AN INEXACT MODIFIED SUBGRADIENT ALGORITHM FOR PRIMAL-DUAL PROBLEMS VIA AUGMENTED LAGRANGIANS*

REGINA S. BURACHIK†, ALFREDO N. IUSEM ‡, AND JEFFERSON G. MELO§

Abstract. We consider a primal optimization problem in a reflexive Banach space and a duality scheme via generalized augmented Lagrangians. For solving the dual problem, we introduce and analyze a new *parameterized* Inexact Modified Subgradient (IMSg) algorithm. The IMSg generates a primal-dual sequence, and we focus on two simple new choices of the stepsize. We prove that every weak accumulation point of the primal sequence is a primal solution and the dual sequence converges weakly to a dual solution, as long as the dual optimal set is nonempty. Moreover, we establish primal convergence even when the dual optimal set is empty. Our second choice of the stepsize gives rise to a variant of IMSg which has finite termination.

Key words. Banach spaces, nonsmooth optimization, nonconvex optimization, duality scheme, augmented Lagrangian, inexact modified subgradient algorithm.

AMS subject classifications. 90C26;49M29;65K10

1. Introduction. The classical Lagrange function is a useful tool for dealing with constrained optimization problems, specially when the problem is convex. When the *primal* problem is nonconvex, and the dual problem is generated by the classical Lagrange function, a nonzero duality gap between the primal and dual problem may exist, compromising the applicability of many methods. In order to avoid this positive gap, an augmented Lagrangian function can be used instead of the classical linear Lagrangian. A primal problem of minimizing a nonsmooth and nonconvex extended real-valued function is considered in [18, Chapter 11], where a dual problem constructed via augmented Lagrangians with convex augmenting functions is proposed and analyzed. Since then, augmented Lagrangians with nonconvex augmenting functions have been intensively studied, see for instance [7, 8, 11, 12, 17, 19, 20, 22, 23]. The dual problem generated by these augmented Lagrangian functions is convex (i.e, the dual problem is the maximization of a concave function over a convex set), and therefore subgradient methods and its variants are suitable methods for solving it.

Subgradient methods have been extensively studied in the context of classical Lagrangian duality, see for instance [15, 16] and references therein. Modified subgradient algorithms (MSG) have been considered for the dual problem constructed via sharp Lagrangian in finite dimensional spaces, see [2, 4, 5, 6, 9]. A property of these algorithms is that the dual values monotonically increase and converge to the dual optimal value. In [6] the authors introduced and analyzed an inexact version of the algorithm proposed in [2]. The stepsize rule for the algorithms considered in [2, 5, 6, 9] depends strongly on the "a priori" knowledge of the primal optimal value. Moreover, no primal convergence is achieved in these papers. These drawbacks were addressed in [4], where the authors introduced a stepsize selection rule with primal convergence which does not need the "a priori" knowledge of the primal optimal value. The method devised in [4] considers the dual problem induced by the sharp Lagrangian in finite dimensional spaces. However, this method assumes

*Regina S. Burachik acknowledges support by the Australian Research Council Discovery Project Grant DP0556685 for this study; Jefferson G. Melo acknowledges support from CNPq, through a scholarship for a one-year long visit to University of South Australia

†School of Mathematics and Statistics, University of South Australia, Mawson Lakes, SA 5095, Australia regina.burachik@unisa.edu.au.

‡IMPA, Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, CEP 22460-320, Brazil iusp@impa.br.

§IMPA, Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, CEP 22460-320, Brazil jeferson@impa.br.

exact solution of the subproblems, which is too strong a requirement in the context of nonsmooth and nonconvex optimization. The present paper addresses this issue, by developing an infinite dimensional modified subgradient method which accepts an inexact solution of the subproblems, and can be applied to duality schemes induced by a wide family of augmented Lagrangians in Banach spaces.

We consider a primal problem of minimizing an extended real-valued function (possibly nonconvex and nondifferentiable) in a reflexive Banach space. A duality scheme is considered via augmented Lagrangian functions which include the sharp Lagrangian as a particular case (see Example 2.1). Our dual variables belong to a Hilbert space. Such duality schemes are suitable for solving constrained optimization problems in which the image space of the constraint function is a Hilbert space, see e.g. [13, 14, 21, 23] and references therein. Moreover, development and analysis of a given algorithm in infinite dimensional spaces gives a deeper insight into the properties of that algorithm. This issue has also practical interest since usually the performance of numerical algorithms in finite dimensions are closely related to the infinite dimensional performance, see for example [21] and references therein.

We propose a *parameterized* inexact modified subgradient algorithm for solving the dual problem. For this purpose we use a dualizing parameterization function (a function $f(\cdot,\cdot)$ such that $f(\cdot,0) = \varphi(\cdot)$, where $\varphi(\cdot)$ is the primal function). To ensure a monotone improvement of the dual values, we consider an augmenting function (not necessarily convex) similar to the one used in [10] (see assumption (A_0) below). We prove that validity of (A_0) is necessary for having a monotone increase of the dual values, see Proposition 3.15. Our method extends in many ways the one proposed in [4]. First, we extend to a reflexive Banach space the (finite dimensional) MSG proposed in [4]. Also our method admits the dual variables to be in a finite or infinite dimensional Hilbert space; as commented above this can have some advantages. Second, the convergence analysis in [4] assumed exact solution of the subproblems, while here we establish convergence accepting inexact iterates, which is in fact the actual situation in computational implementations. Third, the sharp Lagrangian considered in [4] is just a simple particular case of our augmented Lagrangians (see Example 2.1 and assumption (A_0)). Moreover, we consider in our analysis a level-boundedness assumption on the dualizing parameterization function (Definition 2.2) which is weaker than the compactness assumption used in [2, 4, 6, 9, 10].

We show that, in our more general setting, our algorithm generates primal and dual sequences which are weakly convergent to primal and dual solutions. The primal sequence converges in the sense that all its weak accumulation points are primal solutions, even when the dual solution set is empty. We also analyze a stepsize selection rule which ensures that when the dual solution set is nonempty, approximate primal and dual solutions are obtained after a finite number of iterations of the algorithm (see Section 3.2). It is well known that subgradient methods with classical Lagrangian do not always obtain a primal solution, unless ergodic averages or more sophisticated techniques are considered, see for example [15] and references therein.

The paper is organized as follows. In Section 2 we describe the setting of our primal and dual problems, and give some basic definitions, assumptions and examples. We also recall in this section some useful facts. In Section 3 we consider the inexact modified subgradient algorithm (IMSg) and establish its convergence properties which do not depend on the choice of the stepsize. In Section 3.1 we propose a stepsize selection for IMSg and state and prove our main results. In Section 3.2 another stepsize rule for IMSg is proposed and we show that, under this stepsize rule, IMSg converges in a finite number of steps. In the last section we compare our algorithm with the algorithms with sharp Lagrangian considered in [2, 4, 6, 9].

2. Preliminaries. Let X be a reflexive Banach space and H a Hilbert space. We denote by $\langle\cdot,\cdot\rangle$ the scalar product in H, and by $\|\cdot\|$ the norm, where the same notation will be used for the norm both in X and H. We consider the optimization problem

$$\min\ \varphi(x)\ \text{ s.t. } x \text{ in } X, \tag{2.1}$$

where the function $\varphi : X \to \mathbb{R}_{+\infty} := \mathbb{R} \cup \{+\infty\}$ is a proper (i.e., $\operatorname{dom}\varphi \neq \emptyset$ and $\varphi > -\infty$) weakly lower semicontinuous (w-lsc) function. We also assume that φ has weakly compact sublevel sets. In order to introduce our duality scheme, we consider a *dualizing parameterization* for (2.1), which is a function $f : X \times H \to \bar{R} := \mathbb{R}_{+\infty} \cup \{-\infty\}$ that verifies $f(x, 0) = \varphi(x)$ for all $x \in X$. The perturbation function induced by this dualizing parameterization is the function $\beta : H \to \bar{\mathbb{R}}$ defined by

$$\beta(z) := \inf_{x \in X} f(x, z).$$

Because φ is proper, we have $\beta(0) < +\infty$. Next we define a level-bounded augmenting function. Augmenting functions were introduced in [18, Definition 11.55]. See [11, 7] and references therein for other generalizations of augmenting functions.

DEFINITION 2.1. *A function $\sigma : H \to \mathbb{R}_{+\infty}$ is said to be a level-bounded augmenting function if it is proper, w-lsc, level-bounded on H, $\sigma(0) = 0$ and* $\operatorname{argmin}_y \sigma(y) = \{0\}$.

