_\theta$ of the solution of (22),

$$\Phi_\theta(x, y, s; z, \mu, \nu) := \frac{\|\mu(F(x, y) - (0, s)) + \nu((x, y) - z)\|}{\sqrt{2\nu}} + \frac{\|\mu Y s - e\|}{\sqrt{2(1-\theta)}}, \tag{25}$$

$$\mathcal{N}_\theta(t; z, \mu, \nu) := \left\{(x, y, s) \in \mathbb{R}^N \times \mathbb{R}^M_{++} \times \mathbb{R}^M_{++} \mid \Phi_\theta(x, y, s; z, \mu, \nu) \leq t\right\}. \tag{26}$$

Now we are ready to characterize a region where where Newton method for (22) converges.

Theorem 4.4. *Take $z \in \mathbb{R}^N \times \mathbb{R}^M$, $\mu, \nu > 0$, $\theta, \beta \in (0, 1)$ and define*

$$\gamma = \max\left\{\frac{\mu L}{\sqrt{2\nu^3}}, \frac{1}{\sqrt{2(1-\theta)}}\right\}, \quad \kappa = \min\left\{\frac{\beta}{\gamma}, \frac{\theta}{\sqrt{2(1-\theta)}}\right\}.$$

If

$$(x, y, s) \in \mathcal{N}_\theta(\kappa; z, \mu, \nu)$$

then $d = (d^x, d^y, d^s)$ Newton step for (22) at (x, y, s) is well defined, $\gamma\Phi_\theta(x, y, s)^2 \leq \kappa\beta$ and

$$(x + d^x, y + d^y, s + d^s) \in \mathcal{N}_\theta(\gamma\Phi_\theta(x, y, s)^2; z, \mu, \nu) \subset \mathcal{N}_\theta(\kappa\beta; z, \mu, \nu).$$

Proof. Since $\gamma \geq 1/\sqrt{2(1-\theta)}$,

$$\kappa \leq \min\left\{\beta\sqrt{2(1-\theta)}, \frac{\theta}{\sqrt{2(1-\theta)}}\right\} < 1$$

Let $(r^x, r^y, r^s) = H_{z,\mu,\nu}(x, y, s)$. Then, in view of (25) and (26), $y, s > 0$,

$$\Phi_\theta(x, y, s; z, \mu, \nu) = \frac{\|(r^x, r^y)\|}{\sqrt{2\nu}} + \frac{\|r^s\|}{\sqrt{2(1-\theta)}} \leq \kappa, \quad \frac{\|(r^x, r^y)\|^2}{2\nu} + \frac{\|r^s\|^2}{2(1-\theta)} \leq \kappa^2 < 1$$

and $\|\mu Y s - e\| \leq \sqrt{2(1-\theta)}\kappa \leq \theta < 1$. Using these inequalities and Lemma 4.3(items 1 and 3) we conclude that (d^x, d^y, d^s) is well defined, $y + d^y > 0$, $s + d^s > 0$ and

$$\Phi_\theta(x + d^x, y + d^y, s + d^s; z, \mu, \nu) \leq \gamma\Phi_\theta(x, y, s; z, \mu, \nu)^2 \leq \gamma\kappa\Phi_\theta(x, y, s; z, \mu, \nu).$$

To end the proof, note that $\gamma\kappa \leq \beta$. □

Observe that we just proved that Newton's method for (22) is quadratically convergent (to the unique solution of this system) in the region $\mathcal{N}_\theta(\kappa; z, \mu, \nu)$, with κ as in Theorem 4.4. Moreover, a single Newton iteration at a point in this region generates a point in the region $\mathcal{N}_\theta(\kappa\beta; z, \mu, \nu)$. These regions will play the role of "outer" and "inner" neighborhoods of the solution of (22), for updating the parameters z, μ, ν.

5 Updating the parameters z, μ and ν

In this section we discuss to possible updates for the parameters z, μ, ν at a point in $\mathcal{N}_\theta(\kappa\beta; z, \mu, \nu)$ so that this point remains in the updated region $\mathcal{N}_\theta(\kappa; z_+, \mu_+, \nu_+)$ and the ratio $\mu\nu^{-3/2}$ remains constant. We show that either the scalars μ, ν can be increased, or decreased while performing a relaxed hybrid proximal-extragradient step in z.

We will further restrict the parameters θ, β used in Theorem 4.4. From now on in this section

$$
\begin{aligned}
&z \in \mathbb{R}^N \times \mathbb{R}^M, \quad \mu, \nu > 0, \quad \theta, \beta \in (0, 1/2];\\
&\gamma = \max\left\{\frac{\mu L}{\sqrt{2\nu^3}}, \frac{1}{\sqrt{2(1-\theta)}}\right\}; \quad \kappa = \min\left\{\frac{\beta}{\gamma}, \frac{\theta}{\sqrt{2(1-\theta)}}\right\},\\
&h = \frac{\kappa(1-\beta)}{6(\sqrt{n} + 1/2)}.
\end{aligned} \tag{27}
$$

Since $\beta, \theta \leq 1/2$,

$$
\kappa < \frac{1}{2}, \quad h < \frac{1}{12}, \quad \frac{h}{1+h} \geq \frac{\kappa(1-\beta)}{7(\sqrt{n} + 1/2)}. \tag{28}
$$

First we determine under which conditions the parameters μ, ν can be increased

Lemma 5.1. *Suppose that $z, \mu, \nu, \theta, \beta$ and κ are as in* (27),

$$
(x, y, s) \in \mathcal{N}_\theta(\,\kappa\beta\,; z, \mu, \nu) \quad \nu\|(x, y) - z\|^2 \leq 8(\sqrt{n} + 1/2)^2
$$

and define

$$
\mu_+ = (1+h)^3\mu, \quad \nu_+ = (1+h)^2\nu\,.
$$

Then

$$
\|F(x, y) - (0, s)\| \leq \frac{2\sqrt{2\nu}}{\mu}(\sqrt{n} + 1), \quad \langle y, s\rangle \leq \frac{\sqrt{n}(\sqrt{n} + 1/2)}{\mu}\,.
$$

and $(x, y, s) \in \mathcal{N}_\theta(\kappa; z, \mu_+, \nu_+)$.

Proof. To simplify the proof, let

$$
r_1 = \mu(F(x, y) - (0, s)) + \nu((x, y) - z), \qquad r_2 = \mu Y s - e.
$$

Observe that $\Phi_\theta(x, y, s; z, \mu, \nu) = \|r_1\|/\sqrt{2\nu} + \|r_2\|/\sqrt{2(1-\theta)} \leq \kappa\beta$,

$$
F(x, y) - s = \mu^{-1}(r_1 - \nu((x, y) - z)), \qquad \langle y, s\rangle = \mu^{-1}(\langle r_2, e\rangle + n).
$$

Therefore, $\|r_1\| \leq \kappa\beta\sqrt{2\nu}$, $\|r_2\| \leq \kappa\beta\sqrt{2(1-\theta)}$,

$$\begin{aligned}
\|F(x,y)-(0,s)\| &\leq \frac{\|r_1\| + \|\nu((x,y)-z)\|}{\mu} \\
&\leq \frac{\kappa\beta\sqrt{2\nu} + \sqrt{\nu}\sqrt{8}(\sqrt{n}+1/2)}{\mu} = \frac{\sqrt{2\nu}}{\mu}(\kappa\beta + 1 + 2\sqrt{n}), \\
\langle y, s\rangle &\leq \frac{\|r_2\|\sqrt{n} + n}{\mu} \leq \frac{\kappa\beta\sqrt{2(1-\theta)}\sqrt{n} + n}{\mu}.
\end{aligned}$$

The two inequalities of the lemma follows from the above inequalities and (27).

