# Iteration-complexity of a Rockafellar's proximal method of multipliers for convex programming based on second-order approximations

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

April 18, 2018

**Abstract**

This paper studies the iteration-complexity of a new primal-dual algorithm based on Rockafellar's proximal method of multipliers (PMM) for solving smooth convex programming problems with inequality constraints. In each step, either a step of Rockafellar's PMM for a second-order model of the problem is computed or a relaxed extragradient step is performed. The resulting algorithm is a (large-step) relaxed hybrid proximal extragradient (r-HPE) method of multipliers, which combines Rockafellar's PMM with the r-HPE method. We obtain pointwise $\mathcal{O}(1/k)$ and ergodic $\mathcal{O}(1/k^{3/2})$ global convergence rates at the price of solving, at each iteration, quadratic quadratically constrained convex programming problems. These convergence rates are superior to the corresponding pointwise $\mathcal{O}(1/\sqrt{k})$ and ergodic $\mathcal{O}(1/k)$ currently known for standard proximal-point methods, thanks to the incorporation of second-order information. To the best of our knowledge, this is the first time that the above mentioned rates and results are obtained for solving the smooth convex programming problems with inequality constraints.

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

Key words: convex programming, proximal method of multipliers, second-order methods, hybrid extragradient, complexity.

## Introduction

The smooth convex programming problem with (for the sake of simplicity) only inequality constraints is

$$\min \quad f(x) \qquad \text{s.t.} \;\; g(x) \leq 0 \tag{1}$$

*Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (maicon.alves@ufsc.br). This work was done while this author was a postdoc at the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. The work of this author was partially supported by CNPq grants no. 406250/2013-8, 306317/2014-1 and 405214/2016-2.

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

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

where $f : \mathbb{R}^n \to \mathbb{R}$ and the components of $g = (g_1, \ldots, g_m) : \mathbb{R}^n \to \mathbb{R}^m$ are smooth convex functions. Dual methods for this problem solve the associated dual problem

$$\max \left( \inf_{x \in \mathbb{R}^n} f(x) + \langle y, g(x) \rangle \right) \qquad \text{s.t. } \; y \geq 0$$

and, *en passant*, find a solution of the original (primal) problem. Notice that a pair $(x, y)$ satisfies the Karush-Kuhn-Tucker conditions for problem (1) if and only if $x$ is a solution of this problem, $y$ is a solution of the associated dual problem, and there is no duality gap. There are many practical methods for solving (1), e.g., the method of multipliers, sequential quadratic programming, stabilized sequential quadratic programming, semi-smooth Newton methods, active set methods, and barrier/penalization methods, to cite some of them.

The *method of multipliers*, which was proposed by Hestenes [3, 4] and Powel [11] for equality constrained optimization problems and extended by Rockafellar [12] (see also [13]) to inequality constrained convex programming problems, is a typical example of a dual method. It generates iteractively sequences $(x_k)$ and $(y_k)$ as follows:

$$x_k \approx \arg\min_{x \in \mathbb{R}^n} \; \mathscr{L}(x, y_{k-1}, \lambda_k), \qquad y_k = y_{k-1} + \lambda_k \nabla_y \mathscr{L}(x_k, y_{k-1}, \lambda_k)$$

where $\approx$ stands for approximate solution, $\lambda_k > 0$, and $\mathscr{L}(x, y, \lambda)$ is the *augmented Lagrangian*

$$\begin{aligned} \mathscr{L}(x, y, \lambda) &= f(x) + \frac{1}{2\lambda}[\|(y + \lambda g(x))_+\|^2 - \|y\|^2] \\ &= \max_{y' \geq 0} \; f(x) + \langle y', g(x) \rangle - \frac{1}{2\lambda}\|y' - y\|^2. \end{aligned}$$

The method of multipliers is also called the *augmented Lagrangian method*. In the seminal article [13], Rockafellar proved that the method of multipliers is an instance of his proximal point method (hereafter PPM) [14] applied to the dual objective function. Still in [13], Rockafellar proposed a new primal-dual method for (1), which we discuss next and that we will use in this paper to design a new primal-dual method for this problem.

Rockafellar's *proximal method of multipliers* (hereafter PMM) [13] generates, for any starting point $(x_0, y_0)$, a sequence $\big((x_k, y_k)\big)_{k \in \mathbb{N}}$ as the approximate solution of a regularized saddle-point problem

$$(x_k, y_k) \approx \arg\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m_+} \; f(x) + \langle y, g(x) \rangle + \frac{1}{2\lambda}\left[\|x - \mathring{x}\|^2 - \|y - \mathring{y}\|^2\right] \tag{2}$$

where $(\mathring{x}, \mathring{y}) = (x_{k-1}, y_{k-1})$ is the current iterate and $\lambda = \lambda_k > 0$ is a stepsize parameter. Notice that the objective function of the above saddle-point problem is obtained by adding to the augmented Lagrangian a proximal term for the primal variable $x$. If $\inf \lambda_k > 0$ and

$$\sum_{k=1}^{\infty} \|(x_k, y_k) - (x_k^*, y_k^*)\| < \infty \tag{3}$$

where $(x_k^*, y_k^*)$ is the (exact) solution of (2), then $((x_k, y_k))_{k \in \mathbb{N}}$ converges to a solution of the Karush-Kuhn-Tucker conditions for (1) provided that there exist a pair satisfying these conditions. This result follows from the facts that the satisfaction of KKT conditions for (1) can be formulated as a monotone inclusion problem and (2) is the Rockafellar's PPM iteration for this inclusion problem (see comments after Proposition 2.2). Although (2) is a (strongly) convex-concave problem – and hence has a unique solution – the computation of its exact or an approximate solution can be very hard.

We assume in this paper that $f$ and $g_i$ ($i = 1, \dots, m$) are $\mathscr{C}^2$ convex functions with Lipschitz continuous Hessians. The method proposed in this paper either solves a second-order model of (2) in which second-order approximations of $f$ and $g_i$ ($i = 1, \dots, m$) replace these functions in (2) or performs a (relaxed) extragradient step. In its general form, PMM is an inexact PPM in that each iteration approximately solves (2) according to the summable error criterion (3). The method proposed in this paper can also be viewed as an inexact PPM but one based on a relative error criterion instead of the one in (3). More specifically, it can be viewed as an instance of the (large-step) relaxed hybrid proximal extragradient (r-HPE) method [8, 16, 22] which we briefly discuss next.

Given a point-to-set maximal monotone operator $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$, the large-step r-HPE method computes approximate solutions for the monotone inclusion problem $0 \in T(z)$ as extragradient steps

$$z_k = z_{k-1} - \tau \lambda_k v_k, \tag{4}$$

where $z_{k-1}$ is the current iterate, $\tau \in (0, 1]$ is a relaxation parameter, $\lambda_k > 0$ is the stepsize and $v_k$ together with the pair $(\tilde{z}_k, \varepsilon_k)$ satisfy the following conditions

$$\begin{aligned} & v_k \in T^{[\varepsilon_k]}(\tilde{z}_k), \qquad \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2, \\ & \lambda_k \|\tilde{z}_k - z_{k-1}\| \geq \eta \end{aligned} \tag{5}$$

where $\sigma \in [0, 1)$ and $\eta > 0$ are given constants and $T^{\varepsilon}$ denotes the $\varepsilon$-enlargement of $T$. (It has the property that $T^{\varepsilon}(z) \supset T(z)$ for every $z$.) The method proposed in this paper for solving the minimization problem (1) can be viewed as a realization of the above framework where the operator $T$ is the standard saddle-point operator defined as $T(z) := (\nabla f(x) + \nabla g(x)y, -g(x) + N_{\mathbb{R}^m_+}(y))$ for every $z = (x, y)$. More specifically, the method consists of two type of iterations. The ones which perform extragradient steps can be viewed as a realization of (5). On the other hand, each one of the other iterations updates the stepsize by increasing it by a multiplicative factor larger than one and then solves a suitable second-order model of (2). After a few of these iterations, an approximate solution satisfying (5) is then obtained. Hence, in contrast to the PMM which does not specify how to obtain an approximate solution $(x_k, y_k)$ of (2), or equivalently the prox inclusion $0 \in \lambda_k T(z) + z - z_{k-1}$ with $T$ as above, these iterations provide a concrete scheme for computing an approximate solution of this prox inclusion according to the relative criterion in (5). Pointwise and ergodic iteration-complexity bounds are then derived for our method using the fact that the large-step r-HPE method has pointwise and ergodic global convergence rates of $\mathcal{O}(1/k)$ and $\mathcal{O}(1/k^{3/2})$, respectively.

We emphasize that to the best of our knowledge this is the first time that global (pointwise and ergodic) convergence rates are obtained for a second-order type algorithm for solving (1). The proposed method (Algorithm 1) combines two distinct approaches for solving optimization problems: proximal-point type and second-order methods. The latter is known to have good local performance but, even in an unconstrained instance of (1), some sort of regularization is needed to guarantee that the corresponding iteration is well-defined along the whole iterative process (see, e.g., [10]). On the other hand, a standard proximal-point algorithm (e.g., PMM) for solving (1), while globally well-defined, would require at each iteration the solution of potentially numerically expensive subproblems (e.g., (2)). Moreover, this approach would lead to global pointwise $\mathcal{O}(1/\sqrt{k})$ and ergodic $\mathcal{O}(1/k)$ convergence rates. Summarizing, we combine the both just mentioned approaches to obtain a proximal-point second-order type algorithm (Algorithm 1) with superior pointwise $\mathcal{O}(1/k)$ and ergodic $\mathcal{O}(1/k^{3/2})$ global convergence rates at the price of solving, at each iteration, quadratic quadratically constrained subproblems.

The paper is organized as follows. Section 1 reviews some basic properties of $\varepsilon$-enlargements of maximal monotone operators and briefly reviews the basic properties of PPM and the large-step r-HPE method. Section 2 presents the basic properties of the minimization problem of interest and some equivalences between certain saddle-point, complementarity and monotone inclusion problems, as well as of its regularized versions. Section 3 introduces an error measure, shows some of its properties and how it is related to the relative error criterion for the large-step r-HPE method. Section 4 studies the smooth convex programming problem (1) and its second-order approximations. The proposed method (Algorithm 1) is presented in Section 5 and its iteration-complexities (pointwise and ergodic) are studied in Section 6.

# 1 Rockafellar's proximal method and the hybrid proximal extragradient method

This work is based on Rockafellar's proximal point method (PPM). The new method presented in this paper is a particular instance of the (large-step) relaxed hybrid proximal extragradient (r-HPE) method [15]. For these reasons, in this section we review Rockafellar's PPM, the large-step r-HPE method, and review some convergence properties of these methods.

### Maximal monotone operators, the monotone inclusion problem, and Rockafellar's proximal point method

A *point-to-set operator* in $\mathbb{R}^p$, $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$, is a relation $T \subset \mathbb{R}^p \times \mathbb{R}^p$ and

$$T(z) := \{v \mid (z, v) \in T\}, \qquad z \in \mathbb{R}^p.$$

The *inverse* of $T$ is $T^{-1} : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$, $T^{-1} := \{(v, z) \mid (z, v) \in T\}$. The *domain* and the range of $T$ are, respectively,

$$D(T) := \{z \mid T(z) \neq \emptyset\}, \quad R(T) := \{v \mid \exists z \in \mathbb{R}^p,\ v \in T(z)\}.$$

When $T(z)$ is a singleton for all $z \in D(T)$, it is usual to identify $T$ with the map $D(T) \ni z \mapsto v \in \mathbb{R}^p$ where $T(z) = \{v\}$. If $T_1, T_2 : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ and $\lambda \in \mathbb{R}$, then $T_1 + T_2 : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ and $\lambda T_1 : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ are defined as

$$(T_1 + T_2)(z) := \{v_1 + v_2 \mid v_1 \in T_1(z),\ v_2 \in T_2(z)\}, \qquad (\lambda T_1)(z) := \{\lambda v \mid v \in T_1(z)\}.$$

A point-to-set operator $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ is *monotone* if

$$\langle z - z', v - v' \rangle \geq 0, \qquad \forall (z, v), (z', v') \in T$$

and it is *maximal monotone* if it is a maximal element in the family of monotone point-to-set operators in $\mathbb{R}^p$ with respect to the partial order of set inclusion. The subdifferential of a proper closed convex function is a classical example of a maximal monotone operator. Minty's theorem [5] states that if $T$ is maximal monotone and $\lambda > 0$, then the *proximal map* $(\lambda T + I)^{-1}$ is a point-to-point nonexpansive operator with domain $\mathbb{R}^p$.