The *augmented Lagrangian function* $\ell : X \times H \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ is defined as

$$\ell(x, y, r) := \inf_{z \in H} \{f(x, z) - \langle z, y\rangle + r\sigma(z)\}. \tag{2.2}$$

The dual function $q : H \times \mathbb{R}_+ \to \mathbb{R} \cup \{-\infty\}$ is defined as $q(y, r) := \inf_{x \in X} \ell(x, y, r)$ and therefore the dual problem is stated as

$$\max\ q(y, r)\ \text{ s.t. } (y, r) \in H \times \mathbb{R}_+. \tag{2.3}$$

Denote by $M_P := \inf_{x \in X} \varphi(x)$ and by $M_D := \sup_{(y,r) \in H \times \mathbb{R}_+} q(y, r)$ the optimal values of the primal and dual problem, respectively. The primal and dual solution sets are denoted by P_* and D_*, respectively.

EXAMPLE 2.1. Consider the following equality constrained problem

$$\min\ \psi(x)\ \text{ s.t. } x \in K, h(x) = 0, \tag{2.4}$$

where $h : X \to H$ has a weakly closed graph, i.e., $G(h) := \{(x, h(x)) : x \in K\}$ is weakly closed in $X \times H$, $\psi : X \to \mathbb{R}$ is w-lsc, and $K \subset X$ is weakly closed. We consider the following equivalent unconstrained problem:

$$\min\ \phi(x) := \psi(x) + \delta_V(x), \quad \text{s.t. } x \in X,$$

where $V := \{x \in K : h(x) = 0\}$ and $\delta_V(x) = 0$ if $x \in V$, $\delta_V(x) = \infty$ otherwise. Consider the augmenting function given by $\sigma(\cdot) = \|\cdot\|$, and the canonical dualizing parameterization function given by

$$f(x, z) = \begin{cases} \psi(x) & \text{if } x \in K \text{ and } h(x) = z, \\ \infty, & \text{otherwise.} \end{cases}$$

By definition, we have $\ell(x, y, r) = \inf_{z \in H} \{f(x, z) - \langle z, y\rangle + r\sigma(z)\}$, which in this case becomes the sharp Lagrangian proposed in [18, Example 11.58]:

$$\ell(x, y, r) = \begin{cases} \psi(x) - \langle y, h(x)\rangle + r\|h(x)\| & \text{if } x \in K, \\ \infty & \text{otherwise.} \end{cases}$$

The dual function induced by this Lagrangian is

$$\hat{q}(y,r) := \inf_{x\in K}\{\psi(x) - \langle y, h(x)\rangle + r\|h(x)\|\},$$

and the dual problem is

$$\max\ \hat{q}(y,r)\ \text{ s.t. }\ (y,r)\in H\times\mathbb{R}_+.$$

A modified subgradient algorithm has been considered in [9, 2, 6, 4] for the primal-dual scheme described in Example 2.1, under the assumptions that $X=\mathbb{R}^n$, $H=\mathbb{R}^m$ and K is a compact set.

As in [10, Eq. (8.1)], we make the following assumption on the augmenting function.

$(\mathbf{A_0})$: $\sigma(z)\geq\|z\|$ for all $z\in H$.

Next we list some examples of augmenting functions that satisfy (A_0).

i) Let $\sigma_{p,q}: H\to\mathbb{R}$ be defined as

$$\sigma_{p,q}(z) := \begin{cases} \|z\|^p & \text{if } \|z\|\leq 1,\\ \|z\|^q & \text{otherwise,}\end{cases}$$

with $0<p\leq 1\leq q$.

ii) Let $H=\mathbb{R}^n$, and $\sigma_k:\mathbb{R}^n\to\mathbb{R}$ be defined as $\sigma_k(z) := (\sum_{i=1}^n |z_i|^{\frac{1}{k}})^k$, with $k\in\mathbb{N}$.

The next definition has been considered in [7, Section 5] and it is a natural generalization of [18, Definition 1.16].

DEFINITION 2.2. *A function $f: X\times H\to\bar{\mathbb{R}}$ is said to be* weakly level-compact *if for every $\bar{z}\in H$ and $\alpha\in\mathbb{R}$ there exist a weak neighborhood $V\subset H$ of $\bar{z}$, and a weak compact set $B\subset X$, such that*

$$L_{V,f}(\alpha) := \{x\in X : f(x,z)\leq\alpha\}\subset B \ \textit{ for all } z\in V.$$

REMARK 2.1. If f verifies Definition 2.2 then every sequence in $L_{V,f}(\alpha)$ has a weakly convergent subsequence. It is not difficult to see that the canonical dualizing parameterization function considered in Example 2.1 is weakly level-compact if K is a weakly compact set. We will only consider dualizing parameterization functions which are proper, weakly level-compact and w-lsc. Next we consider some basic properties of the dual function.

PROPOSITION 2.3. *i) The dual function q is concave and weakly upper-semicontinuous* (w-usc) *function.*

ii) If $r\geq c$ then $q(y,r)\geq q(y,c)$ for all $y\in H$. In particular, if (y,c) is a dual solution, then also (y,r) is a dual solution for all $r\geq c$.

Proof. Item (i) follows from the fact that q is the infimum of affine functions. Item (ii) follows from the fact that the penalty function σ is nonnegative. □

Augmented Lagrangians are a special case of abstract Lagrangians (see for instance, [7] and [19, Section 5.2]). In [7, Proposition 4.1] the authors obtained strong duality for abstract Lagrangians in the framework of abstract convexity. The next theorem is a consequence of [7, Proposition 4.1]. It ensures that there is no duality gap between the primal problem (2.1) and its dual problem (2.3).

THEOREM 2.4. *Consider the primal problem* (2.1) *and its dual problem* (2.3). *Assume that the dualizing parameterization function $f: X\times H\to\bar{\mathbb{R}}$ for the primal function φ is proper, w-lsc and* *weakly level-compact. Suppose that there exists some* $(y,r) \in H \times \mathbb{R}_+$ *such that* $q(y,r) > -\infty$. *Then zero duality gap holds, i.e.* $M_P = M_D$.

Proof. The result is an immediate consequence of [7, Proposition 4.1]. Our duality scheme is a special case of the one considered in [7]; see [7, Remark 3.5], and observe also that level bounded augmenting functions are a special case of the augmenting functions considered in [7]. It follows from [7, Proposition 5.1] that if the dualizing parameterization function f is weakly level-compact then the perturbation function $\beta(y) = \inf_x f(x,y)$ is w-lsc. Thus we just need to ensure that the support of the perturbation function, in the sense of [7, Remark 3.3], is nonempty, but it is elementary to show that this is equivalent to the existence of $(y,c) \in H \times \mathbb{R}_+$ such that $q(y,c) > -\infty$, which is part of our hypotheses. □

¿From now on we assume that the hypotheses of Theorem 2.4 are verified. We give next some definitions.

DEFINITION 2.5. *Let* H *be a Hilbert space and* $q : H \to \mathbb{R}_{-\infty}$ *a concave function. Take* $r \geq 0$. *The* r*-superdifferential of* q *at* $y_0 \in \mathrm{dom}(q) := \{y \in H \,:\, q(y) > -\infty\}$ *is the set* $\partial_r q(y_0)$ *defined by*

$$\partial_r q(y_0) := \{v \in H : q(y) \leq q(y_0) + \langle v, y - y_0\rangle + r \ \ \forall y \in H\}.$$

DEFINITION 2.6. *We say that* $x_* \in X$ *is an* ϵ_**-optimal primal solution if* $\varphi(x_*) \leq M_P + \epsilon_*$; *we say that* $(y_*, c_*) \in H \times \mathbb{R}_+$ *is an* ϵ_**-optimal dual solution if* $q(y_*, c_*) \geq M_D - \epsilon_*$.

For $r \geq 0$ consider the following set

$$A_r(y,c) = \{(x,z) \in X \times H : f(x,z) - \langle z, y\rangle + c\sigma(z) \leq q(y,c) + r\}. \tag{2.5}$$

By definition of q and (2.2), we see that $A_r(y,c)$ is nonempty for all $r > 0$ and all (y,c) such that $q(y,c) > -\infty$. Fix $(y,c) \in H \times \mathbb{R}_+$ and define $\Phi_{(y,c)} : X \times H \to \bar{\mathbb{R}}$ as

$$\Phi_{(y,c)}(x,z) = f(x,z) - \langle z, y\rangle + c\sigma(z). \tag{2.6}$$

Observe that computation of an element in $A_r(y,c)$ is tantamount to an approximate unconstrained minimization of $\Phi_{(y,c)}(\cdot,\cdot)$, with tolerance r.

3. Inexact Modified Subgradient Algorithm (IMSg).

We state next the Inexact Modified Subgradient Algorithm (IMSg).