Let

$$r_{1+} = \mu_+(F(x,y)-(0,s)) + \nu_+((x,y)-z), \qquad r_{2+} = \mu_+ Ys - e, \quad \tau = (1+h).$$

Since $\mu_+ = \tau^3\mu$ and $\nu_+ = \tau^2\nu$,

$$r_{1+} = \tau^3 r_1 + (\tau^2 - \tau^3)(\nu((x,y)-z), \qquad r_{2+} = \tau^3 r_2 + (\tau^3 - 1)e$$

Therefore

$$\begin{aligned}
\frac{\|r_{1+}\|}{\sqrt{2\nu_+}} &\leq \frac{\tau^2\|r_1\| + |\tau - \tau^2|\nu\|(x,y)-z\|}{\sqrt{2\nu}} \leq \frac{\tau^3\|r_1\|}{\sqrt{2\nu}} + (\tau^2-\tau)2(\sqrt{n}+1/2), \\
\frac{\|r_{2+}\|}{\sqrt{2(1-\theta)}} &\leq \frac{\tau^3\|r_2\| + (\tau^3-1)\sqrt{n}}{\sqrt{2(1-\theta)}}.
\end{aligned}$$

Adding these inequalities and using definition (25) we conclude that

$$\begin{aligned}
\Phi_\theta(x,y,s;z,\mu_+,\nu_+) &= \frac{\|r_{1+}\|}{\sqrt{2\nu_+}} + \frac{\|r_{2+}\|}{\sqrt{2(1-\theta)}} \\
&\leq \tau^3\left(\frac{\|r_1\|}{\sqrt{2\nu}} + \frac{\|r_2\|}{\sqrt{2(1-\theta)}}\right) + 2(\tau^2-\tau)(\sqrt{n}+1/2) + (\tau^3-1)\sqrt{n} \\
&\leq \tau^3\kappa\beta + 2(\tau^2-\tau)(\sqrt{n}+1/2) + (\tau^3-1)\sqrt{n}
\end{aligned}$$

Whence

$$\begin{aligned}
\Phi_\theta(x,y,s;z,\mu_+,\nu_+) &\leq \kappa\beta + 2(\tau^2-\tau)(\sqrt{n}+1/2) + (\tau^3-1)(\sqrt{n}+\kappa\beta) \\
&\leq \kappa\beta + (2(\tau^2-\tau) + \tau^3 - 1)(\sqrt{n}+1/2) \\
&= \kappa\beta + (h^2+5h+5)h(\sqrt{n}+1/2) \\
&= \kappa\beta + (h^2+5h+5)\frac{\kappa(1-\beta)}{6}\ .
\end{aligned}$$

To end the proof of the last inclusion, observe that $h < 1/12$, use definition (26), and observe that $y, s > 0$. □

Lemma 5.2. *Suppose that $z, \mu, \nu, \theta, \beta$ and κ are as in (27), with $\theta, \beta \in (0, 1/2]$,*

$$(x,y,s) \in \mathcal{N}_\theta(\,\kappa\beta\,; z, \mu, \nu), \quad \nu\|(x,y)-z\|^2 \geq 8(\sqrt{n}+1/2)^2$$

and define

$$\lambda = \mu/\nu, \qquad v = F(x,y) - (0,s), \qquad \varepsilon = \langle y, s\rangle,$$
$$z_+ = z - \frac{h}{1+h}\lambda v, \qquad \mu_+ = \frac{\mu}{(1+h)^3}, \qquad \nu_+ = \frac{\nu}{(1+h)^2}.$$

then

$$v \in (F + \mathbf{N}^{[\varepsilon]}_{\mathbb{R}^N\times\mathbb{R}^M_{++}})(x,y) \subset (F + \mathbf{N}_{\mathbb{R}^N\times\mathbb{R}^M_{++}})^{[\varepsilon]}(x,y),$$
$$\|\lambda v + (x,y) - z\|^2 + 2\lambda\varepsilon \le (1/4)\|(x,y) - z\|^2.$$

and $(x,y,s) \in \mathcal{N}_\theta(\kappa; z_+, \mu_+, \nu_+)$.

Proof. Observe that $y, s > 0$. The first two inclusions follows from Lemma 4.1. Define, again,

$$r_1 = \mu(F(x,y) - (0,s)) + \nu((x,y) - z), \qquad r_2 = \mu Y s - e.$$

Observe that $\Phi_\theta(x,y,s;z,\mu,\nu) = \|r_1\|/\sqrt{2\nu} + \|r_2\|/\sqrt{2(1-\theta)} \le \kappa\beta$,

$$\lambda v + (x,y) - z = \nu^{-1} r_1, \qquad \langle y, s\rangle = \mu^{-1}(\langle r_2, e\rangle + n).$$

Therefore,

$$\|\lambda v + (x,y) - z\|^2 + 2\lambda\varepsilon = \frac{\|r_1\|^2}{\nu^2} + \frac{2}{\nu}(\langle r_2, e\rangle + n) \le \frac{2}{\nu}\left[\frac{\|r_1\|^2}{2\nu} + \|r_2\|\sqrt{n} + n\right]$$
$$\le \frac{2}{\nu}[(\kappa\beta)^2 + \kappa\beta\sqrt{2(1-\theta)}\sqrt{n} + n] \le \frac{2}{\nu}(\sqrt{n} + 1/2)^2$$

where the last inequality follows from (27). The first inequality of the lemma follows trivially from its assumptions and the above inequality.