The *monotone inclusion problem* is: given $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ maximal monotone, find $z$ such that

$$0 \in T(z). \tag{6}$$

Rockafellar's PPM [14] generates, for any starting $z_0 \in \mathbb{R}^p$, a sequence $(z_k)$ by the approximate rule

$$z_k \approx (\lambda_k T + I)^{-1} z_{k-1},$$

where $(\lambda_k)$ is a sequence of strictly positive *stepsizes*. Rockafellar proved [14] that if (6) has a solution and

$$\left\| z_k - (\lambda_k T + I)^{-1}(z_{k-1}) \right\| \leq e_k, \ \sum_{k=1}^{\infty} e_k < \infty, \quad \inf \lambda_k > 0, \tag{7}$$

then $(z_k)$ converges to a solution of (6).

In each step of the PPM, computation of the proximal map $(\lambda T + I)^{-1} z$ amounts to solving the *proximal (sub) problem*

$$0 \in \lambda T(z_+) + z_+ - z,$$

a regularized inclusion problem which, although well-posed, is almost as hard as (6). From this fact stems the necessity of using approximations of the proximal map, for example, as prescribed in (7). Moreover, since each new iterate is, hopefully, just a better approximation to the solution than the old one, if it was computed with high accuracy, then the computational cost of each iteration would be too high (or even prohibitive) and this would impair the overall performance of the method (or even make it infeasible).

So, it seems natural to try to improve Rockafellar's PPM by devising a variant of this method that would accept a *relative* error tolerance and wherein the progress of the iterates towards the solution set could be estimated. In the next subsection we discuss the hybrid proximal extragradient (HPE) method, a variant of the PPM which aims to satisfy these goals.

### Enlargements of maximal monotone operators and the hybrid proximal extragradient method

The HPE method [16, 17] is a modification of Rockafellar's PPM wherein: (a) the proximal subproblem, in each iteration, is to be solved within a *relative* error tolerance and (b) the update rule is modified so as to guarantee that the next iterate is closer to the solution set by a quantifiable amount.

An additional feature of (a) is that, in some sense, errors in the inclusion on the proximal subproblems are allowed. Recall that the *ε-enlargement* [2] of a maximal monotone operator $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ is

$$T^{[\varepsilon]}(z) := \{v \mid \langle z - z', v - v' \rangle \geq -\varepsilon \ \forall (z', v') \in T\}, \qquad z \in \mathbb{R}^p, \ \varepsilon \geq 0. \tag{8}$$

From now on in this section $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ is a maximal monotone operator. The r-HPE method [20] for the monotone inclusion problem (6) proceed as follows: choose $z_0 \in \mathbb{R}^p$ and $\sigma \in [0, 1)$; for $i = 1, 2, \dots$

$$\begin{array}{ll} \text{compute } \tilde{z}_i, v_i, \lambda_i > 0, \varepsilon_i \geq 0 \text{ such that} & v_i \in T^{[\varepsilon_i]}(\tilde{z}_i), \;\; \|\lambda_i v_i + \tilde{z}_i - z_{i-1}\|^2 + 2\lambda_i\varepsilon_i \leq \sigma^2 \|\tilde{z}_i - z_{i-1}\|^2, \\ \text{choose } \tau_i \in (0, 1] \text{ and set} & z_i = z_{i-1} - \tau_i \lambda_i v_i. \end{array} \tag{9}$$

In practical applications, each problem has a particular structure which may render feasible the computation of $\lambda_i$, $\tilde{z}_i$, $v_i$, and $\varepsilon_i$ as prescribed above. For example, $T$ may be Lipschitz continuous, it may be differentiable, or it may be a sum of an operator which has a proximal map easily computable with others with some of these properties. Prescription for computing $\lambda_i$, $\tilde{z}_i$, $v_i$, and $\varepsilon_i$ under each one of these assumptions were presented in [6, 7, 8, 9, 16, 19, 21].

Computation of $(\lambda T + I)^{-1}(z)$ is equivalent to the resolution of an inclusion-equation system:

$$z_+ = (\lambda T + I)^{-1}(z) \iff \exists v \in T(z_+), \; \lambda v + z_+ - z = 0.$$

Whence, the error criterion in the first line of (9) relaxes both the inclusion and the equality at the right-hand side of the above equivalence. Altogether, each r-HPE iteration consists in: (1) solving (with a relative error tolerance) a "proximal" inclusion-equation system; (2) updating $z_{i-1}$ to $z_i$ by means of an extragradient step, that is, using $v_i \in T^{[\varepsilon_i]}(\tilde{z}_i)$.

In the remainder part of this section we present some convergence properties of the r-HPE Method which were essentially proved in [15] and revised in [22]. The next proposition shows that $z_i$ is closer than $z_{i-1}$ to the solution set with respect to the square of the norm, by a quantifiable amount, and present some useful estimations.

**Proposition 1.1** ([22, Proposition 2.2])**.** *For any $i \geq 1$ and $z^* \in T^{-1}(0)$,*

(a) $(1-\sigma)\|\tilde{z}_i - z_{i-1}\| \leq \|\lambda_i v_i\| \leq (1+\sigma)\|\tilde{z}_i - z_{i-1}\|$ *and* $2\lambda_i\varepsilon_i \leq \sigma^2\|\tilde{z}_i - z_{i-1}\|^2$*;*

(b) $\|z^* - z_{i-1}\|^2 \geq \|z^* - z_i\|^2 + \tau_i(1-\sigma^2)\|\tilde{z}_i - z_{i-1}\|^2 \geq \|z^* - z_i\|^2$*;*

(c) $\|z^* - z_0\|^2 \geq \|z^* - z_i\|^2 + (1-\sigma^2)\sum_{j=1}^{i} \tau_j \|\tilde{z}_j - z_{j-1}\|^2$*;*

(d) $\|z^* - \tilde{z}_i\| \leq \|z^* - z_{i-1}\|/\sqrt{1-\sigma^2}$ *and* $\|\tilde{z}_i - z_{i-1}\| \leq \|z^* - z_{i-1}\|/\sqrt{1-\sigma^2}$.

The aggregate stepsize $\Lambda_i$ and the ergodic sequences $(\tilde{z}_i^a)$, $(\tilde{v}_i^a)$, $(\varepsilon_i^a)$ associated with the sequences $(\lambda_i)$, $(\tilde{z}_i)$, $(v_i)$, and $(\varepsilon_i)$ are, respectively,

$$\begin{aligned} &\Lambda_i := \sum_{j=1}^{i} \tau_j \lambda_j, \\ &\tilde{z}_i^a := \frac{1}{\Lambda_i} \sum_{j=1}^{i} \tau_j \lambda_j \tilde{z}_j, \quad v_i^a := \frac{1}{\Lambda_i} \sum_{j=1}^{i} \tau_j \lambda_j v_j, \quad \varepsilon_i^a := \frac{1}{\Lambda_i} \sum_{j=1}^{i} \tau_j \lambda_j (\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_i^a, v_j - v_i^a \rangle). \end{aligned} \tag{10}$$

Next we present the pointwise and ergodic iteration-complexities of the large-step r-HPE method, i.e., the r-HPE method with a large-step condition [7, 8]. We also assume that the sequence of relaxation parameters $(\tau_i)$ is bounded away from zero.

**Theorem 1.2** ([22, Theorem 2.4]). *If $d_0$ is the distance from $z_0$ to $T^{-1}(0) \neq \emptyset$ and*

$$\lambda_i \|\tilde{z}_i - z_{i-1}\| \geq \eta > 0, \quad \tau_i \geq \tau > 0 \qquad\qquad i = 1, 2, \ldots$$

*then, for any $i \geq 1$,*

(a) *there exists $j \in \{1, \ldots, i\}$ such that $v_j \in T^{[\varepsilon_j]}(\tilde{z}_j)$ and*

$$\|v_j\| \leq \frac{d_0^2}{i\tau(1-\sigma)\eta}, \qquad \varepsilon_j \leq \frac{\sigma^2 d_0^3}{(i\tau)^{3/2}(1-\sigma^2)^{3/2} 2\eta};$$

(b) $v_i^a \in T^{[\varepsilon_i^a]}(\tilde{z}_i^a)$, $\|v_i^a\| \leq \dfrac{2d_0^2}{(i\tau)^{3/2}(\sqrt{1-\sigma^2})\eta}$, *and* $\varepsilon_i^a \leq \dfrac{2d_0^3}{(i\tau)^{3/2}(1-\sigma^2)\eta}$.

**Remark.** We mention that the inclusion in Item (a) of Theorem 1.2 is in the enlargement of $T$ which appears in the inclusion in (9). To be more precise, in some applications the operator $T$ may have a special structure, like for instance $T = S + N_{\mathcal{X}}$, where $S$ is point-to-point and $N_{\mathcal{X}}$ is the normal cone operator of a closed convex set $\mathcal{X}$, and the inclusion in (9), in this case, is $v_i \in \left(S + N_{\mathcal{X}}^{[\varepsilon_i]}\right)(\tilde{z}_i)$, which is stronger than $v_i \in T^{[\varepsilon_i]}(\tilde{z}_i)$. In such a case, Item (a) would guarantee that $v_j \in \left(S + N_{\mathcal{X}}^{[\varepsilon_j]}\right)(\tilde{z}_j)$. Unfortunately, the observation is not true for the Item (b).

The next result was proved in [18, Corollary 1]

**Lemma 1.3.** *If $\mathring{z} \in \mathbb{R}^p$, $\lambda > 0$, and $v \in T^{[\varepsilon]}(z)$, then*

$$\|\lambda v + z - \mathring{z}\|^2 + 2\lambda\varepsilon \geq \left\|z - (\lambda T + I)^{-1}\mathring{z}\right\|^2 + \left\|\lambda v - \left(\mathring{z} - (\lambda T + I)^{-1}\mathring{z}\right)\right\|^2.$$

## 2 The smooth convex programming problem

Consider the smooth convex optimization problem (1), i.e.,

$$(P) \qquad \min f(x) \qquad \text{s.t. } g(x) \leq 0, \tag{11}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g = (g_1, \ldots, g_m) : \mathbb{R}^n \to \mathbb{R}^m$. From now on we assume that:

O.1) $f, g_1, \ldots, g_m$ are convex $\mathscr{C}^2$ functions;

O.2) the Hessians of $f$ and $g_1, \ldots, g_m$ are Lipschitz continuous with Lipschitz constants $L_0$ and $L_1, \ldots, L_m$, respectively, with $L_i \neq 0$ for some $i \geq 1$;

O.3) there exists $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying Karush-Kuhn-Tucker conditions for (11),

$$\nabla f(x) + \nabla g(x) y = 0, \quad g(x) \leq 0, \quad y \geq 0, \quad \langle y, g(x)\rangle = 0. \tag{12}$$

The *canonical Lagrangian* of problem (11) $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the corresponding *saddle-point operator* $\mathbf{S} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ are, respectively,

$$\mathscr{L}(x, y) := f(x) + \langle y, g(x)\rangle, \quad \mathbf{S}(x, y) := \begin{bmatrix} \nabla_x \mathscr{L}(x, y) \\ -\nabla_y \mathscr{L}(x, y) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + \nabla g(x) y \\ -g(x) \end{bmatrix}. \tag{13}$$

The *normal cone operator* of $\mathbb{R}^n \times \mathbb{R}^m_+$, $\mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+} : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$, is the subdifferential of the indicator function of this set $\delta_{\mathbb{R}^n\times\mathbb{R}^m_+} : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, that is,

$$\delta_{\mathbb{R}^n\times\mathbb{R}^m_+}(x,y) := \begin{cases} 0, & \text{if } y \geq 0; \\ \infty & \text{otherwise}, \end{cases} \qquad \mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+} := \partial\delta_{\mathbb{R}^n\times\mathbb{R}^m_+}. \tag{14}$$

Next we review some reformulations of (12).

**Proposition 2.1.** *The point-to-set operator* $\mathbf{S} + \mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+}$ *is maximal monotone and for any* $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ *the following conditions are equivalent:*

(a) $\nabla f(x) + \nabla g(x) y = 0$, $g(x) \leq 0$, $y \geq 0$, *and* $\langle y, g(x)\rangle = 0$*;*

(b) $(x,y)$ *is a solution of the saddle-point problem* $\max_{y\in\mathbb{R}^m_+} \min_{x\in\mathbb{R}^n} \ f(x) + \langle y, g(x)\rangle$*;*

(c) $(x,y)$ *is a solution of the complementarity problem*

$$(x,y) \in \mathbb{R}^n \times \mathbb{R}^m; \ w \in \mathbb{R}^m; \ \ \mathbf{S}(x,y) - (0,w) = 0; \ \ y, w \geq 0; \ \ \langle y, w\rangle = 0;$$

(d) $(x,y)$ *is a solution of the monotone inclusion problem* $0 \in \left(\mathbf{S} + \mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+}\right)(x,y)$*.*

Next we review some reformulations of the saddle-point problem in (2).