Step 0. Choose $(y_0, c_0) \in H \times \mathbb{R}_+$ such that $q(y_0, c_0) > -\infty$, and exogenous parameters $\epsilon_* > 0$ (a prescribed tolerance), $\delta < 1$, $\{\alpha_k\} \subset (0, \alpha)$ for some $\alpha > 0$, and $\{r_k\} \subset \mathbb{R}_+$ such that $r_k \to 0$. Let $k := 0$.

Step 1. (Subproblem and Stopping Criterion)

a) Find $(x_k, z_k) \in A_{r_k}(y_k, c_k)$,

b) if $z_k = 0$ and $r_k \leq \epsilon_*$ stop,

c) if $z_k = 0$ and $r_k > \epsilon_*$, then $r_k := \delta r_k$ and go to (*a*),

d) if $z_k \neq 0$ go to Step 2.

Step 2. (Selection of the stepsize and Updating the Variables)

Consider s_k a stepsize and define

$y_{k+1} := y_k - s_k z_k$,

$c_{k+1} := c_k + (\alpha_k + 1) s_k \sigma(z_k)$,

$k := k + 1$, go to Step 1.

Note that IMSg has the general form of standard augmented Lagrangian methods: in Step 1 the primal variables are updated through the approximate solution of an unconstrained minimization problem (in this case producing the pair (x, z)), and then in Step 2 the dual variables (y_k, c_k) is updated through an explicit formula, in this case moving along a direction of dual ascent. Observe also that here the penalty parameters $\{c_k\}$ are considered as variables. The parameters $\{\alpha_k\}$ ensure monotonic increase of the dual values; see Theorem 3.3.

First, we present some results which do not depend on the selection of the stepsize. The next proposition establishes the relation between the approximate minimization implicit in $A_r(y, c)$ and the approximate superdifferential $\partial_r q(y, c)$.

PROPOSITION 3.1. *The following facts hold for* IMSg.

i) If $(\hat{x}, \hat{z}) \in A_r(\hat{y}, \hat{c})$, *then* $(-\hat{z}, \sigma(\hat{z})) \in \partial_r q(\hat{y}, \hat{c})$, *for all* $r \geq 0$.

ii) If (A_0) *holds then* IMSg *generates a dual bounded sequence* $\{(y_k, c_k)\}$ *if and only if* $\sum_k s_k \sigma(z_k) < +\infty$.

iii) If IMSg *stops at iteration* k, *then* x_k *is an* ϵ_**-optimal primal solution, and* (y_k, c_k) *is an* ϵ_**-optimal dual solution.*

Proof. i) For all $(y, c) \in H \times \mathbb{R}_+$ we have:

$$\begin{aligned} q(y,c) &= \inf_{(x,z)}\{f(x,z) - \langle z, y\rangle + c\sigma(z)\} \\ &\leq f(\hat{x},\hat{z}) - \langle \hat{z}, y\rangle + c\sigma(\hat{z}) \\ &= f(\hat{x},\hat{z}) - \langle \hat{z}, \hat{y}\rangle + \hat{c}\sigma(\hat{z}) + \langle -\hat{z}, y - \hat{y}\rangle + (c - \hat{c})\sigma(\hat{z}). \end{aligned} \tag{3.1}$$

Using that $(\hat{x}, \hat{z}) \in A_r(\hat{y}, \hat{c})$ in (3.1), we obtain

$$\begin{aligned} q(y,c) &\leq q(\hat{y},\hat{c}) + r + \langle -\hat{z}, y - \hat{y}\rangle + (c - \hat{c})\sigma(\hat{z}) \\ &= q(\hat{y},\hat{c}) + \langle(-\hat{z}, \sigma(\hat{z})), (y,c) - (\hat{y},\hat{c})\rangle + r. \end{aligned}$$

Therefore, $(-\hat{z}, \sigma(\hat{z})) \in \partial_r q(\hat{y}, \hat{c})$.

ii)Using (A_0) and simple manipulations in the definition of $\{y_k\}$, we obtain

$$\|y_{k+1} - y_0\| \leq \sum_{j=0}^{k} \|y_{j+1} - y_j\| = \sum_{j=0}^{k} s_j \|z_j\| \leq \sum_{j=0}^{k} s_j \sigma(z_j). \tag{3.2}$$

On the other hand,

$$c_{k+1} - c_0 = \sum_{j=0}^{k} c_{j+1} - c_j = \sum_{j=0}^{k} (\alpha_j + 1) s_j \sigma(z_j). \tag{3.3}$$

Since $\{\alpha_k\}$ is bounded, (ii) follows from (3.2) and (3.3).

For proving (iii), observe that if IMSg stops at iteration k, then $z_k = 0$ and $r_k \leq \epsilon_*$. Therefore we have (see Theorem 2.4):

$$\begin{aligned} M_D &= M_P \leq \varphi(x_k) = f(x_k, 0) - \langle y_k, 0\rangle + c_k\sigma(0) \leq q(y_k, c_k) + r_k \\ &\leq q(y_k, c_k) + \epsilon_* \leq M_P + \epsilon_*, \end{aligned}$$

which implies that $M_D \leq q(y_k, c_k) + \epsilon_*$, and $\varphi(x_k) \leq M_P + \epsilon_*$. That is to say, x_k is an ϵ_*-optimal primal solution, and (y_k, c_k) is an ϵ_*-optimal dual solution.

□

Next we establish boundedness properties of the sub-level sets of $\Phi_{(y,c)}(\cdot,\cdot)$.

LEMMA 3.2. *Let $(y, c_y) \in H \times \mathbb{R}_+$ be such that $q(y, c_y) > -\infty$. The set*

$$L_r(y,c) = \{(x,z) : \Phi_{(y,c)}(x,z) := f(x,z) - \langle z, y\rangle + c\sigma(z) \leq M_P + r\}$$

is nonempty and weakly-compact for each $r \geq 0$ and $c > c_y$. In particular, for each $r \geq 0$ and $c > c_y$ there exists some $(\tilde{x}, \tilde{z})$ such that

$$q(y,c) = f(\tilde{x},\tilde{z}) - \langle \tilde{z}, y\rangle + c\sigma(\tilde{z}).$$

Proof. Since the function $\varphi(\cdot) = f(\cdot, 0)$ is w-lsc and weakly-level compact, there exists a global minimizer x^* of φ, so that $(x^*, 0) \in L_r(y,c)$ for all y, c and $r \geq 0$, ensuring that $L_r(y,c)$ is nonempty. For proving that $L_r(y,c)$ is bounded, suppose by contradiction that $L_r(y,c)$ is unbounded for some y, c, r with $c > c_y$ and $r \geq 0$, so that there exists some unbounded sequence $\{(x_k, z_k)\} \subset L_r(y,c)$. Therefore

$$\begin{aligned} M_P + r &\geq f(x_k,z_k) - \langle z_k, y\rangle + c\sigma(z_k) = f(x_k,z_k) - \langle z_k, y\rangle + c_y\sigma(z_k) + (c-c_y)\sigma(z_k) \\ &\geq q(y,c_y) + (c-c_y)\sigma(z_k) \geq q(y,c_y) + (c-c_y)\|z_k\|, \end{aligned}$$

implying that

$$\|z_k\| \leq \frac{M_P + r - q(y,c_y)}{c - c_y},$$

and hence $\{z_k\}$ is bounded. Without loss of generality we can assume that the whole sequence $\{z_k\}$ converges weakly to some $\bar{z}$. Since $\{z_k\}$ is bounded and $\sigma(z) \geq 0$ for all z, we have

$$f(x_k, z_k) \leq M_P + r + \|y\|\|z_k\| \leq \tilde{\alpha} \tag{3.4}$$

for some $\tilde{\alpha} \in \mathbb{R}$. Take a weak compact set $B \subset X$ and a weak neighborhood V of $\bar{z}$ given by the level compactness property of f related to $\bar{z}$ and $\tilde{\alpha}$ (see Definition 2.2). We know that there exists k_0 such that $z_k \in V$ for all $k > k_0$. Thus $\{x_k\}_{k>k_0} \subset L_{V,f}(\tilde{\alpha}) \subset B$, by (3.4). Therefore $\{x_k\}$ is bounded, and hence $\{(x_k, z_k)\}$ is bounded, which is a contradiction, establishing boundedness of $L_r(y,c)$. Since the function $\Phi_{(y,c)}(\cdot,\cdot)$ given by (2.6) is w-lsc, $L_r(y,c)$ is also weakly-closed, and so $L_r(y,c)$ is weakly-compact, by Banach-Alaoglu theorem. The last assertion of the Lemma is equivalent to

$$(\tilde{x}, \tilde{z}) \in \operatorname{argmin}_{(x,z)\in X\times H} \Phi_{(y,c)}(x,z).$$