Let

$$r_{1+} = \mu_+(F(x,y) - (0,s)) + \nu_+((x,y) - z_+), \qquad r_{2+} = \mu_+ Y s - e, \quad \tau = (1+h).$$

Direct use of the definitions of z_+, μ_+, ν_+ yields

$$r_{1+} = \tau^{-2} r_1, \qquad r_{2+} = \tau^{-3} r_2 + (\tau^{-3} - 1)e.$$

Therefore

$$\frac{\|r_{1+}\|}{\sqrt{2\nu_+}} = \frac{\|r_1\|}{\tau\sqrt{2\nu}} \le \frac{\|r_1\|}{\sqrt{2\nu}}, \qquad \frac{\|r_{2+}\|}{\sqrt{2(1-\theta)}} \le \frac{\tau^{-3}\|r_2\| + \left|\tau^{-3} - 1\right|\sqrt{n}}{\sqrt{2(1-\theta)}} \le \frac{\|r_2\| + (1-\tau^{-3})\sqrt{n}}{\sqrt{2(1-\theta)}}$$

Since $t \mapsto 1/t^3$ is convex for $t \ge 0$, $1 - \tau^{-3} \le 3h$. Using this inequality, the two above inequalities, definition (25), and (27) we conclude that

$$\Phi_\theta(x,y,s;z_+,\mu_+,\nu_+) = \frac{\|r_{1+}\|}{\sqrt{2\nu_+}} + \frac{\|r_{2+}\|}{\sqrt{2(1-\theta)}}$$
$$\le \left(\frac{\|r_1\|}{\sqrt{2\nu}} + \frac{\|r_2\|}{\sqrt{2(1-\theta)}}\right) + \frac{3h\sqrt{n}}{\sqrt{2(1-\theta)}} \le \kappa\beta + \frac{3\kappa(1-\beta)}{7\sqrt{2(1-\theta)}} \le \kappa.$$

To end the proof, recall that $y, s > 0$. □

6 An Hybrid Proximal Extragradient Primal-Dual interior point method

We will present an algorithm which combines: first, Newton steps in variables x, y and s aimed at solving (22); second, short step interior-point iterations for increasing μ and ν in (22); and third relaxed hybrid extragradient proximal steps in variable $z = (z^x, z^y)$ for the inclusion problem (21).

Algorithm 1

Initialization Choose $\theta, \beta \in (0, 1/2]$, and compute $x_0 \in \mathbb{R}^N$, $y_0, s_0 \in \mathbb{R}^M_{++}$, $z_0 \in \mathbb{R}^N \times \mathbb{R}^M$, $\mu_0, \nu_0 > 0$, such that

$$\Phi_\theta(x_0, y_0, s_0; z_0, \mu_0, \nu_0) \leq \kappa$$

where

$$\gamma = \max\left\{\frac{\mu_0 L}{\sqrt{2\nu_0^3}}, \frac{1}{\sqrt{2(1-\theta)}}\right\}, \qquad \kappa = \min\left\{\frac{\beta}{\gamma}, \frac{\theta}{\sqrt{2(1-\theta)}}\right\}, \qquad h = \frac{\kappa(1-\beta)}{6(\sqrt{n}+1/2)},$$

iterations For $k = 1, 2, \ldots$

0) Set

$$x = x_{k-1}, \quad y = y_{k-1}, \quad s = s_{k-1}, \quad z = z_{k-1}, \quad \mu = \mu_{k-1}, \quad \nu = \nu_{k-1}$$

1) Compute (d_k^x, d_k^y, d_k^s) solution of

$$\begin{bmatrix} \mu DF(x,y) + \nu I & \begin{matrix} 0 \\ -\mu I \end{matrix} \\ 0 \quad \mu S & \mu Y \end{bmatrix} \begin{bmatrix} d_k^x \\ d_k^y \\ d_k^s \end{bmatrix} = -H_{z,\mu,\nu}(x, y, s) \tag{29}$$

and set $x_k = x + d_k^x$, $y_k = y + d_k^y$, $s_k = s + d_k^s$.

2.a) if $\nu\|(x_k, y_k) - z\|^2 \leq 8(\sqrt{n} + 1/2)^2$ **then**

$$\mu_k = (1+h)^3\mu, \quad \nu_k = (1+h)^2\nu, \qquad z_k = z,$$

2.b) else $(\nu\|(x_k, y_k) - z\|^2 > 8(\sqrt{n} + 1/2)^2)$

$$\mu_k = \frac{\mu}{(1+h)^3}, \quad \nu_k = \frac{\nu}{(1+h)^2}, \qquad z_k = z - \frac{h}{(1+h)}\frac{\mu}{\nu}(F(x_k, y_k) - (0, s_k)).$$

Some remarks about the above algorithm are in order.

i) The unique purpose of step **0)** is to simplify the notation in the ensuing steps.

ii) Well definiteness of iteration k depends on the existence and unicity of a solution of the linear system in step **1)**.

iv) At it iteration k, the input is x_{k-1}, y_{k-1}, s_{k-1}, z_{k-1}, μ_{k-1}, ν_{k-1}; the output is x_k, y_k, s_k, z_k, μ_k, ν_k and either step **2.a)** or step **2.b)** is executed.

v) The computational burden of the algorithm is at step **1)**, where a Jacobian shall be computed and a linear system shall be solved.

vi) In view of the update rule for μ_k and ν_k in step **2)**

$$\left(\frac{\nu_k}{\nu_0}\right)^3 = \left(\frac{\mu_k}{\mu_0}\right)^2 \qquad k = 0, 1, \ldots \tag{30}$$

vi) the computational cost of step **2)** is negligible;

vii) Since the initialization of this algorithm impacts in its computational complexity, we leave the discussion of this phase (initialization) to the end of this section and Section 7.

Form now on (x_k), (y_k), (s_k), (z_k), (μ_k) and (ν_k) are the sequences generated by Algorithm I. In principle these sequences may be well defined only up to some k. An abnormal termination would happen if, for some k: the linear system in (29) were singular; or, if t_k as specified in step **2.a)** did not exist.

Proposition 6.1. *Algorithm 3 is well defined and for any k*

$$(x_k, y_k, s_k) \in \mathcal{N}_\theta(\kappa; z_k, \mu_k, \nu_k), \quad (x_{k+1}, y_{k+1}, s_{k+1}) \in \mathcal{N}_\theta(\kappa\beta; z_k, \mu_k, \nu_k). \tag{31}$$

In particular, $y_k, s_k > 0$ for all k.

Proof. We will use induction on k. Suppose that
i) all iterations $k < m$ are well defined and
ii) the *first* inclusion in (31) holds for $k = m - 1$.
It follows from Theorem 4.4 and the definition of step **1)** that, in iteration m, this step is well defined and that

$$(x_m, y_m, s_m) \in \mathcal{N}_\theta(\kappa\beta; z_{m-1}, \mu_{m-1}, \nu_{m-1})$$

Hence iteration m is well defined and the second inclusion in (31) holds for $k = m - 1$.

Next we analyze step **2)** at iteration m. First suppose that

$$\nu_{m-1}\|(x_m, y_m) - z_{m-1}\|^2 \leq 8(\sqrt{n} + 1/2).$$

In this case, step **2.a)** is to be performed and it follows from the above inclusion and Lemma 5.1 that the first inclusion in (31) holds for $k = m$.