**Proposition 2.2.** *Take* $(\mathring{x}, \mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ *and* $\lambda > 0$*. The following conditions are equivalent:*

(a) $(x,y)$ *is the solution of the regularized saddle-point problem*

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m_+} \ f(x) + \langle y, g(x)\rangle + \frac{1}{2\lambda}\left(\|x - \mathring{x}\|^2 - \|y - \mathring{y}\|^2\right);$$

(b) $(x,y)$ *is the solution of the regularized complementarity problem*

$$(x,y) \in \mathbb{R}^n \times \mathbb{R}^m; \ \ \lambda\Big(\mathbf{S}(x,y) - (0,w)\Big) + (x,y) - (\mathring{x}, \mathring{y}) = 0; \ \ y, w \geq 0; \ \ \langle y, w\rangle = 0;$$

(c) $(x,y) = \left(\lambda(\mathbf{S} + \mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+}) + \mathrm{I}\right)^{-1}(\mathring{x}, \mathring{y})$*.*

It follows from Propositions 2.1 and 2.2 that (12) is equivalent to the monotone inclusion problem

$$0 \in (\mathbf{S} + \mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+})(x,y)$$

and that (2) is the PPM iteration for this inclusion problem. Therefore, the convergence analysis of the Rockafellar's Proximal Method of Multipliers (PMM) follows from Rockafellar's classical convergence analysis of the PPM.

## 3 An error measure for regularized saddle-point problems

We will present a modification of Rockafellar's PMM which uses approximate solutions of the regularized saddle-point problem (2) satisfying a relative error tolerance. To that effect, in this section we define a generic instance of problem (2), define an error measure for approximate solutions of this generic instance, and analyze some properties of the proposed error measure.

Consider, for $\lambda > 0$ and $(\mathring{x}, \mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, a generic instance of the regularized saddle-point problem to be solved in each iteration of Rockafellar's PMM,

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m_+} \; f(x) + \langle y, g(x)\rangle + \frac{1}{2\lambda}(\|x-\mathring{x}\|^2 - \|y-\mathring{y}\|^2). \tag{15}$$

Define for $\lambda \in \mathbb{R}$ and $(\mathring{x}, \mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$

$$\begin{aligned}
&\Psi_{\mathbf{S},(\mathring{x},\mathring{y}),\lambda} : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R},\\
&\Psi_{\mathbf{S},(\mathring{x},\mathring{y}),\lambda}(x,y) := \min_{w\in\mathbb{R}^m_+} \left\| \lambda\Big(\mathbf{S}(x,y) - (0,w)\Big) + (x,y) - (\mathring{x},\mathring{y}) \right\|^2 + 2\lambda\langle y, w\rangle.
\end{aligned} \tag{16}$$

For $\lambda > 0$, this function is trivially an error measure for the complementarity problem on Proposition 2.2 (b), a problem which is equivalent to (15), by Proposition 2.2 (a); hence, $\Psi_{\mathbf{S},(\mathring{x},\mathring{y}),\lambda}(x,y)$ is an error measure for (15).

In the context of complementarity problems, the quantity $\langle y, w\rangle$ in (16) is refered to as the *complementarity gap*. Next we show that the complementarity gap is related to the $\varepsilon$-subdifferential of $\delta_{\mathbb{R}^n\times\mathbb{R}^m_+}$ and to the $\varepsilon$-enlargement of the normal cone operator of $\mathbb{R}^n \times \mathbb{R}^m_+$. Direct use of (8) and of the definition of the $\varepsilon$-subdifferential [1] yields

$$\begin{aligned}
&\forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^m_+,\ \varepsilon \geq 0\\
&\partial_\varepsilon \delta_{\mathbb{R}^n\times\mathbb{R}^m_+}(x,y) = \mathbf{N}^{[\varepsilon]}_{\mathbb{R}^n\times\mathbb{R}^m_+}(x,y) = \left\{ -(0,w) \mid w \in \mathbb{R}^m_+,\ \langle y, w\rangle \leq \varepsilon \right\}.
\end{aligned} \tag{17}$$

Since

$$\operatorname*{argmin}_{w\in\mathbb{R}^m_+} \|\lambda\big(\mathbf{S}(x,y) - (0,w)\big) + (x,y) - (\mathring{x},\mathring{y})\|^2 + 2\lambda\langle y, w\rangle = (g(x) + \lambda^{-1}\mathring{y})_-, \tag{18}$$

it follows from definition (16) that

$$\begin{aligned}
\Psi_{\mathbf{S},(\mathring{x},\mathring{y}),\lambda}(x,y) &= \|\lambda(\nabla f(x) + \nabla g(x) y) + x - \mathring{x}\|^2 + \|y - (\lambda g(x) + \mathring{y})_+\|^2 + 2\langle y, (\lambda g(x) + \mathring{y})_-\rangle\\
&= \|\lambda \mathbf{S}(x,y) + (x,y) - (\mathring{x},\mathring{y})\|^2 - \|(\lambda g(x) + \mathring{y})_-\|^2
\end{aligned} \tag{19}$$

for any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m_+$.

**Lemma 3.1.** *If $\lambda > 0$, $\mathring{z} = (\mathring{x}, \mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, $z = (x,y) \in \mathbb{R}^m \times \mathbb{R}^m_+$ and $w, v, \varepsilon$ are defined as*

$$w := (g(x) + \lambda^{-1}\mathring{y})_-, \qquad v := \mathbf{S}(z) - (0,w), \qquad \varepsilon := \langle y, w\rangle,$$

*then*

(a) $-(0,w) \in \partial_\varepsilon \delta_{\mathbb{R}^n\times\mathbb{R}^m_+}(z) = \mathbf{N}^{[\varepsilon]}_{\mathbb{R}^n\times\mathbb{R}^m_+}(z)$;

(b) $v \in (\mathbf{S}+\mathbf{N}^{[\varepsilon]}_{\mathbb{R}^n\times\mathbb{R}^m_+})(z) \subset (\mathbf{S}+\mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+})^{[\varepsilon]}(z)$, $\|\lambda v + z - \mathring{z}\|^2 + 2\lambda\varepsilon = \Psi_{\mathbf{S},\mathring{z},\lambda}(z)$;

(c) $\left\| z - \left(\lambda(\mathbf{S}+\mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+}) + I\right)^{-1}(\mathring{z}) \right\| \le \sqrt{\Psi_{\mathbf{S},\mathring{z},\lambda}(z)}$.

*Proof.* Item (a) follows trivially from the definitions of $w$ and $\varepsilon$, and (17). The first inclusion in item (b) follows from the definition of $v$ and item (a); the second inclusion follows from direct calculations and (8); the identity in item (b) follows from the definitions of $w$ and $\varepsilon$, (16) and (18). Finally, item (c) follows from item (b) and Lemma 1.3. □

Now we will show how to update $\lambda$ so that $\Psi_{\mathbf{S},\mathring{z},\lambda}(x,y)$ does not increase when $\mathring{z}$ is updated like $z_{k-1}$ is updated to $z_k$ in (9).

**Proposition 3.2.** *Suppose that $\lambda > 0$, $\mathring{z} = (\mathring{x},\mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, $z = (x,y) \in \mathbb{R}^n \times \mathbb{R}^m_+$ and define*

$$w := (g(x) + \lambda^{-1}\mathring{y})_-, \qquad v := \mathbf{S}(z) - (0,w), \qquad \mathring{z}(\tau) := \mathring{z} - \tau\lambda v =: (\mathring{x}(\tau), \mathring{y}(\tau)).$$

*For any $\tau \in [0,1]$,*

$$\begin{aligned} &\mathring{x}(\tau) = \mathring{x} - \tau\lambda(\nabla f(x) + \nabla g(x)y), \quad \mathring{y}(\tau) = \mathring{y} + \tau\lambda(g(x) + (g(x)+\lambda^{-1}\mathring{y})_-)\,, \\ &\Psi_{\mathbf{S},\mathring{z}(\tau),(1-\tau)\lambda}(z) \le \Psi_{\mathbf{S},\mathring{z},\lambda}(z). \end{aligned}$$

*If, additionally, $\mathring{y} \ge 0$ then, for any $\tau \in [0,1]$, $\mathring{y}(\tau) \ge 0$.*

*Proof.* Direct use of the definitions of $w$, $v$ and $\mathring{z}(\tau)$ yields the expressions for $\mathring{x}(\tau)$ and $\mathring{y}(\tau)$ as well as the identity

$$(1-\tau)\lambda\left(\mathbf{S}(z) - (0,w)\right) + z - \mathring{z}(\tau) = \lambda\left(\mathbf{S}(z) - (0,w)\right) + z - \mathring{z},$$

which, in turn, combined with (16) gives, for any $\tau \in [0,1]$,

$$\begin{aligned} \Psi_{\mathbf{S},\mathring{z}(\tau),(1-\tau)\lambda}(z) &\le \|(1-\tau)\lambda\left(\mathbf{S}(z)-(0,w)\right) + z - \mathring{z}(\tau)\|^2 + 2(1-\tau)\lambda\langle y, w\rangle \\ &= \|\lambda\left(\mathbf{S}(z)-(0,w)\right) + z - \mathring{z}\|^2 + 2(1-\tau)\lambda\langle y,w\rangle \qquad \le \Psi_{\mathbf{S},\mathring{z},\lambda}(z), \end{aligned}$$

where the second inequality follows from (16), (18), the assumption $\tau \in [0,1]$ and the definition of $w$. To prove the second part of the proposition, observe that, for any $\tau \in [0,1]$, $\mathring{y}(\tau)$ is a convex combination of $\mathring{y}$ and $\mathring{y}(1) = (\lambda g(x) + \mathring{y})_+$. □

The next lemma and the next proposition provide quantitative and qualitative estimations of the dependence of $\Psi_{\mathbf{S},(\mathring{x},\mathring{y}),\lambda}(x,y)$ on $\lambda$.

**Lemma 3.3.** *If $\psi(\lambda) := \Psi_{\mathbf{S},\mathring{z},\lambda}(z)$ for $\lambda \in \mathbb{R}$, where $\mathring{z} = (\mathring{x},\mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, and $z = (x,y) \in \mathbb{R}^n \times \mathbb{R}^m_+$, then*

(a) *$\psi$ is convex, differentiable and piecewise quadratic;*

(b) $\dfrac{d}{d\lambda}\psi(\lambda) = 2\left(\langle \mathbf{S}(z), \lambda(\mathbf{S}(z) - (0,w)) + z - \mathring{z}\rangle\right)$ *where $w = (g(x) + \lambda^{-1}\mathring{y})_-$ ;*

(c) $\psi(\lambda) \le (\|\lambda\mathbf{S}(z)\| + \|z - \mathring{z}\|)^2$*;*

(d) $\lim_{\lambda\to\infty}\psi(\lambda) < \infty$ *if and only if $(x,y)$ is a solution of (12).*

*Proof.* The proof follows trivially from (19). □

**Proposition 3.4.** *If* $\mathring{z} = (\mathring{x}, \mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$ *and* $0 < \mu \leq \lambda$ *then*

$$\sqrt{\Psi_{\mathbf{S},\mathring{z},\lambda}(z)} \leq \frac{\lambda}{\mu}\sqrt{\Psi_{\mathbf{S},\mathring{z},\mu}(z)} + \frac{\lambda-\mu}{\mu}\|z - \mathring{z}\|.$$

*Proof.* Let $w := (g(x) + \mu^{-1}\mathring{y})_-$ and

$$r_\mu := \mu\Big(\mathbf{S}(z) - (0, w)\Big) + z - \mathring{z}, \;\; r_\lambda := \lambda\Big(\mathbf{S}(z) - (0, w)\Big) + z - \mathring{z}.$$

It follows from the latter definitions, (16) and (18) that $\Psi_{\mathbf{S},\mathring{z},\mu}(z) = \|r_\mu\|^2 + 2\mu\langle y, w\rangle$ and

$$\begin{aligned}\Psi_{\mathbf{S},\mathring{z},\lambda}(z) \leq \|r_\lambda\|^2 + 2\lambda\langle y, w\rangle &= \left\|\frac{\lambda}{\mu}r_\mu + \frac{\mu-\lambda}{\mu}(z-\mathring{z})\right\|^2 + \frac{\lambda}{\mu}2\mu\langle y, w\rangle\\ &\leq \left(\frac{\lambda}{\mu}\right)^2\left(\|r_\mu\|^2 + 2\mu\langle y, w\rangle\right) + 2\frac{\lambda}{\mu}\frac{\lambda-\mu}{\mu}\|r_\mu\|\|z-\mathring{z}\| + \left(\frac{\lambda-\mu}{\mu}\|z-\mathring{z}\|\right)^2\\ &\leq \left(\frac{\lambda}{\mu}\right)^2\Psi_{\mathbf{S},\mathring{z},\mu}(z) + 2\frac{\lambda}{\mu}\sqrt{\Psi_{\mathbf{S},\mathring{z},\mu}(z)}\frac{\lambda-\mu}{\mu}\|z-\mathring{z}\| + \left(\frac{\lambda-\mu}{\mu}\|z-\mathring{z}\|\right)^2,\end{aligned}$$

where the first inequality follows from the assumption $0 < \mu \leq \lambda$. The conclusion follows trivially from the latter inequality. □

# 4 Quadratic approximations of the smooth convex programming problem

In this section we use second-order approximations of $f$ and $g$ around a point $\tilde{x} \in \mathbb{R}^n$ to define a second-order approximation of problem (11) around such a point. We also define a local model of (2), where second-order approximations of $f$ and $g$ around $\tilde{x}$ substitute these functions, and give conditions on a point $(\tilde{x}, \tilde{y})$ under which a solution of the local model is a better approximation to the solution of (2) than this point.