Indeed, $(\tilde{x}, \tilde{z})$ verifies the inclusion above if and only if

$$q(y,c) = \inf_{(x,z)\in X\times H} \Phi_{(y,c)}(x,z) = \Phi_{(y,c)}(\tilde{x},\tilde{z}) = f(\tilde{x},\tilde{z}) - \langle \tilde{z}, y\rangle + c\sigma(\tilde{z}), \tag{3.5}$$

where we use the definitions of $\Phi_{(y,c)}(\cdot,\cdot)$ and q. Note also that $X \times H$ is reflexive and $L_r(y,c)$ is a nonempty weakly-compact sub-level set of $\Phi_{(y,c)}(\cdot,\cdot)$, as we have already established. Since $\Phi_{(y,c)}(\cdot,\cdot)$ is w-lsc, we know (see for example [3, Proposition 3.1.15]) that $\Phi_{(y,c)}(\cdot,\cdot)$ attains its minimum $(\tilde{x}, \tilde{z})$ on $L_r(y,c)$, which must coincide with the unconstrained minimum. In view of (3.5), we conclude that $(\tilde{x}, \tilde{z})$ verifies

$$q(y,c) = f(\tilde{x},\tilde{z}) - \langle \tilde{z}, y\rangle + c\sigma(\tilde{z}).$$

□

The following theorem guarantees a special property of IMSg, which is not verified by the classical subgradient algorithm. It states that IMSg guarantees a monotonic increase of the dual function.

THEOREM 3.3. *Let $\{z_k\}$ be the sequence generated by* IMSg. *If $z_k \neq 0$ and (y_k, c_k) is not a dual solution, then $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$.*

Proof. For all k consider $\epsilon_k := \alpha_k s_k$. Using the update rule to the dual variables we have

$$\begin{aligned} q(y_{k+1}, c_{k+1}) &= \inf_{(x,z)} \{f(x,z) - \langle z, y_{k+1}\rangle + c_{k+1}\sigma(z)\} \\ &= \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + s_k\langle z, z_k\rangle + [c_k + (\epsilon_k + s_k)\sigma(z_k)]\sigma(z)\} \\ &= \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + (c_k + \epsilon_k\sigma(z_k))\sigma(z) + (\sigma(z_k)\sigma(z) + \langle z, z_k\rangle)s_k\} \\ &\geq \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + (c_k + \epsilon_k\|z_k\|)\sigma(z) + (\|z_k\|\|z\| + \langle z, z_k\rangle)s_k\} \end{aligned}$$

where the inequality follows from (A_0). Now we obtain, using Cauchy-Schwarz inequality,

$$\begin{aligned} q(y_{k+1}, c_{k+1}) &\geq \inf_{(x,z)} \{f(x,z) - \langle z, y_k\rangle + (c_k + \epsilon_k\|z_k\|)\sigma(z)\} \\ &= q(y_k, c_k + \epsilon_k\|z_k\|) \geq q(y_k, c_k), \end{aligned}$$

where the second inequality follows from Proposition 2.3 (ii). Thus for all k we have

$$q(y_{k+1}, c_{k+1}) \geq q(y_k, c_k + \epsilon_k\|z_k\|) \geq q(y_k, c_k). \tag{3.6}$$

In particular we have $q(y_{k+1}, c_{k+1}) \geq q(y_k, c_k)$ for all k. Since $q_0 = q(y_0, c_0) > -\infty$, we conclude that $q(y_k, c_k) > -\infty$ for all k. Therefore we obtain from Lemma 3.2 that, fixing k such that $z_k \neq 0$, there exists $(\tilde{x}, \tilde{z})$ such that

$$q(y_k, c_k + \epsilon_k\|z_k\|) = f(\tilde{x}, \tilde{z}) - \langle \tilde{z}, y_k\rangle + (c_k + \epsilon_k\|z_k\|)\sigma(\tilde{z}). \tag{3.7}$$

If $\tilde{z} = 0$, then we get from (3.7) that $q(y_k, c_k + \epsilon_k\|z_k\|) = f(\tilde{x}, 0) = \varphi(\tilde{x}) \geq M_D > q(y_k, c_k)$, where the last strict inequality follows from the fact that (y_k, c_k) is not a dual solution. Therefore we conclude from (3.6) that $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$. If $\tilde{z} \neq 0$, then, since $z_k \neq 0$ and $\epsilon_k > 0$, we obtain from (3.6) and (3.7) that

$$q(y_{k+1}, c_{k+1}) \geq q(y_k, c_k) + \epsilon_k\|z_k\|\|\tilde{z}\| > q(y_k, c_k),$$

using (A_0) and definition of q. The proof is complete. □

¿From now on we assume that $z_k \neq 0$ for all k. In other words, we assume from now on that the method generates an infinite sequence.

LEMMA 3.4. *Let $\{(y_k, c_k)\}$ be the sequence generated by* IMSg *and consider a sequence $\{z_k\}$ such that $(x_k, z_k) \in A_{r_k}(y_k, c_k)$ for all k. Then, the sequence $\{\sigma(z_k)\}$ is bounded and in particular $\{z_k\}$ is bounded.*

Proof. ¿From (3.2) and (3.3) we obtain, for all $k \geq 1$,

$$c_k - c_0 \geq \|y_k - y_0\| + \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j) \geq \|y_k - y_0\| + a,$$

for some $a > 0$, (e.g., we may take $a = \alpha_0 s_0 \sigma(z_0)$). Hence we have

$$\|y_k - y_0\| + c_0 - c_k \leq -a. \tag{3.8}$$

On the other hand, by Proposition 3.1, $(-z_k, \sigma(z_k)) \in \partial q_{r_k}(y_k, c_k)$. Thus

$$\begin{aligned} q_0 = q(y_0, c_0) &\leq q(y_k, c_k) + \langle -z_k, y_0 - y_k \rangle + (c_0 - c_k)\sigma(z_k) + r_k \\ &\leq M_D + \|z_k\|\|y_k - y_0\| + (c_0 - c_k)\sigma(z_k) + r_k \\ &\leq M_D + \sigma(z_k)(\|y_k - y_0\| + c_0 - c_k) + r_k \\ &\leq M_D - a\sigma(z_k) + r_k, \end{aligned}$$

using Cauchy-Schwarz inequality in the second inequality, (A_0) in the third one, and (3.8) in the last one. It follows that

$$q_0 \leq M_D - a\sigma(z_k) + r_k. \tag{3.9}$$

Rewriting (3.9) we have

$$\sigma(z_k) \leq \frac{M_D - q_0 + r_k}{a} \leq \frac{M_D - q_0 + \tilde{r}}{a} := b,$$

where $\tilde{r} > 0$ is an upper bound for $\{r_k\}$. Since $\|z_k\| \leq \sigma(z_k) \leq b$ for all $k > k_0$, the proof is complete. □

¿From now on we use the notation $q_k := q(y_k, c_k)$ for all k, and $\bar{q} := M_D$.

LEMMA 3.5. *Consider the sequences $\{(x_k, z_k)\}$, $\{(y_k, c_k)\}$ generated by* IMSg *algorithm.*

a) The following estimates hold for all $k \geq 1$

$$f(x_k, z_k) - \langle z_k, y_0 \rangle \leq q_k + r_k, \ \text{and} \tag{3.10}$$

$$\sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j)\sigma(z_k) \leq q_k - q_0 + r_k. \tag{3.11}$$

b) Assume that the dual solution set D_* is nonempty. If $(\bar{y}, \bar{c}) \in D_*$ then for all k,

$$\|y_{k+1} - \bar{y}\|^2 \leq \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left(\frac{s_k\sigma(z_k)}{2} + \frac{q_k - \bar{q} + r_k}{\sigma(z_k)} + \bar{c} - c_k\right). \tag{3.12}$$

Proof. ¿From the update formula for $\{(y_k, c_k)\}$ we have

$$y_k = y_0 - \sum_{j=0}^{k-1} s_j z_j \ \text{ and } \ c_k = c_0 + \sum_{j=0}^{k-1}(1 + \alpha_j)s_j\sigma(z_j). \tag{3.13}$$

Hence

$$\langle y_k, z_k \rangle = \langle y_0, z_k \rangle - \sum_{j=0}^{k-1} s_j \langle z_j, z_k \rangle.$$

By Cauchy Schwarz inequality and (A_0),

$$\langle y_k, z_k \rangle \leq \langle y_0, z_k \rangle + \sum_{j=0}^{k-1} s_j \sigma(z_j)\sigma(z_k).$$