If the above inequality does not hold, step **2.b)** is performed. In this case, it follows from the above inclusion and from Lemma 5.2 that the first inclusion in (31) holds for $k = m$. This concludes the induction proof and proves the first part of the proposition. The last part of the proposition follows from the firs one and definition (26). □

We need to count the number of steps **2.a)** and **2.b)** executed up to iteration k. Define

$$\begin{aligned} A_k &= \{i \leq k \mid \text{at iteration } i, \text{ step } \textbf{2.a)} \text{ is executed}\}, \\ B_k &= \{i \leq k \mid \text{at iteration } i, \text{ step } \textbf{2.b)} \text{ is executed}\}. \end{aligned} \tag{32}$$

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

Lemma 6.2. *If, in iteration k, step* **2.a)** *is executed and $\#A_k - \#B_k = m+1$, then*

$$\|F(x_k, y_k) - (0, s_k)\| \le \frac{2\sqrt{2\nu_0}}{\mu_0(1+h)^{2m}}(\sqrt{n}+1), \quad \langle y_k, s_k\rangle \le \frac{1}{\mu_0(1+h)^{3m}}\sqrt{n}\left(\sqrt{n}+1/2\right).$$

Proof. First use the defintion of step **2.a)** and Lemma 5.1 to conclude that

$$\|F(x_k, y_k) - (0, s_k)\| \le \frac{2\sqrt{2\nu_{k-1}}}{\mu_{k-1}}(\sqrt{n}+1), \quad \langle y_k, s_k\rangle \le \frac{\sqrt{n}(\sqrt{n}+1/2)}{\mu_{k-1}}.$$

In view of the definition of step **2)**,

$$\nu_j = \nu_0(1+h)^{2(\#A_j - \#B_j)}, \qquad \mu_j = \mu_0(1+h)^{3(\#A_j - \#B_j)}.$$

Since we are assuming that at iteration k, step **2.a)** is executed,

$$\#A_{k-1} - \#B_{k-1} = m.$$

wich, combined with the previous equalities evaluated at $j = k-1$ and the above inequalities proves the lemma. □

In the next proposition we will show that iterations in which step **2.b)** is executed can be regarded as "large stpe" RHPE method iterations in the sequence (z_k) with relaxation parameter $h/(1+h)$. Define

Proposition 6.3. *Suppose that at iteration k step* **2.b)** *is executed and define*

$$v_k = F(x_k, y_k) - (0, s_k), \qquad \varepsilon_k = \langle y_k, s_k\rangle, \qquad \lambda_k = \mu_{k-1}/\nu_{k-1}. \tag{33}$$

Then

$$\begin{aligned}
&v_k \in (F + \mathbf{N}_{\mathbb{R}^N \times \mathbb{R}^m_+}{}^{[\varepsilon_k]})(x_k, y_k) \subseteq (F + \mathbf{N}_{\mathbb{R}^N \times \mathbb{R}^M_+})^{[\varepsilon_k]}(x_k, y_k),\\
&\|\lambda_k v_k + (x_k, y_k) - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \le (1/2)^2 \|(x_k, y_k) - z_{k-1}\|^2,\\
&z_k = z_{k-1} - (h/(1+h))\lambda_k v_k,\\
&\lambda_k \|(x_k, y_k) - z_{k-1}\| \ge \frac{\mu_0}{\sqrt{\nu_0^3}} 2\sqrt{2}(\sqrt{n}+1/2).
\end{aligned}$$

Proof. The inclusions and the first inequality follow from Lemma 5.2. The expression for z_k follows trivially from the definition of step **2.a)** and the one of λ_k.

Since we assumed that at iteration k step **2.b)** is executed,

$$\|(x_k, y_k) - z_k\|^2 \ge \frac{8}{\nu_{k-1}}(\sqrt{n}+1/2)^2.$$

To prove the last inequality, multiply the above inequality by λ_k^2 and use (30). □

To use Proposition 6.3, define

$$c = \frac{\mu_0}{\sqrt{\nu_0^3}} 2\sqrt{2}(\sqrt{n}+1/2), \quad t = \frac{h}{1+h}, \quad \lambda_k = \frac{\mu_{k-1}}{\nu_{k-1}} \quad k = 0, 1, \ldots \tag{34}$$

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

$$\Lambda_k = \sum_{i\in B_k} t\lambda_i, \qquad \bar{x}_k = \sum_{i\in B_k} \frac{t\lambda_i}{\Lambda_k} x_i, \;\; \bar{y}_k = \sum_{i\in B_k} \frac{t\lambda_i}{\Lambda_k} y_i, \qquad \bar{v}_k = \sum_{i\in B_k} \frac{t\lambda_i}{\Lambda_k}(F(x_i,y_i) - (0,s_i)), \tag{35}$$

$$\bar{\varepsilon}_k = \sum_{i\in B_k} \frac{t\lambda_i}{\Lambda_k}[\langle y_i, s_i\rangle + \langle (x_i,y_i) - (\bar{x}_k,\bar{y}_k), F(x_i,y_i) - (0,s_i) - \bar{v}_k\rangle]. \tag{36}$$

Lemma 6.4. *Suppose that the solution set of* (19) *is nonempty, and let d be the distance of z_0 to this set. If $\#B_k = m \geq 1$ then*

1. *there is an $i \leq k$ such that*

$$\|F(x_i,y_i) - (0,s_i)\| \leq \frac{7}{\sqrt{2}} \frac{\sqrt{\nu_0^3}}{\mu_0\kappa(1-\beta)} \frac{d^2}{m}, \qquad \langle y_i, s_i\rangle \leq \frac{7\sqrt{7}}{6\sqrt{6}} \frac{\sqrt{\nu_0^3}\sqrt{\sqrt{n}+1/2}}{\mu_0\sqrt{\kappa^3(1-\beta)^3}} \frac{d^3}{m^{3/2}};$$

2. $\bar{v}_k \in (F + \mathbf{N}_{\mathbb{R}^N\times\mathbb{R}^M_+})^{[\bar{\varepsilon}_k]}(\bar{x}_k,\bar{y}_k)$,

$$\|\bar{v}_k\| \leq \frac{7\sqrt{2*7}}{\sqrt{3}} \frac{\sqrt{\nu_0^3}\sqrt{\sqrt{n}+1/2}}{\mu_0\sqrt{\kappa^3(1-\beta)^3}} \frac{d^2}{m^{3/2}}, \qquad \bar{\varepsilon}_k \leq \frac{4*7\sqrt{2*7}}{\sqrt{3}} \frac{\sqrt{\nu_0^3}\sqrt{\sqrt{n}+1/2}}{\mu_0\sqrt{\kappa^3(1-\beta)^3}} \frac{d^3}{m^{3/2}}.$$

Proof. Let c, t be as in (34) and define. Direct use of the last inequality in (28) yields

$$ct \geq \frac{2\sqrt{2}}{7} \frac{\mu_0\kappa(1-\beta)}{\nu_0^{3/2}}, \qquad ct\sqrt{t} \geq \frac{2\sqrt{2}}{7} \frac{\mu_0[\kappa(1-\beta)]^{3/2}}{\nu_0^{3/2}\sqrt{\sqrt{n}+1/2}} \tag{37}$$

Let $k_1 < k_2 < \cdots < k_m$ be the m elements of B_k and define

$$\mathbf{z}_0 = z_0,$$
$$\mathbf{z}_i = z_{k_i}, \quad \mathbf{x}_i = (x_{k_i}, y_{k_i}) \qquad \mathbf{v}_i = F(x_{k_i}, y_{k_i}) - (0, s_{k_i}), \quad \boldsymbol{\epsilon}_i = \langle y_{k_i}, s_{k_i}\rangle, \quad \boldsymbol{\lambda}_i = \lambda_{k_i}, \quad i = 1,\ldots,m$$