For $\tilde{x} \in \mathbb{R}^n$, let $f_{[\tilde{x}]}$ and $g_{[\tilde{x}]} = (g_{1,[\tilde{x}]}, \ldots, g_{m,[\tilde{x}]})$ be the quadratic approximations of $f$ and $g = (g_1, \ldots, g_m)$ around $\tilde{x}$, that is,

$$\begin{aligned} f_{[\tilde{x}]}(x) &:= f(\tilde{x}) + \nabla f(\tilde{x})^T(x-\tilde{x}) + \frac{1}{2}(x-\tilde{x})^T\nabla^2 f(\tilde{x})(x-\tilde{x})\\ g_{i,[\tilde{x}]}(x) &:= g_i(\tilde{x}) + \nabla g_i(\tilde{x})^T(x-\tilde{x}) + \frac{1}{2}(x-\tilde{x})^T\nabla^2 g_i(\tilde{x})(x-\tilde{x}), \quad i = 1, \ldots, m.\end{aligned} \tag{20}$$

We define

$$(P_{[\tilde{x}]}) \qquad \min f_{[\tilde{x}]}(x) \qquad \text{s.t. } \; g_{[\tilde{x}]}(x) \leq 0 \tag{21}$$

as the quadratic approximation of problem (11) around $\tilde{x}$. The canonical Lagrangian of (21), $\mathscr{L}_{[\tilde{x}]} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and the corresponding saddle-point operator, $\mathbf{S}_{[\tilde{x}]} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$, are, respectively,

$$\begin{aligned}\mathscr{L}_{[\tilde{x}]}(x, y) &:= f_{[\tilde{x}]}(x) + \langle y, g_{[\tilde{x}]}(x)\rangle,\\ \mathbf{S}_{[\tilde{x}]}(x, y) &:= \begin{bmatrix}\nabla_x\mathscr{L}_{[\tilde{x}]}(x, y)\\ -\nabla_y\mathscr{L}_{[\tilde{x}]}(x, y)\end{bmatrix} = \begin{bmatrix}\nabla f_{[\tilde{x}]}(x) + \nabla g_{[\tilde{x}]}(x)y\\ -g_{[\tilde{x}]}(x)\end{bmatrix}.\end{aligned} \tag{22}$$

Since $\mathscr{L}_{[\tilde{x}]}(x,y)$ is a 3rd-degree polynomial in $(x,y)$ and the components of $\mathbf{S}_{[\tilde{x}]}(x,y)$ are 2nd-degree polynomials in $(x,y)$, neither $\mathscr{L}_{[\tilde{x}]}$ is a quadratic approximation of $\mathscr{L}$ nor $\mathbf{S}_{[\tilde{x}]}$ is a linear approximation of $\mathbf{S}$; nevertheless, this 3-rd degree functional and that componentwise quadratic operator are, respectively, the canonical Lagrangian and the associated saddle-point operator of a quadratic approximation of $(P)$ around $\tilde{x}$, namely, $(P_{[\tilde{x}]})$. So, we may say that $\mathscr{L}_{[\tilde{x}]}$ and $\mathbf{S}_{[\tilde{x}]}$ are approximations of $\mathscr{L}$ and $\mathbf{S}$ *based on quadratic approximations* of $f$ and $g$.

Each iteration of Rockafellar's PMM *applied to problem* $(P_{[\tilde{x}]})$ requires the solution of an instance of the generic regularized saddle-point problem

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m_+}\ f_{[\tilde{x}]}(x)+\big\langle y, g_{[\tilde{x}]}(x)\big\rangle+\frac{1}{2\lambda}(\|x-\mathring{x}\|^2-\|y-\mathring{y}\|^2), \tag{23}$$

where $\lambda>0$ and $(\mathring{x},\mathring{y})\in\mathbb{R}^n\times\mathbb{R}^m$. It follows from Proposition 2.2 that this problem is equivalent to the complementarity problem

$$(x,y)\in\mathbb{R}^n\times\mathbb{R}^m_+;\quad \lambda\mathbf{S}_{[\tilde{x}]}(x,y)+(x,y)-(\mathring{x},\mathring{y})=(0,w);\quad y,w\geq 0;\quad \langle y,w\rangle=0.$$

To analyze the error of substituting $\mathbf{S}$ by $\mathbf{S}_{[\tilde{x}]}$ we introduce the notation:

$$L_g=(L_1,\dots,L_m);\quad |(y_1,\dots,y_m)|=(|y_1|,\dots,|y_m|),\ \ (y_1,\dots,y_m)\in\mathbb{R}^m. \tag{24}$$

**Lemma 4.1.** *For any $(x,y)\in\mathbb{R}^n\times\mathbb{R}^m$ and $\tilde{x}\in\mathbb{R}^n$*

$$\big\|\mathbf{S}(x,y)-\mathbf{S}_{[\tilde{x}]}(x,y)\big\|\leq\frac{L_0+\langle L_g,|y|\rangle}{2}\|x-\tilde{x}\|^2+\frac{\|L_g\|}{6}\|x-\tilde{x}\|^3.$$

*Proof.* It follows from triangle inequality, (20) and assumption (O.2) that

$$\begin{aligned}\|\nabla_x\mathscr{L}_{[\tilde{x}]}(x,y)-\nabla_x\mathscr{L}(x,y)\|&\leq\|\nabla f_{[\tilde{x}]}(x)-\nabla f(x)\|+\|(\nabla g_{[\tilde{x}]}(x)-\nabla g(x))y\|\\&\leq\left(\frac{L_0}{2}+\sum_{i=1}^m|y_i|\frac{L_i}{2}\right)\|x-\tilde{x}\|^2\end{aligned}$$

and

$$\|g_{[\tilde{x}]}(x)-g(x)\|=\sqrt{\sum_{i=1}^m\big(g_{i,[\tilde{x}]}(x)-g_i(x)\big)^2}\leq\sqrt{\sum_{i=1}^m\left(\frac{L_i}{6}\|x-\tilde{x}\|^3\right)^2}=\frac{\|L_g\|}{6}\|x-\tilde{x}\|^3.$$

To end the proof, use the above inequalities, (13) and (22). □

Define, for $(\mathring{x},\mathring{y})\in\mathbb{R}^n\times\mathbb{R}^m$ and $\lambda>0$,

$$\mathcal{N}_\theta((\mathring{x},\mathring{y}),\lambda):=\left\{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m_+\ \middle|\ \begin{array}{l}\lambda\left(\dfrac{L_0+\langle L_g,|y|\rangle}{2}+\dfrac{2\|L_g\|}{3}\rho\right)\rho\leq\theta,\\ \text{where } \rho=\sqrt{\Psi_{\mathbf{S},(\mathring{x},\mathring{y}),\lambda}(x,y)}\end{array}\right\}. \tag{25}$$

The next proposition shows that, if $(\tilde{x},\tilde{y})\in\mathcal{N}_\theta((\mathring{x},\mathring{y}),\lambda)$ with $\theta\leq 1/4$, then the solution of the regularized saddle-point problem (23) is a better approximation than $(\tilde{x},\tilde{y})$ to the solution of the regularized saddle-point problem (15), with respect to the merit function $\Psi_{\mathbf{S},(x,y),\lambda}$.

**Proposition 4.2.** *If* $\lambda > 0$, $(\mathring{x}, \mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, $(\tilde{x}, \tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ *and*

$$(x, y) = \arg \min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m_+} \; f_{[\tilde{x}]}(x) + \left\langle y, g_{[\tilde{x}]}(x) \right\rangle + \frac{1}{2\lambda} \left( \|x - \mathring{x}\|^2 - \|y - \mathring{y}\|^2 \right),$$

*then*

$$\|(\tilde{x}, \tilde{y}) - (x, y)\| \leq \sqrt{\Psi_{\mathbf{S}, (\mathring{x}, \mathring{y}), \lambda}(\tilde{x}, \tilde{y})}.$$

*If, additionally,* $(\tilde{x}, \tilde{y}) \in \mathcal{N}_\theta((\mathring{x}, \mathring{y}), \lambda)$ *with* $0 \leq \theta \leq 1/4$, *then*

$$\sqrt{\Psi_{\mathbf{S}, (\mathring{x}, \mathring{y}), \lambda}(x, y)} \leq \theta \sqrt{\Psi_{\mathbf{S}, (\mathring{x}, \mathring{y}) \lambda}(\tilde{x}, \tilde{y})}$$

*and* $(x, y) \in \mathcal{N}_{\theta^2}((\mathring{x}, \mathring{y}), \lambda)$.

*Proof.* Applying Lemma 3.1 to (23) and using (22) we conclude that

$$\|(\tilde{x}, \tilde{y}) - (x, y)\| \leq \sqrt{\Psi_{\mathbf{S}_{[\tilde{x}]}, (\mathring{x}, \mathring{y}), \lambda}(\tilde{x}, \tilde{y})}.$$

It follows from (20), (22), and (13) that $\mathbf{S}_{[\tilde{x}]}(\tilde{x}, \tilde{y}) = \mathbf{S}(\tilde{x}, \tilde{y})$, which, combined with (16), implies that $\Psi_{\mathbf{S}_{[\tilde{x}]}, (\mathring{x}, \mathring{y}), \lambda}(\tilde{x}, \tilde{y}) = \Psi_{\mathbf{S}(\mathring{x}, \mathring{y}), \lambda}(\tilde{x}, \tilde{y})$. To prove the first part of the proposition, combine this result with the above inequality.

To simplify the proof of the second part of the proposition, define

$$\tilde{\rho} = \sqrt{\Psi_{\mathbf{S}_{[\tilde{x}]}, (\mathring{x}, \mathring{y}), \lambda}(\tilde{x}, \tilde{y})}, \qquad w = (g(x) + \lambda^{-1}\mathring{y})_-, \;\; r = \lambda \Big( \mathbf{S}(x, y) - (0, w) \Big) + (x, y) - (\mathring{x}, \mathring{y}),$$

Since $(x, y)$ is the solution of (23),

$$\lambda \Big( \mathbf{S}_{[\tilde{x}]}(x, y) - (0, w) \Big) + (x, y) - (\mathring{x}, \mathring{y}) = 0, \quad y, w \geq 0, \quad \langle y, w \rangle = 0.$$

Therefore, $r = \lambda(\mathbf{S}(x, y) - \mathbf{S}_{[\tilde{x}]}(x, y))$. Using also (16), Lemma 4.1 and the first part of the proposition we conclude that

$$\sqrt{\Psi_{\mathbf{S}, (\mathring{x}, \mathring{y}), \lambda}(x, y)} \leq \sqrt{\|r\|^2 + 2\lambda \langle y, w \rangle} = \lambda \left\| \mathbf{S}(x, y) - \mathbf{S}_{[\tilde{x}]}(x, y) \right\| \leq \lambda \left( \frac{L_0 + \langle L_g, |y| \rangle}{2} + \frac{\|L_g\|}{6} \tilde{\rho} \right) \tilde{\rho}^2.$$

Moreover, it follows from the Cauchy-Schwarz inequality, the first part of the proposition and the definition of $\tilde{\rho}$ that

$$\langle L_g, |y| \rangle \leq \langle L_g, |\tilde{y}| \rangle + \|L_g\| \|y - \tilde{y}\| \leq \langle L_g, |\tilde{y}| \rangle + \|L_g\| \tilde{\rho}.$$

Therefore

$$\sqrt{\Psi_{\mathbf{S}, (\mathring{x}, \mathring{y}), \lambda}(x, y)} \leq \lambda \left( \frac{L_0 + \langle L_g, |\tilde{y}| \rangle}{2} + \frac{2\|L_g\|}{3} \tilde{\rho} \right) \tilde{\rho}^2.$$