Using the expression for c_k given in (3.13) in the inequality above, we obtain

$$\langle y_k, z_k \rangle \leq \langle y_0, z_k \rangle - c_0\sigma(z_k) - \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j)\sigma(z_k) + c_k\sigma(z_k).$$

Adding $f(x_k, z_k)$ to both sides of this inequality, and observing that $\sigma \geq 0$, we have, after some simple algebra,

$$\begin{aligned} f(x_k, z_k) - \langle y_0, z_k\rangle &\leq f(x_k, z_k) - \langle y_0, z_k\rangle + c_0\sigma(z_k) + \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j)\sigma(z_k) \\ &\leq f(x_k, z_k) - \langle y_k, z_k\rangle + c_k\sigma(z_k) \leq q_k + r_k. \end{aligned}$$

¿From these inequalities, using the definition of q, we obtain

$$f(x_k, z_k) - \langle y_0, z_k\rangle \leq q_k + r_k \quad \text{and} \quad q_0 + \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j)\sigma(z_k) \leq q_k + r_k,$$

which are the statements of (a). For proving (b), take $(\bar{y}, \bar{c}) \in D_*$. For all k we have

$$\begin{aligned} \|y_{k+1} - \bar{y}\|^2 &= \|y_k - s_k z_k - \bar{y}\|^2 \\ &= \|y_k - \bar{y}\|^2 + s_k^2\|z_k\|^2 + 2s_k\langle \bar{y} - y_k, z_k\rangle \\ &\leq \|y_k - \bar{y}\|^2 + s_k^2\sigma(z_k)^2 + 2s_k(q_k - \bar{q} + r_k + \sigma(z_k)(\bar{c} - c_k)) \\ &= \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left(\frac{s_k\sigma(z_k)}{2} + \frac{q_k - \bar{q} + r_k}{\sigma(z_k)} + \bar{c} - c_k\right) \end{aligned}$$

where the inequality follows from (A_0) and the supergradient inequality. The result follows. □

LEMMA 3.6. *If the sequence $\{z_k\}$ converges weakly to 0, then $\{q_k\}$ converges to $\bar{q}$, the primal sequence $\{x_k\}$ is bounded, and all its weak accumulation points are primal solutions.*

Proof. Take an upper bound r of $\{r_k\}$. By Lemma 3.5(a), we have, for all k,

$$f(x_k, z_k) - \langle y_0, z_k\rangle \leq q_k + r_k \leq \bar{q} + r.$$

Rearranging this inequality and using Cauchy Schwarz inequality, we obtain

$$f(x_k, z_k) \leq \|y_0\|\|z_k\| + \bar{q} + r \leq \tilde{b} := \|y_0\|b + \bar{q} + r \quad \text{for all } k, \tag{3.14}$$

where b is an upper bound for $\{\|z_k\|\}$. Now by the level compactness assumption on f, there exists a weak open neighborhood $V \subset H$ of 0 and a weakly compact set $B \subset X$ such that

$$L_{V,f}(\tilde{b}) = \{x : f(x, z) \leq \tilde{b}\} \subset B, \text{ for all } z \in V.$$

Since $\{z_k\}$ is weakly convergent to 0, $z_k \in V$ for k sufficiently large. Hence $x_k \in L_{V,f}(\tilde{b})$ for k sufficiently large, by (3.14). Therefore $\{x_k\}$ is bounded. Take a weak accumulation point $\bar{x}$ of $\{x_k\}$. Thus there exists a subsequence $\{x_{k_j}\}$ which converges weakly to $\bar{x}$. In particular $\{(x_{k_j}, z_{k_j})\}$ converges weakly to $(\bar{x}, 0)$. Since $f(\cdot, \cdot)$ is w-lsc, we obtain

$$\varphi(\bar{x}) = f(\bar{x}, 0) \leq \liminf_j (f(x_{k_j}, z_{k_j}) - \langle y_0, z_{k_j}\rangle) \leq \liminf_j (q_{k_j} + r_{k_j}) \leq \bar{q}, \tag{3.15}$$

where the second inequality follows from Lemma 3.5 (a), and the third follows from the fact that $\{r_k\}$ converges to 0. Since $\bar{q} = M_P$ by Theorem 2.4, we obtain from (3.15) that $\varphi(\bar{x}) = M_P$, and then $\bar{x}$ is a primal solution. In particular, all inequalities in (3.15) are equalities. Since $\{r_k\}$ converges to 0, we obtain that $\liminf_j q_{k_j} = \bar{q}$. Since $\{q_k\}$ is increasing by Theorem 3.3, we conclude that $\{q_k\}$ converges to $\bar{q}$. The proof is complete. □

In order to obtain our results we consider the following assumption on the error sequence.

$(\mathbf{A_1})$: There exists $R > 0$ such that $r_k \leq \bar{q} - q_k + R\sigma(z_k)$ for all k.

REMARK 3.1. We mention that the verification of condition (A_1) is not immediate, since in general at iteration k we ignore the values of both $\bar{q}$ and q_k. One alternative is to think of an "a posteriori" verification of (A_1), meaning that we check the boundedness of the sequence $\{\frac{r_k}{\sigma(z_k)}\}$. Indeed, note that when the latter sequence is bounded, then (A_1) holds, because $\frac{\bar{q}-q_k}{\sigma(z_k)} \geq 0$. The situation improves considerably when we know the optimal dual value $\bar{q}$, a situation which occurs in many "real life" problems. In this case, we can verify condition (A_1) along the iterations of the algorithm. In order to do this, we observe that the value $L_k := f(x_k, z_k) - \langle y_k, z_k\rangle + c_k\sigma(z_k)$ is computable, and satisfies $\hat{r}_k := L_k - q_k \leq r_k$ for each k. The (unknown) value $\hat{r}_k$ is in fact the actual error in the kth-iteration, while the (known) value r_k can be seen as an estimate of $\hat{r}_k$. We assert that we can take $\hat{r}_k$ instead of r_k in condition (A_1). Observe that, since $L_k = q_k + \hat{r}_k$, we should verify the following condition:

$(\mathbf{\hat{A}_1})$: There exists $R > 0$ such that $L_k \leq \bar{q} + R\sigma(z_k)$ for all k.

This condition is checkable when we know the optimal dual value $\bar{q} = M_P$. Another possibility is to think of IMSg as "measuring" at each iteration the boundedness of $\frac{r_k}{\sigma(z_k)}$, meaning that we observe whether the condition $\frac{r_k}{\sigma(z_k)} \leq R$ is satisfied, where $R > 0$ is an "a priori" given parameter. For those values of k such that this inequality does not hold, one should consider an exact step $(r_k = 0)$. Another option consists of applying the inexact algorithm IMSg just for a finite number of iterations and then switch to the exact version of IMSg, i.e., with $r_k = 0$.

LEMMA 3.7. *If $\{c_k\}$ is bounded then $\{y_k\}$ is also bounded. If the dual solution set is nonempty then the converse of the previous statement holds.*

Proof. The first statement follows directly from (3.2) and (3.3). For proving the last statement, we rewrite the supergradient inequality as follows:

$$c_k \leq \bar{c} + \frac{q_k - \bar{q} + r_k - \langle \bar{y} - y_k, z_k\rangle}{\sigma(z_k)},$$

where $(\bar{y}, \bar{c}) \in D_*$. Using $(A_0), (A_1)$ and Cauchy-Schwarz inequality we obtain

$$c_k \leq \bar{c} + R + \|\bar{y} - y_k\| \tag{3.16}$$

where the constant R is given by (A_1). The last statement now follows from (3.16). □

Next we propose and analyze two algorithms related to IMSg. We remark that the difference between them lies in the stepsize selection rule.

3.1. Algorithm 1. Take two parameters $\beta > \eta > 0$. In Step-2 of the k-th iteration of IMSg, take $\eta_k := \min\{\eta, \|z_k\|\}$ and $\beta_k := \max\{\beta, \sigma(z_k)\}$, and choose a stepsize $s_k \in [\eta_k, \beta_k]$, for all k. We denote this algorithm by IMSg-1.

REMARK 3.2. By definition of η_k, β_k and (A_0) we have $\eta_k \leq \|z_k\| \leq \sigma(z_k) \leq \beta_k$. In particular we see that $\|z_k\|$ and $\sigma(z_k)$ are simple choices for the stepsize s_k. Observe that since $[\eta, \beta] \subset [\eta_k, \beta_k]$ for all k, we can choose any stepsize $s \in [\eta, \beta]$. In particular, a constant stepsize for all iterations is admissible.

The next theorem establishes some basic convergence properties of the dual sequence generated by IMSg-1.