Define also

$$\boldsymbol{\Lambda}_m = \sum_{i=1}^m \boldsymbol{\lambda}_i, \;\; \bar{\mathbf{x}}_m = \sum_{i=1}^m \frac{t\boldsymbol{\lambda}_i}{\boldsymbol{\Lambda}_m}\mathbf{x}_i, \;\; \bar{\mathbf{v}}_m = \sum_{i=1}^m \frac{t\boldsymbol{\lambda}_i}{\boldsymbol{\Lambda}_m}\mathbf{v}_i, \;\; \bar{\boldsymbol{\epsilon}}_m = \sum_{i=1}^m \frac{t\boldsymbol{\lambda}_i}{\boldsymbol{\Lambda}_m}[\boldsymbol{\epsilon}_i + \langle \mathbf{x}_i - \bar{\mathbf{x}}_m, \mathbf{v}_i - \bar{\mathbf{v}}_i\rangle].$$

If step **2.a)** executed at iteration j, then $z_j = z_{j-1}$. Therefore $\mathbf{z}_{i-1} = z_{k_i-1}$ for $i = 1,\ldots,m$. Using this result, Proposition 6.3 for $k = k_i$, (34), and the above defintions we conclude that

$$\mathbf{v}_i \in (F + \mathbf{N}_{\mathbb{R}^N\times\mathbb{R}^M_+})^{[\boldsymbol{\epsilon}_i]}(\mathbf{z}_i), \quad \|\boldsymbol{\lambda}_i\mathbf{v}_i + \mathbf{x}_i - \mathbf{z}_{i-1}\|^2 + 2\boldsymbol{\lambda}_i\boldsymbol{\epsilon}_i \leq (1/2)^2\|\mathbf{x}_i - \mathbf{z}_{i-1}\|^2,$$
$$\mathbf{z}_i = \mathbf{z}_{i-1} - t\boldsymbol{\lambda}_i\mathbf{v}_i, \quad \boldsymbol{\lambda}_i\|\mathbf{x}_i - \mathbf{z}_{i-1}\| \geq c > 0,$$

for $i = 1,2,\ldots,m$. Therefore, the sequences $(\sigma_i = 1/2)$, $(\boldsymbol{\lambda}_i)$, $(t_i = t)$, $(\mathbf{z}_i)$, $(\mathbf{x}_i)$, $(\mathbf{v}_i)$, $(\boldsymbol{\epsilon}_i)$ can be regarded as sequences generater by the RHPE, as discussed in Section 3, for operator $F + \mathbf{N}_{\mathbb{R}^N\times\mathbb{R}^M_+}$. To end the proof, use the above relations, Proposition 3.4 with $T = F + \mathbf{N}_{\mathbb{R}^N\times\mathbb{R}^M_+}$,

$$t_k = t,\; \sigma_k = 1/2,\; x_k = \mathbf{x}_k,\; z_k = \mathbf{z}_k,\; v_k = \mathbf{v}_k,\; \varepsilon_k = \boldsymbol{\epsilon}_k$$

c and t as in (34); take in to account (37); and observe that

$$\boldsymbol{\Lambda}_m = \Lambda_k, \quad \bar{\mathbf{x}}_m = (\bar{x}_k, \bar{y}_k), \quad \bar{\mathbf{v}}_m = \bar{v}_k, \quad \bar{\boldsymbol{\epsilon}}_m = \bar{\varepsilon}_k.$$

□

We will consider a particular case of Lemma 6.4, with a particular choice of the parameters μ_0, ν_0, β, and θ. Since $\kappa \leq \beta/\gamma$ and $\gamma \geq \mu_0 L/\sqrt{2\nu_0^3}$

$$\frac{\sqrt{\nu_0^3}}{\mu_0 \kappa} \geq \frac{L}{\beta\sqrt{2}}, \qquad \frac{\sqrt{\nu_0^3}}{\mu_0\sqrt{\kappa^3}} \geq \frac{L}{\beta},$$

where the second inequality follows from the first one and the first inequality in (28). Therefore, the best one can expect is to have the two inequalities close to equalities. With this aim, we design an initialization and a phase one method to reach such a goal.

7 A Phase I procedure for Algorithm 1

From now on we will use

$$\theta = \beta = \frac{1}{2}, \qquad z_0 = (\tilde{x}, 0), \tag{38}$$

where $\tilde{x} \in \mathbb{R}^N$ is an user-supplied initial point. To initialize Algorithm I with these parameter we will use the following procedure, which aim is to compute $x \in \mathbb{R}^N$, $y, s \in \mathbb{R}^M$, $z \in \mathbb{R}^N \times \mathbb{R}^M$, $\mu, \nu > 0$ such that

$$\frac{\mu L}{\sqrt{2\nu^3}} = 1, \quad y > 0, \quad s > 0, \quad \Phi_{1/2}(x, y, s; z, \mu, \nu) \leq \frac{1}{2}. \tag{39}$$

Phase I method: input $\tilde{x} \in \mathbb{R}^N$

1.) if $2\|F(\tilde{x}, e)\| \leq L$ **then**

$$\mu = \frac{\sqrt{2}}{L}, \quad \nu = 1, \quad x = \tilde{x}, \quad y = e, \quad s = \mu^{-1}e, \quad z = (\tilde{x}, 0),$$

2.) else

$k \leftarrow 1$

$\theta = \frac{1}{2}, \quad \beta = \frac{1}{2}, \quad \kappa = \frac{1}{2}, \qquad \tau = \frac{1}{4\sqrt{n}},$

$\mu'_0 = \frac{1}{\sqrt{2}\,\|F(\tilde{x}, e)\|}, \qquad \nu'_0 = 1, \quad x'_0 = \tilde{x}, \quad y'_0 = e, \quad s'_0 = (\mu'_0)^{-1}e, \quad z' = (\tilde{x}, 0),$

while $\mu'_{k-1} L/\sqrt{2{\nu'_{k-1}}^3} < 1$

$u_k = \max\left\{1 - \tau, \frac{(\mu'_{k-1}L)^2}{2{\nu'_{k-1}}^3}\right\}, \quad \mu'_k = u_k \mu'_{k-1}, \quad \nu'_k = u_k \nu'_{k-1}$

set $x = x'_{k-1}$, $y = y'_{k-1}$, $s = s'_{k-1}$ and compute (d_k^x, d_k^y, d_k^s) solution of

$$\begin{bmatrix} \mu'_k DF(x,y) + \nu'_k I & & 0 \\ & & -\mu'_k I \\ 0 & \mu'_k S & \mu'_k Y \end{bmatrix} \begin{bmatrix} d_k^x \\ d_k^y \\ d_k^s \end{bmatrix} = -H_{z', \mu'_k, \nu'_k}(x, y, s) \tag{40}$$

$x'_k = x + d_k^x$, $y'_k = y + d_k^y$, $s'_k = s + d_k^s$

$k \leftarrow k + 1$

end while

$x = x'_{k-1}$, $y = y'_{k-1}$, $s = s'_{k-1}$, $z = z'$, $\mu = \mu'_{k-1}$, $\nu = \nu'_{k-1}$

endif

The computational burden of the above procedure is in the **while** loop, in which one Jacobian shall be computed and a linear system shall be solved.