Suppose that $(\tilde{x}, \tilde{y}) \in \mathcal{N}_\theta((\mathring{x}, \mathring{y}), \lambda)$ with $0 \leq \theta \leq 1/4$. It follows trivially from this assumption, (25), the definition of $\tilde{\rho}$, and the above inequality, that the inequality in the second part of the proposition holds. To end the proof of the second part, let $\rho = \sqrt{\Psi_{\mathbf{S}_{[\tilde{x}]},(\mathring{x},\mathring{y}),\lambda}(x,y)}$. Since $\rho \le \theta\tilde{\rho} \le \tilde{\rho}/4$ and $\langle L_g, |y|\rangle \le \langle L_g, |\tilde{y}|\rangle + \|L_g\|\tilde{\rho}$,

$$\lambda\left(\frac{L_0 + \langle L_g, |y|\rangle}{2} + \frac{2\|L_g\|}{3}\rho\right)\rho \le \lambda\left(\frac{L_0 + \langle L_g, |\tilde{y}|\rangle + \|L_g\|\tilde{\rho}}{2} + \frac{2\|L_g\|}{3}\frac{\tilde{\rho}}{4}\right)\theta\rho$$
$$= \lambda\left(\frac{L_0 + \langle L_g, |\tilde{y}|\rangle}{2} + \frac{2\|L_g\|}{3}\tilde{\rho}\right)\theta\tilde{\rho} \le \theta^2,$$

where the last inequality follows from the assumption $(\tilde{x},\tilde{y}) \in \mathcal{N}_\theta((\mathring{x},\mathring{y}),\lambda)$ and (25). To end the proof use the definition of $\rho$, the above inequality and (25). □

In view of the preceding proposition, for a given $(\mathring{x},\mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ and $\theta > 0$, it is natural to search for $\lambda > 0$ and $(x,y) \in \mathcal{N}_\theta((\mathring{x},\mathring{y}),\lambda)$.

**Proposition 4.3.** *For any $(\mathring{x},\mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m_+$, and $\theta > 0$ there exists $\bar{\lambda} > 0$ such that $(x,y) \in \mathcal{N}_\theta((\mathring{x},\mathring{y}),\lambda)$ for any $\lambda \in (0,\bar{\lambda}]$.*

*Proof.* The proof follows from the definition (25) and Lemma 3.3(c). □

The neighborhoods $\mathcal{N}_\theta$ as well as the next defined function will be instrumental in the definition and analysis of Algorithm 1, to be presented in the next section.

**Definition 4.4.** *For $\alpha > 0$ and $y \in \mathbb{R}^m$, $\rho(y,\alpha)$ stands for the largest root of*

$$\left(\frac{L_0 + \langle L_g, |y|\rangle}{2} + \frac{2\|L_g\|}{3}\rho\right)\rho = \alpha.$$

Observe that for any $\lambda, \theta > 0$ and $(\mathring{x},\mathring{y}) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$\mathcal{N}_\theta((\mathring{x},\mathring{y}),\lambda) = \left\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m_+ \;\middle|\; \sqrt{\Psi_{\mathbf{S},(\mathring{x},\mathring{y}),\lambda}(x,y)} \le \rho\left(y, \theta/\lambda\right)\right\}. \tag{26}$$

Moreover, since $\rho(y,\alpha)$ is the largest root of a quadratic it follows that it has an explicit formula.

## 5 A relaxed hybrid proximal extragradient method of multipliers based on second-order approximations

In this section we consider the smooth convex programming problem (11) where assumptions O.1, O.2 and O.3 are assumed to hold. Aiming at finding approximate solutions of the latter problem, we propose a new method, called (relaxed) hybrid proximal extragradient method of multipliers based on quadratic approximations (hereafter rHPEMM-2o), which is a modification of Rockafellar's PMM in the following senses: in each iteration either a relaxed extragradient step is executed or a second order approximation of (15) is solved. More specifically, each iteration $k$ uses the (available) variables

$$(x_{k-1}, y_{k-1}), \quad (\tilde{x}_k, \tilde{y}_k) \in \mathbb{R}^n \times \mathbb{R}^m_+, \text{ and } \lambda_k > 0$$

to generate

$$(x_k, y_k), \quad (\tilde{x}_{k+1}, \tilde{y}_{k+1}) \in \mathbb{R}^n \times \mathbb{R}^m_+, \text{ and } \lambda_{k+1} > 0$$

in one of two ways. Either
(1) $(x_k, y_k)$ is obtained from $(x_{k-1}, y_{k-1})$ via a relaxed extragradient step,

$$(x_k, y_k) = (x_{k-1}, y_{k-1}) - \tau\lambda_k v_k, \quad v_k \in (\mathbf{S} + \mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+})^{\varepsilon_k}(\tilde{x}_k, \tilde{y}_k)$$

in which case $(\tilde{x}_{k+1}, \tilde{y}_{k+1}) = (\tilde{x}_k, \tilde{y}_k)$ and $\lambda_{k+1} < \lambda_k$; or
(2) $(x_k, y_k) = (x_{k-1}, y_{k-1})$ and the point $(\tilde{x}_{k+1}, \tilde{y}_{k+1})$ is the outcome of one iteration (at $(x_k, y_k)$) of Rockafellar's PMM for problem (21) with $\tilde{x} = \tilde{x}_k$ and $\lambda = \lambda_{k+1}$.

Next we present our algorithm, where $\mathcal{N}_\theta$, $f_{[\tilde{x}]}, g_{[\tilde{x}]}$ and $\rho(y, \alpha)$ are as in (25), (20), and Definition 4.4, respectively.

---

**Algorithm 1:** Relaxed hybrid proximal extragradient method of multipliers based on 2nd ord. approx. (r-HPEMM-2o)

---

**initialization:** choose $(x_0, y_0) = (\tilde{x}_1, \tilde{y}_1) \in \mathbb{R}^n \times \mathbb{R}^m_+$, $0 < \sigma < 1$, $0 < \theta \leq 1/4$;
define $h :=$ positive root of $\theta(1 + h')\left(1 + h'\left(1 + 1/\sigma\right)\right)^2 = 1$, $\tau = h/(1 + h)$;
choose $\lambda_1 > 0$ such that $(\tilde{x}_1, \tilde{y}_1) \in \mathcal{N}_{\theta^2}((x_0, y_0), \lambda_1)$ and set $k \leftarrow 1$

**1** **if** $\rho(\tilde{y}_k, \theta^2/\lambda_k) \leq \sigma\|(\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1})\|$ **then**
**2** $\quad\lambda_{k+1} := (1 - \tau)\lambda_k$;
**3** $\quad(\tilde{x}_{k+1}, \tilde{y}_{k+1}) := (\tilde{x}_k, \tilde{y}_k)$;
**4** $\quad x_k := x_{k-1} - \tau\lambda_k[\nabla f(\tilde{x}_k) + \nabla g(\tilde{x}_k)\tilde{y}_k], \quad y_k := y_{k-1} + \tau[\lambda_k g(\tilde{x}_k) + (\lambda_k g(\tilde{x}_k) + y_{k-1})_-]$;
**5** **else**
**6** $\quad\lambda_{k+1} := (1 - \tau)^{-1}\lambda_k$;
**7** $\quad(x_k, y_k) := (x_{k-1}, y_{k-1})$;
**8** $\quad(\tilde{x}_{k+1}, \tilde{y}_{k+1}) := \arg\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m_+} \; f_{[\tilde{x}_k]}(x) + \langle y, g_{[\tilde{x}_k]}(x)\rangle + \dfrac{\|x - x_k\|^2 - \|y - y_k\|^2}{2\lambda_{k+1}}$;
**9** **end if**
**10** set $k \leftarrow k + 1$ and go to step 1;

---

To simplify the presentation of Algorithm 1, we have omitted a stopping test. First we discuss its initialization. In the absence of additional information on the dual variables $y$, one shall consider the initialization

$$(x_0, y_0) = (\tilde{x}_1, \tilde{y}_1) = (x, 0), \tag{27}$$

where $x$ is "close" to the feasible set. If $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$ an approximated solution of (12) is available, one can do a "warm start" by setting $(x_0, y_0) = (\tilde{x}_1, \tilde{y}_1) = (x, y)$. Note that $h > 0$ and $0 < \tau < 1$. Existence of $\lambda_1 > 0$ as prescribed in this step follows from the inclusion $(\tilde{x}_1, \tilde{y}_1) \in \mathbb{R}^n \times \mathbb{R}^m_+$ and from Proposition 4.3. Moreover, if we compute $\lambda = \lambda_1 > 0$ satisfying the inequality

$$\left(\frac{2\|L_g\|\|\mathbf{S}(x_0, y_0)\|^2}{3}\right)\lambda^3 + \left(\frac{L_0 + \langle L_g, |y_0|\rangle}{2}\|\mathbf{S}(x_0, y_0)\|\right)\lambda^2 - \theta^2 \leq 0,$$

where the operator $\mathbf{S}$ is defined in (13), use Lemma 3.3 (c) and Definition 4.4 we find $\sqrt{\Psi_{\mathbf{S},(x_0,y_0),\lambda_1}(x_0, y_0)} \leq \lambda_1\|\mathbf{S}(x_0, y_0)\| \leq \rho(y_0, \theta^2/\lambda_1)$, which, in turn, combined with the fact that $(\tilde{x}_1, \tilde{y}_1) = (x_0, y_0)$, gives the inclusion in the initialization of Algorithm 1.

The computational cost of block of steps [**2**,**3**,**4**] is negligible. The initialization $\lambda_1 > 0$, together with the update of $\lambda_k$ by step **2** or **6** guarantee that $\lambda_k > 0$ for all $k$. Therefore, the saddle-point problem to be solved in step **8** is strongly convex-concave and hence has a unique solution. The computational burden of the algorithm is in the computation of the solution of this problem.

We will assume that $(\tilde{x}_1, \tilde{y}_1)$ does not satisfy (12), i.e., the KKT conditions for (11), otherwise we would already have a solution for the KKT system and $\tilde{x}_1$ would be a solution of (11). For the sake of conciseness we introduce, for $k = 1, \dots,$ the notation

$$z_{k-1} = (x_{k-1}, y_{k-1}), \qquad \tilde{z}_k = (\tilde{x}_k, \tilde{y}_k), \qquad \rho_k = \rho(\tilde{y}_k, \theta^2/\lambda_k). \tag{28}$$

Since there are two kinds of iterations in Algorithm 1, its is convenient to have a notation for them. Define

$$A := \{k \in \mathbb{N} \setminus \{0\} \mid \rho_k \leq \sigma \|\tilde{z}_k - z_{k-1}\|\}, \quad B := \{k \in \mathbb{N} \setminus \{0\} \mid \rho_k > \sigma \|\tilde{z}_k - z_{k-1}\|\}. \tag{29}$$

Observe that in iteration $k$, either $k \in A$ and steps **2**, **3**, **4** are executed, or $k \in B$ and steps **6**, **7**, **8** are executed.

**Proposition 5.1.** *For* $k = 1, \dots,$

(a) $\tilde{z}_k \in \mathcal{N}_{\theta^2}(z_{k-1}, \lambda_k)$;

(b) $\sqrt{\Psi_{\mathbf{S}, z_{k-1}, \lambda_k}(\tilde{z}_k)} \leq \rho_k$.

(c) $z_{k-1} \in \mathbb{R}^n \times \mathbb{R}^m_+$ .

*Proof.* We will use induction on $k \geq 1$ for proving (a). In view of the initialization of Algorithm 1, this inclusion holds trivially for $k = 1$. Suppose that this inclusion holds for $k = k_0$. We shall consider two possibilities.

(i) $k_0 \in A$: It follows from Proposition 3.2 and the update rules in steps **2** and **4** that

$$\sqrt{\Psi_{\mathbf{S}, z_{k_0}, \lambda_{k_0+1}}(\tilde{z}_{k_0})} \leq \sqrt{\Psi_{\mathbf{S}, z_{k_0-1}, \lambda_{k_0}}(\tilde{z}_{k_0})} \leq \rho(\tilde{y}_{k_0}, \theta^2/\lambda_{k_0}) \leq \rho(\tilde{y}_{k_0}, \theta^2/\lambda_{k_0+1})$$

where the second inequality follows from the inclusion $\tilde{z}_{k_0} \in \mathcal{N}_{\theta^2}(z_{k_0-1}, \lambda_{k_0})$ and (26); and the third inequality follows from step **2** and Definition 4.4. It follows from the above inequalities and (26) that $\tilde{z}_{k_0} \in \mathcal{N}_{\theta^2}(z_{k_0}, \lambda_{k_0+1})$. By step **3**, $\tilde{z}_{k_0+1} = \tilde{z}_{k_0}$. Therefore, the inclusion of Item (a) holds for $k = k_0 + 1$ in case (i).