THEOREM 3.8. *Assume that* IMSg-1 *generates an infinite dual sequence $\{(y_k, c_k)\}$. If the dual optimal set is nonempty then $\{(y_k, c_k)\}$ is bounded and all its weak accumulation points are dual solutions; if the dual optimal set is empty then $\{(y_k, c_k)\}$ is unbounded.*

Proof. First, we prove that $\{(y_k, c_k)\}$ is bounded when $D_* \neq \emptyset$. Observe that $s_k \leq \beta_k \leq \max\{\beta, b\}$, where b is an upper bound for $\sigma(z_k)$ (see Lemma 3.4). Thus $s_k\sigma(z_k) \leq \hat{b} := \max\{b\beta, b^2\}$

for all k. Let R be as in (A_1), and take $(\bar{y}, \bar{c}) \in D_*$. If we show that $\{c_k\}$ is bounded, then $\{(y_k, c_k)\}$ will be bounded, by Lemma 3.7. Suppose by contradiction that $\{c_k\}$ is not bounded. Thus there exists k_0 such that $c_k \geq M := \frac{\hat{b}}{2} + R + \bar{c}$ for all $k \geq k_0$. Using these estimates in (3.12), we obtain

$$\|y_{k+1} - \bar{y}\| \leq \|y_k - \bar{y}\| + 2s_k\sigma(z_k)\left(\frac{\hat{b}}{2} + R + \bar{c} - c_k\right) \leq \|y_{k_0} - \bar{y}\|,$$

for all $k \geq k_0$. It follows that $\{y_k\}$ is bounded. This entails a contradiction, in view of Lemma 3.7. Therefore the dual sequence is bounded.

Let us prove now that all weak accumulation points of $\{(y_k, c_k)\}$ are dual solutions. In particular this also proves the last statement of the theorem by contradiction. Since $\{(y_k, c_k)\}$ is bounded, we know that $\sum_k s_k\sigma(z_k) < \infty$, by Proposition 3.1. In particular $\{s_k\sigma(z_k)\}$ converges to 0. On the other hand, using (A_0) and the fact that $s_k \geq \min\{\eta, \|z_k\|\}$ we obtain

$$s_k\sigma(z_k) \geq \min\{\eta\|z_k\|, \|z_k\|^2\} > 0.$$

Since $\eta > 0$, we conclude that $\{\|z_k\|\}$ converge to 0. In particular $\{z_k\}$ converges weakly to 0. Now Lemma 3.6 ensures that $\{q_k\}$ converges to $\bar{q}$. Take a weak accumulation point $(\bar{y}, \bar{c})$ of $\{(y_k, c_k)\}$, so that there exists a subsequence $\{(y_{k_j}, c_{k_j})\}$ weakly convergent to $(\bar{y}, \bar{c})$. Since the dual function is w-usc (see Proposition 2.3), we have

$$\bar{q} \geq q(\bar{y}, \bar{c}) \geq \limsup_j q(y_{kj}, c_{kj}) = \lim_j q_{kj} = \bar{q}.$$

Hence $q(\bar{y}, \bar{c}) = \bar{q}$ and we conclude that $(\bar{y}, \bar{c})$ is a dual optimal solution. In particular, boundedness of the dual sequence implies that the dual solution set D_* is nonempty, which establishes the theorem. □

Theorem 3.8 establishes dual convergence results of IMSg-1. The next theorem establishes primal convergence results.

THEOREM 3.9. *Consider the primal sequence $\{x_k\}$ generated by* IMSg-1. *Suppose that there exists $\bar{\alpha} > 0$ such that $\bar{\alpha} \leq \alpha_k$ for all k. Then $\{q_k\}$ converges to $\bar{q}$, the primal sequence $\{x_k\}$ is bounded and all its weak accumulation points are primal solutions.*

Proof. Take the dual sequence $\{(y_k, c_k)\}$ generated by IMSg-1. If $\{(y_k, c_k)\}$ is bounded, we can use the same argument used in the second part of the proof of Theorem 3.8 for ensuring that $\{z_k\}$ converges weakly to 0. Thus, in the case that $\{(y_k, c_k)\}$ is bounded the result follows from Lemma 3.6. Hence we just need to consider the case in which the dual sequence is unbounded. In this case we get from Lemma 3.5(a)

$$\sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j)\sigma(z_k) \leq q_k - q_0 + r_k. \tag{3.17}$$

On the other hand, $\{r_k\}$ is bounded and $q_k - q_0 \leq \bar{q} - q_0$ for all k. Thus there exists $\hat{M} > 0$ such that $q_k - q_0 + r_k \leq \hat{M}$ for all k. Using this estimate in (3.17), together with the fact that $\alpha_k \geq \bar{\alpha}$ for all k, we obtain

$$\bar{\alpha}\left(\sum_{j=0}^{k-1} s_j\sigma(z_j)\right)\sigma(z_k) \leq \hat{M} \tag{3.18}$$

for all $k \geq 1$. Since the dual sequence is unbounded, Proposition 3.1(ii) implies that $\sum_{j=0}^{\infty} s_j\sigma(z_j) = \infty$. Using this fact in (3.18), since $\bar{\alpha} > 0$, it follows that $\{\sigma(z_k)\}$ converges to zero. By assumption

(A_0) we get that $\{\|z_k\|\}$ converges to 0, and thus $\{z_k\}$ converges weakly to 0. The result now follows from Lemma 3.6. □

In order to establish convergence of the whole dual sequence, we need some preliminary material on Fejér convergence.

DEFINITION 3.10. *Let H be a Hilbert space and V a nonempty subset of H. A sequence $\{z_k\} \subset H$ is said to be* quasi-Fejér *convergent to V if and only if for all $\bar{z} \in V$ there exists some sequence $\{\mu_k\} \subset \mathbb{R}_+$ such that $\sum_k \mu_k < \infty$ and*

$$\|z_{k+1} - \bar{z}\|^2 \leq \|z_k - \bar{z}\|^2 + \mu_k.$$

LEMMA 3.11. *Consider a Hilbert space H and a sequence $\{\xi_k\} \subset H$. If $\{\xi_k\}$ is* quasi-Fejér *convergent to some set $V \neq \emptyset$, then*
a) The sequence $\{\xi_k\}$ is bounded;

b) $\{\|\xi_k - \bar{v}\|\}$ is convergent for all $\bar{v} \in V$;

c) if all the weak accumulation points of $\{\xi_k\}$ are in V, then the sequence $\{\xi_k\}$ is weakly convergent to some $\bar{v} \in V$.

Proof. See for example [1, Proposition 1]. □

Next we establish Fejér convergence of the dual sequence generated by IMSg-1 to an appropriate subset of the dual solution.

PROPOSITION 3.12. *Consider the dual sequence $\{(y_k, c_k)\}$ generated by* IMSg-1. *If D_* is nonempty, then $\{(y_k, c_k)\}$ is quasi-Fejér convergent to the set $V_* = \{(y, c) \in D_* : c \geq c_k\ \forall k\}$.*

Proof. Since D_* is nonempty, it follows from Theorem 3.8 and Proposition 2.3 (ii) that there exists some $(\bar{y}, \bar{c}) \in D_*$ such that $\bar{c} \geq c_k$ for all k, i.e., V_* is nonempty. Take any $(\bar{y}, \bar{c}) \in V_*$. Consider $d_k := \|(\bar{y}, \bar{c}) - (y_k, c_k)\|$. Using the updating formula for the dual sequence, we have, for all k,

$$\begin{aligned}
d_{k+1}^2 &= \left\|(\bar{y}, \bar{c}) - \left(y_k - s_k z_k, c_k + (1+\alpha_k) s_k \sigma(z_k)\right)\right\|^2 \\
&= d_k^2 + s_k^2 \|z_k\|^2 + (1+\alpha_k)^2 s_k^2 \sigma(z_k)^2 \\
&+ 2 s_k \left[\langle \bar{y} - y_k, z_k \rangle - (1+\alpha_k)\sigma(z_k)(\bar{c} - c_k)\right] \\
&\leq d_k^2 + s_k^2 \sigma(z_k)^2 + (1+\alpha)^2 s_k^2 \sigma(z_k)^2 \\
&+ 2 s_k \left[\langle \bar{y} - y_k, z_k \rangle - \sigma(z_k)(\bar{c} - c_k)\right],
\end{aligned}$$

where the inequality follows from (A_0) and the fact that $\alpha > \alpha_k > 0$. Now, using the supergradient inequality, we obtain

$$d_{k+1}^2 \leq d_k^2 + \left(1 + (1+\alpha)^2\right) s_k^2 \sigma(z_k)^2 + 2 s_k (q_k - \bar{q} + r_k). \tag{3.19}$$

By (A_1) we get $R > 0$ such that $q_k - \bar{q} + r_k \leq R\sigma(z_k)$. Using this estimate in (3.19) and considering $\hat{\alpha} := 1 + (1+\alpha)^2$, we have

$$d_{k+1}^2 \leq d_k^2 + \hat{\alpha} s_k^2 \sigma(z_k)^2 + 2R s_k \sigma(z_k). \tag{3.20}$$

On the other hand, Theorem 3.8 ensures boundedness of the dual sequence. Hence we have $\sum_k s_k \sigma(z_k) < \infty$, by Proposition 3.1, which in turn implies that $\sum_k s_k^2 \sigma(z_k)^2 < \infty$. Consider

$\mu_k := \hat{\alpha} s_k^2 \sigma(z_k)^2 + 2R s_k \sigma(z_k)$ for all k. We see that $\sum_k \mu_k < \infty$ and by (3.20) we obtain

$$d_{k+1}^2 \leq d_k^2 + \mu_k$$

for all k. The result follows from Definition 3.10. □

Now we establish weak convergence of the whole dual sequence generated by IMSg-1 to a dual solution.