Proposition 7.1. *After at most*

$$\left\lceil 8\sqrt{n}\log\left(\frac{2\|F(\tilde{x},e)\|}{L}\right)\right\rceil_+$$

loops of the **while***, Phase I method outputs* $x \in \mathbb{R}^N$*,* $y, s \in \mathbb{R}^M_{++}$*,* $\mu, \nu > 0$ *satisfying* (39) *for* $z = (\tilde{x}, 0)$ *and*

$$\mu = \frac{\sqrt{2}}{L}\min\left\{1, \frac{L}{2\|F(\tilde{x},e)\|}\right\}^3, \quad \nu = \min\left\{1, \frac{L}{2\|F(\tilde{x},e)\|}\right\}^2.$$

Proof. It is trivial to verify that if $4\|F(\tilde{x},e)\| \leq L$ then the **while** loop is not reached and μ, ν, x, y, s, z as prescribed in the method satisfy (39). Suppose that

$$2\|F(\tilde{x},e)\| > L. \tag{41}$$

We claim that all steps in the **while** loop are well defined and that after k loops

$$\frac{\mu'_k L}{\sqrt{2\nu'^3_k}} \leq 1, \quad y'_k > 0, \quad s'_k > 0, \quad \Phi_{1/2}(x'_k, y'_k, s'_k; z', \mu'_k, \nu'_k) \leq 1/4. \tag{42}$$

In view of (41) and the definitions of μ'_0, ν'_0, x'_0, y'_0, s'_0 and z', the above inequalities hold for $k = 0$, and the first one holds as a strict inequality. Assume that the **while** loop is well define up to $k-1$ iterations and that these relations holds for $k-1$. If the first above inequality holds as an equality, then there will be no more loop iterations and the output of Phase I method satisfies (39). If the first inequality in (42) holds as an (strict) inequality, then $1 - \tau \leq u_k < 1$,

$$\frac{\mu'_{k-1} L}{\sqrt{2\nu'^3_{k-1}}} < \frac{1}{\sqrt{u_k}}\frac{\mu'_{k-1} L}{\sqrt{2\nu'^3_{k-1}}} = \frac{\mu'_k L}{\sqrt{2\nu'^3_k}} \leq 1.$$

and

$$\begin{aligned}
\Phi_{1/2}(x'_{k-1}, y'_{k-1}, s'_{k-1}; z', \mu'_k, \nu'_k) &= \sqrt{u_k}\frac{\|\mu'_{k-1}(F(x'_{k-1}, y'_{k-1}) - (0, s'_{k-1})) + \nu'_k((x'_{k-1}, y'_{k-1}) - z')\|}{\sqrt{2\nu'_{k-1}}} \\
&\quad + \|u_k \mu'_{k-1} Y'_{k-1} s'_{k-1} - e\| \\
&\leq \frac{\|\mu'_{k-1}(F(x'_{k-1}, y'_{k-1}) - (0, s'_{k-1})) + \nu'_k((x'_{k-1}, y'_{k-1}) - z')\|}{\sqrt{2\nu'_{k-1}}} \\
&\quad + u_k\|\mu'_{k-1} Y'_{k-1} s'_{k-1} - e\| + (1 - u_k)\sqrt{n} \\
&\leq 1/4 + \tau\sqrt{n} = 1/2.
\end{aligned}$$

It follows from two above equations and Theorem 4.4 that the linear system (40) has a unique solution and that (42) holds for $i = k$.

Once we know that the loop is well defined and that (42) holds in all **while** loop iteration, the bound on the number of loops and the last statement of the proposition follows by standard arguments and the definition of u_k. □

From now on in this section we make the following assumptions.

a.3) The solution set of problem (19) is non-empty and d is the distance of $(\tilde{x}, 0)$ to this set.
a.4) Phase I method is executed, and its output x, y, s, z, μ, ν is used in the initialization of Algorithm I with $\theta = \beta = 1/2$

$$x_0 = x, \quad y_0 = y, \quad s_0 = s, \quad z_0 = z = (\tilde{x}, 0), \quad \mu = \mu_0, \quad \nu_0 = \nu.$$

Let (x_k), (y_k), (s_k), (z_k), (μ_k) and (ν_k) be the sequences generated by Algorithm I with the initialization as above described. In view of Proposition 7.1

$$\frac{\mu_0 L}{\sqrt{2\nu_0^3}} = 1, \quad x_0 > 0, \quad y_0 > 0, \quad \Phi_{1/2}(x_0, y_0, s_0; z_0, \mu_0, \nu_0) \leq \frac{1}{2}, \tag{43}$$

$$\mu_0 = \frac{\sqrt{2}}{L} \min\left\{1, \frac{L}{2\|F(\tilde{x}, e)\|}\right\}^3, \quad \nu_0 = \min\left\{1, \frac{L}{2\|F(\tilde{x}, e)\|}\right\}^2. \tag{44}$$

This initialization, together with the use of $\beta = \theta = 1$, allows a considerable simplification of Lemmas 6.4 and 6.2. Let $\#A_k$, $\#B_k$ be as in (32), and let $\bar{x}_k$, $\bar{v}_k$ and $\bar{\varepsilon}_k$ be as in (35), (36).

Corollary 7.2. *If, after k iterations of Algorithm 1, step* **2.b)** *(of Alg. 1) is executed $m \geq 1$ or more times, then*

1. *there is an $i \leq k$ such that*

$$\|F(x_i, y_i) - (0, s_i)\| \leq 14\, L\frac{d^2}{m}, \qquad \langle y_i, s_i \rangle \leq \frac{14\sqrt{7}}{3\sqrt{3}} L\sqrt{\sqrt{n} + 1/2}\, \frac{d^3}{m^{3/2}};$$

2. $\bar{v}_k \in (F + \mathbf{N}_{\mathbb{R}^N \times \mathbb{R}^M_{++}})^{[\bar{\varepsilon}_k]}(\bar{x}_k, \bar{y}_k)$,

$$\|\bar{v}_k\| \leq \frac{2^3 7\sqrt{7}}{\sqrt{3}} L\sqrt{\sqrt{n} + 1/2}\, \frac{d^2}{m^{3/2}}, \qquad \bar{\varepsilon}_k \leq \frac{2^5 7\sqrt{7}}{3\sqrt{3}} L\sqrt{\sqrt{n} + 1/2}\, \frac{d^3}{m^{3/2}}.$$

Proof. Use Lemma 6.4 and (43). □

Corollary 7.3. *If in iteration k of Algorithm I step* **2.a)** *is executed and*

$$\#A_k - \#B_k = m + 1,$$

then

$$\|F(x_k, y_k) - (0, s_k)\| \leq \frac{2L(\sqrt{n} + 1)}{(1 + h)^{2m}} \max\left\{1, \frac{2\|F(\tilde{x}, e)\|}{L}\right\}^2, \quad \langle y_k, s_k \rangle \leq \frac{L\sqrt{n}(\sqrt{n} + 1/2)}{\sqrt{2}(1 + h)^{3m}} \max\left\{1, \frac{2\|F(\tilde{x}, e)\|}{L}\right\}^3.$$

Proof. Use Lemma 6.2, (43) and (44). □

First we analyze the point-wise complexity of the tandem combination of the Phase I method with Algorithm I.