(ii) $k_0 \in B$: In this case, by step **7**, $z_{k_0} = z_{k_0-1}$ and, using definition (29), the notation (28), and the assumption that the inclusion in Item (a) holds for $k = k_0$ we conclude that

$$\|\tilde{z}_{k_0} - z_{k_0}\| < \rho_{k_0}/\sigma, \quad \tilde{z}_{k_0} \in \mathcal{N}_{\theta^2}(z_{k_0}, \lambda_{k_0}), \quad \sqrt{\Psi_{\mathbf{S}, z_{k_0}, \lambda_{k_0}}(\tilde{z}_{k_0})} \leq \rho_{k_0}.$$

Direct use of the definitions of $h$, $\tau$, and step **6** gives $\lambda_{k_0+1} = (1+h)\lambda_{k_0}$. Defining $\rho' = \sqrt{\Psi_{\mathbf{S}, z_{k_0}, \lambda_{k_0+1}}(\tilde{z}_{k_0})}$, it follows from the above inequalities and from Proposition 3.4 that,

$$\rho' \leq (1+h)\sqrt{\Psi_{\mathbf{S}, z_{k_0}, \lambda_{k_0}}(\tilde{z}_{k_0})} + h\|\tilde{z}_{k_0} - z_{k_0}\| \leq (1 + h(1 + 1/\sigma))\rho_{k_0}.$$

Therefore,

$$\begin{aligned}\lambda_{k_0+1}\left(\frac{L_0+\langle L_g,|\tilde y_{k_0}|\rangle}{2}+\frac{2\|L_g\|}{3}\rho'\right)\rho' &\le (1+h)\left(1+h\left(1+\frac{1}{\sigma}\right)\right)^2\\ &\quad\times\lambda_{k_0}\left(\frac{L_0+\langle L_g,|\tilde y_{k_0}|\rangle}{2}+\frac{2\|L_g\|}{3}\rho_{k_0}\right)\rho_{k_0}\\ &=(1+h)\left(1+h\left(1+\frac{1}{\sigma}\right)\right)^2\theta^2=\theta,\end{aligned}$$

where we also have used Definition 4.4 and the definition of $h$ (in the initialization of Algorithm 1). It follows from the above inequality, the definition of $\rho'$ and (25) that

$$\tilde z_{k_0}\in\mathcal{N}_\theta(z_{k_0},\lambda_{k_0+1}).$$

Using this inclusion, step **8** and Proposition 4.2 we conclude that the inclusion in Item (a) also holds for $k=k_0+1$.

Item (b) follows trivially from Item (a), (26) and (28). Item (c) follows from the fact that $y_0\ge 0$, the definitions of steps **3**, **4**, and the last part of Proposition 3.2. □

## Algorithm 1 as a realization of the large-step r-HPE Method

In this subsection, we will show that a subsequence generated by Algorithm 1 happens to be a sequence generated by the large-step r-HPE Method described in (9) for solving a monotone inclusion problem associated with (11). This result will be instrumental for evaluating (in the next section) the iteration-complexity of Algorithm 1. In fact, we will prove that iterations with $k\in A$, where steps **2**, **3**, **4** are executed, are large-step r-HPE iterations for the monotone inclusion problem

$$0\in T(z):=\left(\mathbf{S}+\mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+}\right)(z),\qquad z=(x,y)\in\mathbb{R}^n\times\mathbb{R}^m, \tag{30}$$

where the operator $\mathbf{S}$ is defined in (13).

Define, for $k=1,2,\dots,$

$$w_k=(g(\tilde x_k)+\lambda_k^{-1}y_{k-1})_-,\qquad v_k=\mathbf{S}(\tilde z_k)-(0,w_k),\qquad \varepsilon_k=\langle\tilde y_k,w_k\rangle, \tag{31}$$

where $\tilde z_k$ is defined in (28). We will show that, whenever $k\in A$, the variables $\tilde z_k$, $v_k$, and $\varepsilon_k$ provide an approximated solution of the proximal inclusion-equation system

$$v\in(\mathbf{S}+\mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+})(z),\qquad \lambda_k v+z-z_{k-1}=0,$$

as required in the first line of (9). We divided the proof of this fact in two parts, the next proposition and the subsequent lemma.

**Proposition 5.2.** *For $k=1,2,\dots,$*

(a) $-(0,w_k)\in\partial_{\varepsilon_k}\delta_{\mathbb{R}^n\times\mathbb{R}^m_+}(\tilde z_k)=\mathbf{N}^{[\varepsilon_k]}_{\mathbb{R}^n\times\mathbb{R}^m_+}(\tilde z_k)$;

(b) $v_k\in(\mathbf{S}+\mathbf{N}^{[\varepsilon_k]}_{\mathbb{R}^n\times\mathbb{R}^m_+})(\tilde z_k)\subset(\mathbf{S}+\mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+})^{[\varepsilon_k]}(\tilde z_k)$;

(c) $\|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k = \Psi_{\mathbf{S},z_{k-1},\lambda_k}(\tilde{z}_k) \leq \rho_k^2$.

*Proof.* Items (a), (b) and the equality in Item (c) follow from definitions (28), (31) and Items (a) and (b) of Lemma 3.1. The inequality in Item (c) follows from Proposition 5.1(b). □

Define

$$
\begin{aligned}
A_k &:= \{j \in \mathbb{N} \mid j \leq k, \text{ steps } \mathbf{2, 3, 4} \text{ are executed at iteration } j\},\\
B_k &:= \{j \in \mathbb{N} \mid j \leq k, \text{ steps } \mathbf{6, 7, 8} \text{ are executed at iteration } j\}
\end{aligned}
\tag{32}
$$

and observe that

$$
A = \bigcup_{k\in\mathbb{N}} A_k, \qquad B = \bigcup_{k\in\mathbb{N}} B_k.
$$

From now on, $\#C$ stands for the number of elements of a set $C$. To further simplify the converge analysis, define

$$
I = \{i \in \mathbb{N} \mid 1 \leq i \leq \#A\}, \quad k_0 = 0, \quad k_i = i\text{-th element of } A. \tag{33}
$$

Note that $k_0 < k_1 < k_2 \cdots$, $A = \{k_i \mid i \in I\}$ and, in view of (29) and step **7** of Algorithm 1,

$$
z_k = z_{k_{i-1}}, \qquad \text{for} \quad k_{i-1} \leq k < k_i, \;\; \forall i \in I. \tag{34}
$$

In particular, we have

$$
z_{k_i - 1} = z_{k_{i-1}} \qquad \forall i \in I. \tag{35}
$$

In the next lemma we show that for indexes in the set $A$, Algorithm 1 generates a subsequence which can be regarded as a realization of the large-step r-HPE Method described in (9), for solving the problem (30).

**Lemma 5.3.** *The sequences $(z_{k_i})_{i\in I}$, $(\tilde{z}_{k_i})_{i\in I}$, $(v_{k_i})_{i\in I}$, $(\varepsilon_{k_i})_{i\in I}$, $(\lambda_{k_i})_{i\in I}$ are generated by a realization of the* r-HPE Method *described in* (9) *for solving* (30)*, that is,* $0 \in (\mathbf{S}+\mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+})(z)$*, in the following sense: for all $i \in I$,*

$$
\begin{aligned}
& v_{k_i} \in (\mathbf{S} + \mathbf{N}^{[\varepsilon_{k_i}]}_{\mathbb{R}^n\times\mathbb{R}^m_+})(\tilde{z}_{k_i}) \subset (\mathbf{S} + \mathbf{N}_{\mathbb{R}^n\times\mathbb{R}^m_+})^{[\varepsilon_{k_i}]}(\tilde{z}_{k_i}),\\
& \|\lambda_{k_i} v_{k_i} + \tilde{z}_{k_i} - z_{k_{i-1}}\|^2 + 2\lambda_{k_i}\varepsilon_{k_i} \leq \rho_{k_i}^2 \leq \sigma^2 \|\tilde{z}_{k_i} - z_{k_{i-1}}\|^2,\\
& z_{k_i} = z_{k_{i-1}} - \tau\lambda_{k_i} v_{k_i}.
\end{aligned}
\tag{36}
$$

*Moreover, if $I$ is finite and $i_{\mathrm{M}} := \max I$ then $z_k = z_{k_{i_{\mathrm{M}}}}$ for $k \geq k_{i_{\mathrm{M}}}$.*

*Proof.* The two inclusions in the first line of (36) follow trivially from Proposition 5.2(b). The first inequality in the second line of (36) follows from (35) and Proposition 5.2(c); the second inequality follows from the inclusion $k_i \in A$, (29), step **1** of Algorithm 1 and (35). The equality in the last line of (36) follows from the inclusion $k_i \in A$, (29) step **4** of Algorithm 1, (35) and (31). Finally, the last statement of the lemma is a direct consequence of (28), (29) and step **7** of Algorithm 1. □

As we already observed in Proposition 2.1, (30) and (12) are equivalent, in the sense that both problems have the same solution set. From now on we will use the notation $\mathcal{K}$ for this solution set, that is,

$$\begin{aligned}\mathcal{K} &= \left(\mathbf{S} + \mathbf{N}_{\mathbb{R}^n \times \mathbb{R}^m_+}\right)^{-1}(0) \\ &= \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \mid \nabla f(x) + \nabla g(x) y = 0,\ g(x) \le 0,\ y \ge 0,\ \langle y, g(x)\rangle = 0\}\ .\end{aligned} \tag{37}$$

We assumed in (O.3) that this set is nonempty. Let $z^* = (x^*, y^*)$ be the projection of $z_0 = (x_0, y_0)$ onto $\mathcal{K}$ and $d_0$ the distance from $z_0$ to $\mathcal{K}$,

$$z^* \in \mathcal{K}, \quad d_0 = \|z^* - z_0\| = \min_{z \in \mathcal{K}} \|z - z_0\|. \tag{38}$$

To complement Lemma 5.3, we will prove that the *large-step condition* for the large-step r-HPE Method (stated in Theorem 1.2) is satisfied for the realization of the method presented in Lemma 5.3. Define

$$c := \frac{L_0 + \langle L_g, |y_0|\rangle}{2} + \left[\frac{1}{2} + \frac{1/2 + 2\sigma/3}{\sqrt{1-\sigma^2}}\right] d_0 \|L_g\|, \qquad \eta := \frac{\theta^2}{\sigma c}. \tag{39}$$

**Proposition 5.4.** *Let $z^* \in \mathcal{K}$ and $d_0$, and $\eta$ as in* (38)*, and* (39)*, respectively. For all $i \in I$,*

$$\|z^* - z_{k_i}\| \le d_0, \quad \|z^* - \tilde{z}_{k_i}\| \le \frac{d_0}{\sqrt{1-\sigma^2}}, \quad \|\tilde{z}_{k_i} - z_{k_{i-1}}\| \le \frac{d_0}{\sqrt{1-\sigma^2}}. \tag{40}$$

*As a consequence,*

$$\lambda_{k_i} \|\tilde{z}_{k_i} - z_{k_{i-1}}\| \ge \eta. \tag{41}$$

*Proof.* Note first that (40) follows from Lemma 5.3, items (c) and (d) of Proposition 1.1, (35) and (38). Using (28), (29), (33) and step **1** of Algorithm 1 we obtain

$$\sigma \|\tilde{z}_{k_i} - z_{k_i - 1}\| \ge \rho(\tilde{y}_{k_i}, \theta^2/\lambda_{k_i}) \qquad \forall i \in I,$$

which, in turn, combined with the definition of $\rho(\cdot,\cdot)$ (see Definition 4.4) yields

$$\left(\frac{L_0 + \langle L_g, |\tilde{y}_{k_i}|\rangle}{2} + \frac{2\|L_g\|}{3} \sigma \|\tilde{z}_{k_i} - z_{k_i - 1}\|\right) \sigma \|\tilde{z}_{k_i} - z_{k_i - 1}\| \ge \frac{\theta^2}{\lambda_{k_i}} \quad \forall i \in I. \tag{42}$$

Using the triangle inequality, (38) and the second inequality in (40) we obtain

$$\|z_0 - \tilde{z}_{k_i}\| \le d_0 + \|z^* - \tilde{z}_{k_i}\| \le d_0 \left(1 + \frac{1}{\sqrt{1-\sigma^2}}\right).$$

Now, using the latter inequality, the fact that $\|z_0 - \tilde{z}_{k_i}\| \ge \|y_0 - \tilde{y}_{k_i}\|$ ($\forall i \in I$) and the triangle inequality we find

$$\begin{aligned}\langle L_g, |\tilde{y}_{k_i}|\rangle &\le \|L_g\| \|z_0 - \tilde{z}_{k_i}\| + \langle L_g, |y_0|\rangle \\ &\le d_0 \|L_g\| \left(1 + \frac{1}{\sqrt{1-\sigma^2}}\right) + \langle L_g, |y_0|\rangle \qquad \forall i \in I.\end{aligned} \tag{43}$$

To finish the proof of (41), use (35), substitute the terms in the right hand side of the last inequalities in (40) and (43) in the term inside the parentheses in (42) and use (39). □

# 6 Complexity analysis

In this section we study the pointwise and ergodic iteration-complexity of Algorithm 1. The main results are (essentially) a consequence of Lemma 5.3 and Proposition 5.4 which guarantee that the (sub)sequences $(z_{k_i})_{i\in I}$, $(\tilde z_{k_i})_{i\in I}$, ... can be regarded as realizations of the large-step r-HPE method of Section 1, for which pointwise and ergodic iteration-complexity results are known.