THEOREM 3.13. *If the dual solution set is nonempty, then the dual sequence generated by* IMSg-1 *is weakly convergent to some dual solution.*

Proof. By Theorem 3.8, the dual sequence $\{(y_k, c_k)\}$ is bounded and all its weak accumulation points belong to $V_* = \{(y_*, c_*) \in D_* : c_* \geq c_k \ \forall k\}$ (observe that $\{c_k\}$ is increasing). By Proposition 3.12 this sequence is quasi-Fejér convergent to V_*. By Lemma 3.11(c), the sequence is weakly convergent to some $(\bar{y}, \bar{c}) \in V_* \subset D_*$. □

3.2. Algorithm 2. In this section we propose a stepsize selection which ensures that IMSg converges in a finite number of steps.

Take $\beta > 0$ and a sequence $\{\theta_k\} \subset \mathbb{R}_+$ such that $\sum_j \theta_j = \infty$, and $\theta_k \leq \beta$ for all k. In Step-2 of the k-th iteration of IMSg, consider $\eta_k := \dfrac{\theta_k}{\sigma(z_k)}$ and $\beta_k := \dfrac{\beta}{\sigma(z_k)}$, and choose a stepsize $s_k \in [\eta_k, \beta_k]$, for all k. IMSg with this stepsize selection is denoted by IMSg-2.

THEOREM 3.14. *a) Suppose that the dual solution set D_* is nonempty. Let $\{(x_k, z_k)\}$ and $\{(y_k, c_k)\}$ be the primal and dual sequences generated by* IMSg-2. *Then there exists a $\bar{k}$ such that* IMSg-2 *stops at iteration $\bar{k}$. As a consequence $x_{\bar{k}}$ and $(y_{\bar{k}}, c_{\bar{k}})$ are ϵ_*-optimal primal and ϵ_*-optimal dual solutions respectively.*

b) Suppose that IMSg-2 *generates infinite primal dual sequences $\{(x_k, z_k)\}$ and $\{(y_k, c_k)\}$, (in this case D_* is empty by (a)). Then $\{(y_k, c_k)\}$ is unbounded, $\{\|z_k\|\}$ converges to 0, and $\{q_k\}$ converges to the optimal value $\bar{q}$. The primal sequence $\{x_k\}$ is bounded and all its weak accumulation points are primal solutions.*

Proof. a) Taking an upper bound $\hat{b}$ for $\{s_k \sigma(z_k)\}$ and repeating the first part of the proof of Theorem 3.8, it follows that $\{(y_k, c_k)\}$ is bounded. In particular, we obtain $\sum_j s_j \sigma(z_j) < \infty$, in view of Proposition 3.1. Observe now that the criterion $r_k < \varepsilon_*$ in Step-1 of IMSg-2 is satisfied after a finite number of iterations, because $\{r_k\}$ converges to 0. Take $\hat{k}$ such that $r_k < \varepsilon_*$ for all $k \geq \hat{k}$, and suppose that the stopping criterion of IMSg-2 is not satisfied for $k < \hat{k}$. In this situation IMSg-2 stops at iteration $\bar{k}$ (with $\bar{k} \geq \hat{k}$) if and only if $z_{\bar{k}} = 0$. On the other hand we know, by the stepsize selection rule of IMSg-2, that if $z_j \neq 0$ at some iteration j then $s_j \sigma(z_j) \geq \theta_j$. Hence the fact that $\sum_j s_j \sigma(z_j) < \infty$ and $\sum_j \theta_j = \infty$, ensures that there exists $\bar{k} > \hat{k}$ such that $z_{\bar{k}} = 0$. Therefore IMSg-2 stops at iteration $\bar{k}$, and by Theorem 3.1 we conclude that $x_{\bar{k}}$ is an ϵ_*-optimal primal solution and $\{(y_{\bar{k}}, c_{\bar{k}})\}$ is an ϵ_*-optimal dual solution.

For proving (b) we observe that since IMSg-2 generates infinite primal and dual sequences, we have $z_k \neq 0$ for all k. By the stepsize selection rule of IMSg-2 we have $s_k \sigma(z_k) \geq \theta_k$ for all k. In particular $\sum_k s_k \sigma(z_k) = \infty$, which is equivalent to $\{(y_k, c_k)\}$ be unbounded, by Proposition 3.1. Now the result follows by using the same argument of the second part of Theorem 3.9. □

The next proposition establishes that if Assumption (A_0) does not hold, then the conclusion of Theorem 3.3 may fail.

PROPOSITION 3.15. *Let H be a Hilbert space. Suppose that there exists some $0 \neq \bar{u} \in H$ such that $\sigma(\bar{u}) = \gamma_1 \|\bar{u}\|$ with $\gamma_1 < 1$. Moreover suppose that $\sigma(-\bar{u}) = \gamma_2 \|\bar{u}\|$, with $\gamma_1 \gamma_2 < 1$. In this situation, the conclusion of* Theorem 3.3 *may fail.*

Proof. Observe that we only need to find a problem such that $q_1 < q_0 < M_D$. First, consider a w-lsc function $g : H \to \mathbb{R}$, such that $\bar{u}, -\bar{u} \in \operatorname{argmin}(g+\sigma)$ and $\min(g+\sigma) = 0 < g(0)$, (for example $g(x) = -\sigma(x)$ if $x \in \{\bar{u}, -\bar{u}\}$, $g(x) = 1$ otherwise). Let $K \subset H$ be a weakly compact set such that $\{\bar{u}, -\bar{u}, 0\} \subset K$. Consider the following primal problem

$$\min \ f(x) := g(x) + \langle \bar{u}, x\rangle \ \text{ s.t. } x \in K, \ h(x) := x = 0. \tag{3.21}$$

Let $C = \{x \in K : h(x) = 0\}$ and $\delta_C(x) = 0$ if $x \in C$, $\delta_C(x) = \infty$ otherwise (observe that $C = \{0\}$, because $0 \in K$ and $h(x) = x$). Take $\bar{\phi}(x) = f(x) + \delta_C(x)$ for each $x \in H$. Therefore the problem (3.21) is equivalent to

$$\min \ \bar{\phi}(x) \ \text{ s.t. } x \in H,$$

Consider now a dualizing parameterization function given by

$$\bar{f}(x,u) = \begin{cases} f(x) & \text{if } x \in K \text{ and } x = u,\\ \infty & \text{otherwise.}\end{cases}$$

Then

$$\ell(x,y,c) = \inf_u \{\bar{f}(x,u) - \langle y, u\rangle + c\sigma(u)\} = \begin{cases} f(x) - \langle y, x\rangle + c\sigma(x) & \text{if } x \in K\\ \infty & \text{otherwise,}\end{cases}$$

and therefore

$$q(y,c) = \inf_{x\in H} \ell(x,y,c) = \inf_{x\in K}\{f(x) - \langle y, x\rangle + c\sigma(x)\} = \inf_{x\in K}\{g(x) - \langle y - \bar{u}, x\rangle + c\sigma(x)\}.$$

Since K is weakly compact and $\bar{f}$ is w-lsc, the hypotheses of Theorem 2.4 are satisfied, and therefore $\sup_{(y,c)\in H\times\mathbb{R}_+} q(y,c) = \inf_{x\in H} \bar{\phi}(x) =: v$ (observe that $v = g(0)$). We can easily verify that

$$q(\overline{u}, 1) = \inf_{x\in K}\{g(x) + \sigma(x)\} = g(\overline{u}) + \sigma(\overline{u}) = 0.$$