Theorem 7.4. *Let j be the total number of iterations of the composition of Phase I method with Algorithm 1, that is, j is the sum of the number of* **while** *loops iterations of Phase I with the number of iterations of Algorithm 1.*

For any $\delta, \varepsilon > 0$ an approximate solution

$$x_i \in \mathbb{R}^N, \quad y_i > 0, \quad s_i > 0, \qquad \|F(x_i, y_i) - (0, s_i)\| \leq \delta, \quad \langle y_i, s_i \rangle \leq \varepsilon$$

is reached after at most a total of

$$j = 1 + \tilde{m} + 2\tilde{n} + \left\lceil 8\sqrt{n} \log\left(\frac{4\|F(e)\|}{L}\right)\right\rceil_+$$

iterations, where

$$\widetilde{n} = \left\lceil 14\, d^2 \max\left\{\frac{L}{\delta}, \frac{L^{2/3}(\sqrt{n}+1/2)^{1/3}}{3 * 2^{1/3}\, \varepsilon^{2/3}}\right\}\right\rceil,$$

$$\tilde{m} = \left\lceil 4 * 7(\sqrt{n}+1/2)\left\lceil \max\left\{\frac{1}{2}\log\left(\frac{2L(\sqrt{n}+1)}{\delta}\right), \frac{1}{3}\log\left(\frac{L(\sqrt{n}+1/2)^2}{\varepsilon\sqrt{2}}\right)\right\}_+ + \left(\log\frac{2\|F(e)\|}{L}\right)_+\right\rceil\right\rceil$$

and d is the distance of $z_0 = (\tilde{x}, 0)$ to the solution set of (19).

Proof. The number of **while** loops in Phase 1 is bounded by

$$\left\lceil 8\sqrt{n} \log\left(\frac{4\|F(e)\|}{L}\right)\right\rceil_+.$$

Let

$$\tilde{k} = 1 + 2\widetilde{m} + \widetilde{n}.$$

If $\#B_{\tilde{k}} \geq \widetilde{n}$, then the conclusion follows from Corollary 7.2, item 1. Suppose that

$$\#B_{\tilde{k}} < \tilde{n}.$$

In this case $\#A_{\tilde{k}} = \tilde{k} - \#B_{\tilde{k}} > \tilde{n} + \tilde{m}$. Therefore

$$\#A_{\tilde{k}} - \#B_{\tilde{k}} \geq \tilde{m} + 1.$$

Let $k \leq \tilde{k}$ be the smallest integer such that

$$\#A_k - \#B_k = \tilde{m} + 1.$$

Then, $k \in A_{\tilde{k}}$ and the conclusion follows from Corollary 7.3. □

Next we provide a complexity estimation which takes in to account the ergodic means defined in (35), (36). Its proof, being quite similar to the one of Theorem 7.4, will be omitted.

Theorem 7.5. *Let j be the total number of iterations of the composition of Phase I method with Algorithm 1, that is, j is the sum of the number of* **while** *loops iterations of Phase I with the number of iterations of Algorithm 1.*

Let $\delta, \varepsilon > 0$ and define

$$\widetilde{n} = \left\lceil 4 * 7L^{2/3}(\sqrt{n} + 1/2)^{1/3} \max\left\{ \frac{d^{4/3}}{3^{1/3}\delta^{2/3}}, \frac{2^{2/3}d^2}{3\varepsilon^{2/3}} \right\} \right\rceil,$$

$$\tilde{m} = \left\lceil 4 * 7(\sqrt{n} + 1/2) \left\lceil \max\left\{ \frac{1}{2}\log\left(\frac{2L(\sqrt{n}+1)}{\delta}\right), \frac{1}{3}\log\left(\frac{L(\sqrt{n}+1/2)^2}{\varepsilon\sqrt{2}}\right) \right\}_+ + \left(\log\frac{2\|F(e)\|}{L}\right)_+ \right\rceil \right\rceil.$$

Them after at most $\tilde{j} = 1 + \tilde{m} + \tilde{n}$ (total) iterations, either

1. *an ergodic mean $\bar{x}_k, \bar{v}_k, \bar{\varepsilon}_k$ is found, satisfying*

$$\bar{x}_k \in \mathbb{R}^N, \quad \bar{y}_k \geq 0, \quad \bar{v}_k \in (F + \mathbf{N}_{\mathbb{R}^N \times \mathbb{R}^M_+})^{[\bar{\varepsilon}_k]}(\bar{x}_k, \bar{y}_k), \quad \|\bar{v}_k\| \leq \delta, \quad \bar{\varepsilon}_k \leq \varepsilon;$$

2. *an iterate x_k, y_k, s_k is found, satisfying*

$$x_i \in \mathbb{R}^N,\ y_i, s_i > 0, \quad \|F(x_i, y_i) - (0, s_i)\| \leq \delta, \quad \langle y_i, x_i \rangle \leq \varepsilon.$$

Proof. The proof is similar to that of Theorem 7.4, using Corollary 7.2, item 2 (instead of item 1) and Corollary 7.3. □

A A Basic Inequality

In this section we prove a basic inequality used for estimating the ergodic-mean complexity of the large-step relaxed HPE method in Section 3.

Lemma A.1. *If $\alpha_1, \ldots, \alpha_k > 0$ and $\sum_{i=1}^k \alpha_i^2 \leq a$ then*

$$\sum_{i=1}^k \frac{1}{\alpha_i} \geq \sqrt{\frac{k^3}{a}}.$$

Proof. Define $f : \mathbb{R}^k_{++} \to \mathbb{R}$ and $\alpha^* \in \mathbb{R}^k_{++}$

$$f(\alpha) = \sum_{i=1}^k \frac{1}{\alpha_i} + \frac{1}{2}\left(\frac{k}{a}\right)^{3/2}\left[\sum_{i=1}^k \alpha_i^2 - a\right], \qquad \alpha_1^* = \alpha_2^* = \cdots = \alpha_k^* = \sqrt{a/k}.$$

Observe that f is strictly convex in $\mathbb{R}^k_{++}$, $\nabla f(\alpha^*) = 0$, and $\sum_{i=1}^k (\alpha_i^*)^2 = a$; therefore, if

$$\alpha_1, \ldots, \alpha_k > 0, \quad \sum_{i=1}^k \alpha_i^2 \leq a$$

then

$$\sum_{i=1}^k \frac{1}{\alpha_i} \geq f(\alpha_1, \ldots, \alpha_k) \geq f(\alpha^*) = \sum_{i=1}^k \frac{1}{\alpha_i^*} = \sqrt{k^3/a}.$$

which concludes the proof. □

B Newton step for the NLMCP

To prove Lemma 4.3 we will use the next elementary linear algebra result

Lemma B.1. *If $y, s \in \mathbb{R}^M_{++}$, $\mu, \nu > 0$, $(b^x, b^y, b^s) \in \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^M$ and $A \in \mathbb{R}^{(N+M)\times(N+M)}$ is positive semi-definite, then the linear system*

$$\begin{bmatrix} A + \nu I & & 0 \\ & & -\mu I \\ 0 & \mu S & \mu Y \end{bmatrix} \begin{bmatrix} d^x \\ d^y \\ d^s \end{bmatrix} = \begin{bmatrix} b^x \\ b^y \\ b^s \end{bmatrix} \tag{45}$$

has a unique solution (d^x, d^y, d^s).