To study the ergodic iteration-complexity of Algorithm 1 we need to define the ergodic sequences associated to $(\lambda_{k_i})_{i\in I}$, $(\tilde z_{k_i})_{i\in I}$, $(v_{k_i})_{i\in I}$ and $(\varepsilon_{k_i})_{i\in I}$, respectively (see (10)), namely

$$\begin{aligned}
&\Lambda_i := \tau\sum_{j=1}^{i}\lambda_{k_j}, \qquad \tilde z_i^a = (\tilde x_i^a, \tilde y_i^a) := \frac{1}{\Lambda_i}\,\tau\sum_{j=1}^{i}\lambda_{k_j}\tilde z_{k_j}, \qquad v_i^a := \frac{1}{\Lambda_i}\,\tau\sum_{j=1}^{i}\lambda_{k_j} v_{k_j},\\
&\varepsilon_i^a := \frac{1}{\Lambda_i}\,\tau\sum_{j=1}^{i}\lambda_{k_j}(\varepsilon_{k_j} + \langle \tilde z_{k_j} - \tilde z_i^a, v_{k_j} - v_i^a\rangle).
\end{aligned} \tag{44}$$

Define also

$$\overline{\mathscr{L}}(x,y) := \begin{cases} f(x) + \langle y, g(x)\rangle, & y \geq 0\\ -\infty, & \text{otherwise.}\end{cases} \tag{45}$$

Observe that that a pair $(x,y)\in\mathcal{K}$, i.e., it is a solution of the KKT system (12) if and only if $(0,0)\in\partial(\overline{\mathscr{L}}(\cdot,y) - \overline{\mathscr{L}}(x,\cdot))(x,y)$. Since (30) and (12) are equivalent, the latter observation leads us to consider in this section the notion of approximate solution for (30) which consists in: for given tolerances $\overline{\delta} > 0$ and $\overline{\varepsilon} > 0$ find $((x,y),v,\varepsilon)$ such that

$$v \in \partial_\varepsilon(\overline{\mathscr{L}}(\cdot,y) - \overline{\mathscr{L}}(x,\cdot))(x,y), \quad \|v\| \leq \overline{\delta}, \quad \varepsilon \leq \overline{\varepsilon}. \tag{46}$$

We will also consider as approximate solution of (30) any triple $((x,y),(p,q),\varepsilon)$ such that $\|(p,q)\| \leq \overline{\delta}$, $\varepsilon \leq \overline{\varepsilon}$ and

$$p = \nabla f(x) + \nabla g(x)y, \quad g(x) + q \leq 0, \quad y \geq 0, \quad \langle y, g(x)+q\rangle = -\varepsilon \tag{47}$$

or

$$p \in \partial_{x,\varepsilon'}\overline{\mathscr{L}}(x,y), \quad g(x) + q \leq 0, \quad y \geq 0, \quad \langle y, g(x)+q\rangle \geq -\varepsilon, \tag{48}$$

where $\varepsilon' := \varepsilon + \langle y, g(x)+q\rangle$.

It is worthing to compare the latter two conditions with (12) and also note that whenever $\varepsilon' = 0$ then (48) reduces to (47), that is, the latter condition is a special case of (48). Moreover, as Theorems 6.3 and 6.4 will show, (47) and (48) are related to the pointwise and ergodic iteration-complexity of Algorithm 1, respectively.

We start by studying rates of convergence of Algorithm 1.

**Theorem 6.1.** *Let $(\tilde z_{k_i})_{i\in I} = ((\tilde x_{k_i}, \tilde y_{k_i}))_{i\in I}$, $(v_{k_i})_{i\in I}$ and $(\varepsilon_{k_i})_{i\in I}$ be (sub)sequences generated by* Algorithm 1 *where the the set of indexes $I$ is defined in* (33). *Let also $(\tilde z_i^a)_{i\in I} = ((\tilde x_i^a, \tilde y_i^a))_{i\in I}$, $(v_i^a)_{i\in I}$ and $(\varepsilon_i^a)_{i\in I}$ be as in* (44). *Then, for any $i\in I$,*

(a) **[pointwise]** *there exists* $j \in \{1, \ldots, i\}$ *such that*

$$v_{k_j} \in \partial_{\varepsilon_{k_j}} \left( \overline{\mathscr{L}}(\cdot, \tilde{y}_{k_j}) - \overline{\mathscr{L}}(\tilde{x}_{k_j}, \cdot) \right) (\tilde{x}_{k_j}, \tilde{y}_{k_j}) \tag{49}$$

*and*

$$\|v_{k_j}\| \leq \frac{d_0^2}{i\tau(1-\sigma)\eta}, \qquad \varepsilon_{k_j} \leq \frac{\sigma^2 d_0^3}{(i\tau)^{3/2}(1-\sigma^2)^{3/2} 2\eta}; \tag{50}$$

(b) **[pointwise]** *there exists* $j \in \{1, \ldots, i\}$ *and* $(p_j, q_j) \in \mathbb{R}^n \times \mathbb{R}^m$ *such that*

$$p_j = \nabla f(\tilde{x}_{k_j}) + \nabla g(\tilde{x}_{k_j}) \tilde{y}_{k_j}, \quad g(\tilde{x}_{k_j}) + q_j \leq 0, \quad \tilde{y}_{k_j} \geq 0, \quad \langle \tilde{y}_{k_j}, g(\tilde{x}_{k_j}) + q_j \rangle = -\varepsilon_{k_j} \tag{51}$$

*and*

$$\|(p_j, q_j)\| \leq \frac{d_0^2}{i\tau(1-\sigma)\eta}, \qquad \varepsilon_{k_j} \leq \frac{\sigma^2 d_0^3}{(i\tau)^{3/2}(1-\sigma^2)^{3/2} 2\eta}; \tag{52}$$

(c) **[ergodic]** *we have*

$$v_i^a \in \partial_{\varepsilon_i^a} \left( \overline{\mathscr{L}}(\cdot, \tilde{y}_i^a) - \overline{\mathscr{L}}(\tilde{x}_i^a, \cdot) \right) (\tilde{x}_i^a, \tilde{y}_i^a) \tag{53}$$

*and*

$$\|v_i^a\| \leq \frac{2d_0^2}{(i\tau)^{3/2}(\sqrt{1-\sigma^2})\eta}, \qquad \varepsilon_i^a \leq \frac{2d_0^3}{(i\tau)^{3/2}(1-\sigma^2)\eta}; \tag{54}$$

(d) **[ergodic]** *there exists* $(p_i^a, q_i^a) \in \mathbb{R}^n \times \mathbb{R}^m$ *such that*

$$p_i^a \in \partial_{x, \varepsilon_i'} \overline{\mathscr{L}}(\tilde{x}_i^a, \tilde{y}_i^a), \qquad g(\tilde{x}_i^a) + q_i^a \leq 0, \quad \tilde{y}_i^a \geq 0, \quad \langle \tilde{y}_i^a, g(\tilde{x}_i^a) + q_i^a \rangle \geq -\varepsilon_i^a \tag{55}$$

*and*

$$\|(p_i^a, q_i^a)\| \leq \frac{2d_0^2}{(i\tau)^{3/2}(\sqrt{1-\sigma^2})\eta}, \qquad \varepsilon_i^a \leq \frac{2d_0^3}{(i\tau)^{3/2}(1-\sigma^2)\eta}, \tag{56}$$

*where* $\varepsilon_i' := \varepsilon_i^a + \langle \tilde{y}_i^a, g(\tilde{x}_i^a) + q_i^a \rangle$.

*Proof.* We first prove Items (a) and (c). Using Lemma 5.3, the last statement in Proposition 5.4 and (30) we have that Items (a) and (b) of Theorem 1.2 hold for the sequences $(\tilde{z}_{k_i})_{i \in I}$, $(v_{k_i})_{i \in I}$ and $(\varepsilon_{k_i})_{i \in I}$. As a consequence, to finish the proof of Items (a) and (c) of the theorem, it remains to prove the inclusions (49) and (53). To this end, note first that from the equivalence between Items (a) and (c) of Proposition A.1 (with $\varepsilon' = 0$) we have the equivalence

$$v_{k_i} \in (\mathbf{S} + \mathbf{N}^{[\varepsilon_{k_i}]}_{\mathbb{R}^n \times \mathbb{R}^m_+})(\tilde{x}_{k_i}, \tilde{y}_{k_i}) \iff v_{k_i} \in \partial_{\varepsilon_{k_i}} \left( \overline{\mathscr{L}}(\cdot, \tilde{y}_{k_i}) - \overline{\mathscr{L}}(\tilde{x}_{k_i}, \cdot) \right) (\tilde{x}_{k_i}, \tilde{y}_{k_i}) \qquad \forall i \in I.$$

Hence, using the latter equivalence, the first inclusion in (the first line) of (36), the inclusion in Theorem 1.2(a), the remark after the latter theorem, and (30) we obtain (49). Likewise, using an analogous reasoning and Proposition A.2 we also obtain (53), which finishes the proof of Items (a) and (c).

We claim that Item (b) follows from Item (a). Indeed, letting $(p_i, q_i) := v_{k_i}$ (for all $i \in I$), using the definition of $v_{k_j}$ and $\varepsilon_{k_j}$ in (31), the definition of $\mathbf{S}$ in (13) and the equivalence between Items (a) and (b) of Proposition A.1 (with $\varepsilon' = 0$) we obtain that $(p_j, q_j) := v_{k_j}$ satisfies (51) and (52). Using an analogous reasoning we obtain that Item (d) follows from Item (c). □

Next we analyze the sequence generated by Algorithm 1 for the set of indexes $k \in B$. Direct use of Algorithm 1's definition shows that

$$\lambda_{k+1} = \left(\frac{1}{1-\tau}\right)^{\#B_k - \#A_k} \lambda_1 \qquad \forall k \geq 1. \tag{57}$$

Define

$$\overline{\rho} = \frac{2\theta^2}{\lambda_1 \left(\dfrac{L_0}{2} + \sqrt{\left(\dfrac{L_0}{2}\right)^2 + \dfrac{8\|L_g\|\theta^2}{3\lambda_1}}\right)}. \tag{58}$$

In the next proposition we obtain a rate of convergence result for the sequence generated by Algorithm 1 with $k \in B$.

**Proposition 6.2.** *Let $\rho_k$ for all $k \geq 1$ be as in* (28) *and let also $\bar{\rho} > 0$ be as in* (58)*. Then, for all $k \in B$,*

$$v_k \in \partial_{\varepsilon_k} \left( \overline{\mathscr{L}}(\cdot, \tilde{y}_k) - \overline{\mathscr{L}}(\tilde{x}_k, \cdot)\right) (\tilde{x}_k, \tilde{y}_k)$$

*and*

$$\|v_k\| \leq \frac{(1+1/\sigma)\rho_k}{\lambda_k}, \qquad \varepsilon_k \leq \frac{\rho_k^2}{2\lambda_k}.$$

*Moreover, if $\lambda_k \geq \lambda_1$ then $\rho_k \leq \overline{\rho}$.*

*Proof.* First note that the desired inclusion follows from Proposition 5.2(b) and the equivalence between items (a) and (c) of Proposition A.1. Moreover, by Proposition 5.2 (c) we have

$$\|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \rho_k^2 \qquad \forall k \geq 1.$$

Note that the desired bound on $\varepsilon_k$ is a direct consequence of the latter inequality. Moreover, this inequality combined with the definition of $B$ (see (29)) gives $\|\lambda_k v_k\| \leq \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\| + \|\tilde{z}_k - z_{k-1}\| \leq (1+1/\sigma)\rho_k$ for all $k \in B$, which proves the desired bound on $\|v_k\|$.