1. *If additionally $\|\mu Y s - e\| < 1$ then*

$$\mu\|D^y d^s\| \leq \frac{\|(b^x, b^y)\|^2}{4\nu} + \frac{\|b^s\|^2}{2(1 - \|\mu Y s - e\|)}$$

$$\frac{\nu\|(d^x, d^y)\|^2}{2} + \mu\|D^y d^s\| \leq \frac{\|(b^x, b^y)\|^2}{2\nu} + \frac{\|b^s\|^2}{2(1 - \|\mu Y s - e\|)}.$$

2. *If additionally $\|\mu Y s - e\| < 1$, $b^s = -(\mu Y s - e)$ and*

$$\frac{\|(b^x, b^y)\|^2}{4\nu} + \frac{\|b^s\|^2}{2(1 - \|Y s - e\|)} < 1$$

   then $y + d^y > 0$ and $s + d^s > 0$.

Proof. Pre-multiplying the last M lines of the coefficient matrix on (45) by S^{-1} we obtain a positive definite matrix, because $y, s > 0$ $\mu, \nu > 0$ and A is positive definite. Therefore, this coefficient matrix is non-singular and (45) has a unique solution.

To prove item 1, (left) multiply both sides of (45) by $(d^x, d^y, 0)^T$ and use the assumption of A being positive semi-definite to conclude that

$$\nu\|(d^x, d^y)\|^2 - \mu\langle d^y, d^s\rangle \leq \langle (d^x, d^y), (b^x, b^y)\rangle.$$

Since the 2-norm is bounded by the 1-norm and $2ab \leq a^2 + b^2$ for any $a, b \in \mathbb{R}$,

$$\begin{aligned}
\mu\|D^y d^s\| \leq \mu \sum |d_i^2 d_i^3| &\leq \frac{\mu}{2} \sum \frac{s_i}{y_i}(d_i^2)^2 + \frac{y_i}{s_i}(d_i^3)^2 \\
&= \frac{\mu}{2} \left\| S^{1/2} Y^{-1/2} d^2 + S^{-1/2} Y^{1/2} d^3 \right\|^2 - \mu\langle d^2, d^3\rangle \\
&= \frac{(\mu Y S)^{-1}}{2} \left\| \mu S d^2 + \mu Y d^3 \right\|^2 - \mu\langle d^2, d^3\rangle \\
&\leq \frac{\|b^2\|^2}{2(1 - \|\mu Y S - e\|)} - \mu\langle d^s, d^3\rangle,
\end{aligned}$$

where the last inequality follows from the assumption $\|\mu Y s - e\| < 1$. Adding the above inequalities we conclude that

$$\nu\|(d^x, d^y)\|^2 + \mu\|D^y d^s\| \leq \langle (d^x, d^y), (b^x, b^y)\rangle + \frac{\|b^s\|^2}{2(1 - \|\mu Y s - e\|)}.$$

To end the proof of item 1, observe that

$$\langle (d^x, d^y), (b^x, b^y)\rangle - \nu\|(d^x, d^y)\|^2 \le \frac{\|(b^x, b^y)\|^2}{4\nu}, \quad \langle (d^x, d^y), (b^x, b^y)\rangle \le \frac{\nu\|(d^x, d^y)\|^2}{2} + \frac{\|(b^x, b^y)\|^2}{2\nu}$$

and combine the above inequalities

To prove item 2, first observe that under its assumptions it follows from item 2 that $\mu\|D^y d_s\| < 1$. Define $(y_\alpha, s_\alpha) = (y, s) + \alpha(d^y, d^s)$ for $\alpha \ge 0$ and observe that since $\mu S d^y + \mu Y d^s = b^s = -(\mu Y s - e)$,

$$\begin{aligned} \mu Y_\alpha s_\alpha - e &= \mu Y s + \alpha\mu(Y d^s + S d^y) + \alpha^2 \mu D^y d^s - e \\ &= \mu Y s + \alpha b^s + \alpha^2 \mu D^y d^s - e \\ &= (1-\alpha)(\mu Y s - e) + \alpha^2 \mu D^y d^s \end{aligned}$$

Therefore, for any $\alpha \in [0, 1]$

$$\|\mu Y_\alpha s_\alpha - e\| = (1-\alpha)\|\mu Y s - e\| + \alpha^2 \mu\|D^y d^s\| \le \max\{\|\mu Y s - e\|, \mu\|D^y d^s\|\} < 1$$

As a consequence, for any $\alpha \in [0, 1]$, all components of $\mu Y_\alpha s_\alpha$ are strictly positive. Since the components of y_α, s_α are continuous in α and are strictly positive for $\alpha = 0$, they do not vanish for any $\alpha \in [0, 1]$ and must be strictly positive for any such an α. In particular, for $\alpha = 1$, $y_{\alpha=1} = y + d^y > 0$ and $s_{\alpha=1} = s + d^s > 0$. □

References

[1] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

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

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

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

[5] Jim Douglas, Jr. and H. H. Rachford, Jr. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.

[6] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):1740, 1976.

[7] R. Glowinski and A. Marrocco. Sur lapproximation par el ements finis et la r esolution par p enalisation-dualit e dune classe de probl emes de dirichlet non lin eaires. *Revue Fran caise dAutomatique, Informatique, Recherche Operationnelle*, 2:4176, 1975.

[8] L. G. Hačijan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244(5):1093–1096, 1979.

[9] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

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

[11] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.

[12] George J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, 29:341–346, 1962.

[13] Renato D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim.*, 20(6):2755–2787, 2010.

[14] Renato 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(4):1688–1720, 2011.

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

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

[17] Arkadi 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(1):229–251 (electronic), 2004.

[18] R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33:209–216, 1970.

[19] Mauricio R. Sicre. *Trajetrias log-quadrticas e algoritmos de path-following para o Problema de Complementaridade Montono.* phd thesis, impa, estrada dona castorina 110, rio de janeiro, cep 22460-320, rj brasil. 2003.

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

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

[22] M. V. Solodov and B. F. Svaiter. Error bounds for proximal point subproblems and associated inexact proximal point algorithms. *Math. Program.*, 88(2, Ser. B):371–389, 2000. Error bounds in mathematical programming (Kowloon, 1998).

[23] M. V. Solodov and B. F. Svaiter. A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM J. Optim.*, 10(2):605–625, 2000.

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

[25] B. F. Svaiter. A family of enlargements of maximal monotone operators. *Set-Valued Anal.*, 8(4):311–328, 2000.

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