Assume now that $\lambda_k \geq \lambda_1$. Using Definition 4.4 and (28) we obtain

$$\rho_k = \rho(\tilde{y}_k, \theta^2/\lambda_k) = \frac{2\theta^2}{\lambda_k \left(\dfrac{L_0 + \langle L_g, |\tilde{y}_k|\rangle}{2} + \sqrt{\left(\dfrac{L_0 + \langle L_g, |\tilde{y}_k|\rangle}{2}\right)^2 + \dfrac{8\|L_g\|\theta^2}{3\lambda_k}}\right)} \qquad \forall k \geq 1,$$

which, in turn, combined with (58), the assumption that $\lambda_k \geq \lambda_1$ and the fact that $\langle L_g, |\tilde{y}_k|\rangle \geq 0$ gives $\rho_k \leq \bar{\rho}$. □

Next we present the two main results of this paper, namely, the pointwise and ergodic iteration-complexities of Algorithm 1.

**Theorem 6.3** (pointwise iteration-complexity)**.** *For given tolerances $\overline{\delta} > 0$ and $\overline{\varepsilon} > 0$, after at most*

$$M := 2\left\lceil \max\left\{\frac{d_0^2}{\overline{\delta}\tau(1-\sigma)\eta},\ \frac{\sigma^{4/3}d_0^2}{\overline{\varepsilon}^{2/3}\tau(1-\sigma^2)(2\eta)^{2/3}}\right\}\right\rceil + \left\lceil \frac{\max\left\{\log^+\left((1+1/\sigma)\bar{\rho}/(\overline{\delta}\lambda_1)\right), \log^+\left(\bar{\rho}^2/(2\overline{\varepsilon}\lambda_1)\right)\right\}}{\log(1/(1-\tau))}\right\rceil \tag{59}$$

*iterations,* Algorithm 1 *finds $((x,y),v,\varepsilon)$ satisfying* (46) *with the property that $((x,y),(p,q),\varepsilon)$ where $(p,q) := v$ also satisfies*

$$\begin{aligned} &p = \nabla f(x) + \nabla g(x)y, \quad g(x)+q \le 0, \quad y \ge 0, \quad \langle y, g(x)+q\rangle = -\varepsilon, \\ &\|(p,q)\| \le \overline{\delta},\ \varepsilon \le \overline{\varepsilon}. \end{aligned} \tag{60}$$

*Proof.* First define

$$M_1 := \left\lceil \max\left\{\frac{d_0^2}{\overline{\delta}\tau(1-\sigma)\eta},\ \frac{\sigma^{4/3}d_0^2}{\overline{\varepsilon}^{2/3}\tau(1-\sigma^2)(2\eta)^{2/3}}\right\}\right\rceil \quad \text{and} \quad M_2 := M - 2M_1. \tag{61}$$

The proof is divided in two cases: (i) $\#A \ge M_1$ and (ii) $\#A < M_1$ . In the first case, the existence of $((x,y),v,\varepsilon)$ (resp. $((x,y),(p,q),\varepsilon)$) satisfying (46) (resp. (60)) in at most $M_1$ iterations follows from Theorem 6.1(a) (resp. Theorem 6.1(b)). Since $M = 2M_1 + M_2 \ge M_1$, it follows that, in this case, the number of iterations is not bigger than $M$.

Consider now the case (ii) and let $k^* \ge 1$ be such that $\#A = \#A_{k^*} = \#A_k$ for all $k \ge k^*$. As a consequence of the latter property and the fact that $\#A < M_1$ we conclude that if $\#B_k \ge M_1 + M_2$, for some $k \ge k^*$, then

$$\beta_k := \#B_k - \#A_k \ge \#B_k - M_1 \ge M_2. \tag{62}$$

Using the latter inequality, (59) and (61) we find

$$\beta_k \ge M_2 \ge \frac{\max\left\{\log^+\left((1+1/\sigma)\bar{\rho}/(\overline{\delta}\lambda_1)\right), \log^+\left(\bar{\rho}^2/(2\overline{\varepsilon}\lambda_1)\right)\right\}}{\log(1/(1-\tau))},$$

which is clearly equivalent to

$$\log\left(\left(\frac{1}{1-\tau}\right)^{\beta_k}\lambda_1\right) + \log\left(\frac{1}{(1+1/\sigma)\bar{\rho}}\right) \ge \log\left(\frac{1}{\bar{\overline{\delta}}}\right), \tag{63}$$

$$\log\left(\left(\frac{1}{1-\tau}\right)^{\beta_k}\lambda_1\right) + \log\left(\frac{2}{\bar{\rho}^2}\right) \ge \log\left(\frac{1}{\bar{\varepsilon}}\right). \tag{64}$$

Now using the definition in (62), (63) (resp. (64)) and (57) we obtain $\log(\lambda_k/[(1+1/\sigma)\bar{\rho}]) \ge \log(1/\bar{\delta})$ (resp. $\log(2\lambda_k/\bar{\rho}^2) \ge \log(1/\bar{\varepsilon})$) which yields

$$\frac{(1+1/\sigma)\bar{\rho}}{\lambda_k} \le \bar{\delta} \qquad \left(\text{resp. } \frac{\bar{\rho}^2}{2\lambda_k} \le \bar{\varepsilon}\right).$$

It follows from the latter inequality and Proposition 6.2 that $((x,y),v,\varepsilon) := ((\tilde{x}_k,\tilde{y}_k),v_k,\varepsilon_k)$ satisfies (46) and, due to Proposition A.1, that $((x,y),(p,q),\varepsilon) := ((\tilde{x}_k,\tilde{y}_k),v_k,\varepsilon_k)$ satisfies (60). Since the index $k$ has been chosen to satisfy $\#A_k < M_1$ and $\#B_k \ge M_1 + M_2$ we conclude that the total number of iterations is at most $M_1 + (M_1 + M_2) = M$. □

**Theorem 6.4** (ergodic iteration-complexity)**.** *For given tolerances $\bar{\delta} > 0$ and $\bar{\varepsilon} > 0$, after at most*

$$\widetilde{M} := 2\left\lceil \max\left\{ \frac{2^{2/3} d_0^{4/3}}{\bar{\delta}^{2/3}\tau\left(\eta\sqrt{1-\sigma}\right)^{2/3}},\ \frac{2^{2/3} d_0^{2}}{\bar{\varepsilon}^{2/3}\tau\left(\eta(1-\sigma^2)\right)^{2/3}} \right\}\right\rceil + \left\lceil \frac{\max\left\{\log^+\left((1+1/\sigma)\bar{\rho}/(\bar{\delta}\lambda_1)\right), \log^+\left(\bar{\rho}^2/(2\bar{\varepsilon}\lambda_1)\right)\right\}}{\log(1/(1-\tau))} \right\rceil \tag{65}$$

*iterations,* Algorithm 1 *finds* $((x,y),v,\varepsilon)$ *satisfying* (46) *with the property that* $((x,y),(p,q),\varepsilon)$ *where* $(p,q) := v$ *also satisfies*

$$\begin{aligned} &p \in \partial_{x,\varepsilon'}\overline{\mathscr{L}}(x,y), \quad g(x)+q \le 0, \quad y \ge 0, \quad \langle y, g(x)+q\rangle \ge -\varepsilon, \\ &\|(p,q)\| \le \bar{\delta},\ \varepsilon \le \bar{\varepsilon}, \end{aligned} \tag{66}$$

*where* $\varepsilon' := \varepsilon + \langle y, g(x)+q\rangle$.

*Proof.* The proof follows the same outline of Theorem 6.3's proof. □

# A Appendix

**Proposition A.1.** *Let* $(\tilde{x},\tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$, $v = (p,q) \in \mathbb{R}^n \times \mathbb{R}^m$ *and* $\varepsilon \ge 0$ *be given and define*

$$w := -(g(\tilde{x})+q), \qquad \varepsilon' := \varepsilon - \langle \tilde{y}, w\rangle. \tag{67}$$

*The following conditions are equivalent:*

(a) $v \in \partial_\varepsilon\left(\overline{\mathscr{L}}(\cdot,\tilde{y}) - \overline{\mathscr{L}}(\tilde{x},\cdot)\right)(\tilde{x},\tilde{y})$;

(b) $w \ge 0, \quad \langle \tilde{y}, w\rangle \le \varepsilon, \quad p \in \partial_{x,\varepsilon'}\overline{\mathscr{L}}(\tilde{x},\tilde{y})$;

(c) $0 \le \varepsilon' \le \varepsilon$ *and* $-w \in N^{[\varepsilon-\varepsilon']}_{\mathbb{R}^m_+}(\tilde{y}), \quad p \in \partial_{x,\varepsilon'}\overline{\mathscr{L}}(\tilde{x},\tilde{y})$.

*Proof.* $(a) \iff (b)$. Note that the inclusion in (a) is equivalent to

$$\overline{\mathscr{L}}(x,\tilde{y}) - \overline{\mathscr{L}}(\tilde{x},y) \ge \langle p, x-\tilde{x}\rangle + \langle q, y-\tilde{y}\rangle - \varepsilon \qquad \forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^m_+, \tag{68}$$

which, in view of (45) and (67), is equivalent to

$$\overline{\mathscr{L}}(x,\tilde{y}) - \overline{\mathscr{L}}(\tilde{x},\tilde{y}) + \inf_{y\ge 0}\langle w, y\rangle \ge \langle p, x-\tilde{x}\rangle - \varepsilon' \qquad \forall x \in \mathbb{R}^n.$$

The latter inequality is clearly equivalent to (b).

$(b) \iff (c)$. Using (17), the fact that $\tilde{y} \ge 0$ and the definition of $\varepsilon'$ in (67) we obtain that the first inequality in (b) is equivalent to $\varepsilon' \le \varepsilon$ and the second inequality in (c). To finish the proof note that the second inequality in (b) is equivalent to $\varepsilon' \ge 0$. □

**Proposition A.2.** *Let* $X \subset \mathbb{R}^n$ *and* $Y \subset \mathbb{R}^m$ *be given convex sets and* $\Gamma : X \times Y \to \mathbb{R}$ *be a function such that, for each* $(x,y) \in X \times Y$, *the function* $\Gamma(\cdot,y) - \Gamma(x,\cdot) : X \times Y \to \mathbb{R}$ *is convex. Suppose that, for* $j = 1,\ldots,i$, $(\tilde{x}_j,\tilde{y}_j) \in X \times Y$ *and* $(p_j,q_j) \in \mathbb{R}^n \times \mathbb{R}^m$ *satisfy*

$$(p_j,q_j) \in \partial_{\varepsilon_j}\left(\Gamma(\cdot,\tilde{y}_j) - \Gamma(\tilde{x}_j,\cdot)\right)(\tilde{x}_j,\tilde{y}_j).$$

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

$$(\tilde{x}_i^a, \tilde{y}_i^a) = \sum_{j=1}^{i} \alpha_j (\tilde{x}_j, \tilde{y}_j), \quad (p_i^a, q_i^a) = \sum_{j=1}^{i} \alpha_j (p_j, q_j),$$

$$\varepsilon_i^a = \sum_{j=1}^{i} \alpha_j \left[ \varepsilon_j + \langle \tilde{x}_j - \tilde{x}_i^a, p_j \rangle + \langle \tilde{y}_j - \tilde{y}_i^a, q_j \rangle \right] .$$

*Then,* $\varepsilon_i^a \geq 0$ *and*

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

*Proof.* See [6, Proposition 5.1] . □

# References


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

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

[3] M.R. Hestenes. Multiplier and Gradient Methods. In *Computing Methods in Optimization Problems-2, Edited by L.A.Zadeh, L.W. Neustadt, and A.V. Balakrishnan.* Academic Press, New York, 1969.

[4] M.R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

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

[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 Journal on Optimization*, 21:1688–1720, 2010.

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

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

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

[10] Yu. Nesterov and B. T. Polyak. Cubic regularization of Newton method and its global performance. *Math. Program.*, 108(1, Ser. A):177–205, 2006.

[11] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In *Optimization, Edited by Fletcher*. Academic Press, New York, 1972.

[12] R. T. Rockafellar. The multiplier method of Hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications*, 12:555–562, 1973.

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

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

[15] 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 mixed complementarity problems. *Computational and Applied Mathematics*, 2017.

[16] 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.

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

[18] 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).

[19] 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 (electronic), 2000.

[20] 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.

[21] M. V. Solodov and B. F. Svaiter. A new proximal-based globalization strategy for the Josephy-Newton method for variational inequalities. *Optim. Methods Softw.*, 17(5):965–983, 2002.

[22] B. F. Svaiter. Complexity of the relaxed hybrid proximal-extragradient method under the large-step condition. Preprint A766/2015, IMPA - Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ Brasil 22460-320, 2015.