Observe also that $0 = g(\overline{u}) + \sigma(\overline{u}) = \bar{f}(\overline{u}, \overline{u}) - \langle \overline{u}, \overline{u}\rangle + \sigma(\overline{u})$. Therefore, $(\overline{u}, \overline{u}) \in A(\overline{u}, 1) := A_0(\overline{u}, 1)$, (see definition of $A_r(y,c)$ in (2.5)). Consider $y_0 = \overline{u}$, $c_0 = 1$ and $r_0 = 0$ in Step-0 of IMSg. If we take $(\overline{u}, \overline{u}) \in A(y_0, c_0)$ as the solution of the subproblem (see IMSg) then

$$y_1 = y_0 - s_0\bar{u}, \text{ and } c_1 = c_0 + (1+\alpha_0)s_0\sigma(\bar{u}),$$

where $s_0 > 0$ is an initial stepsize and α_0 satisfies $\alpha_0 < \dfrac{1}{\gamma_1\gamma_2} - 1$. Hence

$$\begin{aligned} \ell(-\bar{u}, y_1, c_1) &= f(-\bar{u}) + \langle y_1, \bar{u}\rangle + c_1\sigma(-\bar{u})\\ &= f(-\bar{u}) + \langle y_0 - s_0\bar{u}, \bar{u}\rangle + [c_0 + (1+\alpha_0)s_0\sigma(\bar{u})]\,\sigma(-\bar{u})\\ &= f(-\bar{u}) + \|\bar{u}\|^2 + \sigma(-\bar{u}) + s_0\left[(1+\alpha_0)\sigma(\bar{u})\sigma(-\bar{u}) - \|\bar{u}\|^2\right]\end{aligned}$$

where we use $y_0 = \overline{u}$ and $c_0 = 1$. Observing that $g(-\overline{u}) + \sigma(-\overline{u}) = 0$ and using the definition of f we obtain

$$\ell(-\bar{u}, y_1, c_1) = s_0\left[(1+\alpha_0)\sigma(\bar{u})\sigma(-\bar{u}) - \|\bar{u}\|^2\right] = s_0\left[(1+\alpha_0)\gamma_1\gamma_2 - 1\right]\|\bar{u}\|^2 < 0,$$

where we use the assumptions on $\sigma(\overline{u}), \sigma(-\overline{u})$ and α_0. Now by definition of $q(y_1, c_1)$ we get

$$q(y_1, c_1) \le \ell(-\bar{u}, y_1, c_1) < 0 = q(y_0, c_0) < g(0) = v,$$

that is, $q(y_1, c_1) < q(y_0, c_0) < v = M_D$. The proof is complete. □

REMARK 3.3. The assumption $\gamma_1\gamma_2 < 1$ used in Proposition 3.15 is satisfied when σ is even (i.e., when $\sigma(z) = \sigma(-z)$ for all z) and there exists some $\bar{u}$ such that $\sigma(\bar{u}) < \|\bar{u}\|$. An example of such a function is $\sigma(z) = \|z\|^t$ for $t > 1$. Another choice of σ for which Theorem 3.3 may be false is when $\sigma(\cdot) \geq \gamma\|\cdot\|$ with $0 < \gamma < 1$. In the latter case, the following simple modification of the algorithm ensures the increasing property of the dual values. Consider a sequence $\{t_k\}$ such that $t_k \geq \frac{1}{\gamma^2}$ for all k, and update the parameter c_k as follows,

$$c_{k+1} := c_k + (\alpha_k + t_k)s_k\sigma(z_k).$$

The proof of the increasing property of the dual values is similar to the one given in Theorem 3.3 and it is omitted. We claim also that if we consider $\{t_k\} \subset [\frac{1}{\gamma^2}, d_1]$ for some $d_1 > \frac{1}{\gamma^2} > 0$, then Theorems 3.8, 3.9, 3.13 and Theorem 3.14 remain valid for this modification with essentially the same proofs.

3.3. Modified Subgradient Algorithm with Sharp Lagrangian. Some versions of the modified subgradient algorithm (MSG) with sharp Lagrangian in finite dimensional spaces were proposed in [2, 4, 9, 6]. In [10] the authors considered an optimization problem which contains just one inequality constraint and analyzed a version of MSG for the dual problem generated via augmented Lagrangian. The MSG proposed in [2, 6, 9, 10] depends strongly on the knowledge of the primal optimal value. This drawback was avoided in [4]. In the present paper we propose an inexact version of the MSG proposed in [4] and we extend the convergence results to augmented Lagrangians more general than the sharp Lagrangian. In [6] the authors proposed an inexact version of MSG proposed in [2]. In this section we compare our algorithm with these previous versions of MSG, giving special attention to the search direction. We also compare our assumption (A_1) with the assumption on the error sequence $\{r_k\}$ used in [6]. For this purpose, consider Example 2.1, for which we have

$$\begin{aligned} A_r(y,c) &= \{(x,z) \in X \times H : f(x,z) - \langle y,z\rangle + c\|z\| \leq q(y,c) + r\} \\ &= \{(x,h(x)) \in K \times H : \psi(x) - \langle y,h(x)\rangle + c\|h(x)\| \leq q(y,c) + r\} \\ &= \{(x,h(x)) : x \in \Gamma_r(y,c)\}, \end{aligned}$$

where $\Gamma_r(y,c) = \{x \in K : L(x,y,c) := \psi(x) - \langle y, h(x)\rangle + c\|h(x)\| \leq q(y,c) + r\}$, which is precisely the set $X_r(y,c)$ considered in [6, Equation (6)] in a finite dimensional setting. Moreover, the set $A_r(y,c)$ is completely determined if we know $\Gamma_r(y,c)$. Therefore, at iteration k we update the search direction as $(-z_k, \sigma(z_k)) = (-h(x_k), \|h(x_k)\|)$ with $x_k \in \Gamma_{r_k}(y_k, c_k)$, which is the same search direction considered in [6]. In particular, for $r_k = 0$ for all k, we obtain the MSG proposed in [4], which is the exact version of IMSg with sharp Lagrangian. Thus we have also extended to reflexive Banach spaces the MSG method proposed in [4]. We also consider more general augmented Lagrangian functions than the sharp Lagrangian considered in [4, 2, 9, 6].

Next, we compare our assumption on the error sequence $\{r_k\}$ (condition (A_1)) with the assumption considered in [6]. First, we look again at Example 2.1 in a finite dimensional setting, with $\sigma(\cdot) = \|\cdot\|$, and $z_k = h(x_k)$. It is easy to see that assumptions (A_1) and (A_2) considered in [6, Section 4] are equivalent to the following assumption: there exist $\eta > 0$ and $M > 0$ such that, for all k,

$$\eta\frac{(\bar{q} - q_k + r_k)}{\sigma(z_k)} \leq s_k\sigma(z_k) \leq M.$$

In particular, from these inequalities we obtain $r_k < \frac{M}{\eta}\sigma(z_k)$ for all k. We remark that our assumption (A_1) on the error sequence $\{r_k\}$ is an improvement over this last estimate.

REFERENCES

[1] Ya.I. Alber, A.N. Iusem and M.V. Solodov, *On the projected subgradient method for nonsmooth convex optimization in a Hilbert space.* -Math. Program., 81 (1998), pp. 23-35.

[2] R.S. Burachik, R.N. Gasimov, N.A. Ismayilova and C.Y. Kaya, *On a modified subgradient algorithm for dual problems via sharp augmented Lagrangian.* J. Global Optim., 34 (2006), pp. 55-78.

[3] R.S. Burachik, A.N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer, Berlin, 2008.

[4] R.S. Burachik, A.N. Iusem and J.G. Melo, *A Primal Dual Modified Subgradient Algorithm with Sharp Lagrangian,* (submitted).

[5] R.S. Burachik and C.Y. Kaya, *An update rule and a convergence result for a penalty function method,* J. Ind. Manag. Optim. 3 (2007), pp. 381-398.

[6] R.S. Burachik, C.Y. Kaya and M. Mammadov, *An inexact modified subgradient algorithm for nonconvex optimization,* Comput. Optim. Appl. (2008), DOI 10.1007/s10589-008-9168-7.

[7] R.S. Burachik and A.M. Rubinov, *Abstract convexity and augmented Lagrangians,* SIAM J. Optim., 18 (2007), pp. 413-436.

[8] R.S. Burachik and A.M. Rubinov, *On the absence of duality gap for Lagrange-type functions* J. Ind. Manag. Optim. 1 (2005), pp. 33-38.

[9] R.N. Gasimov, *Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming,* J. Global Optim., 24 (2002), pp. 187-203.

[10] R.N. Gasimov and A.M. Rubinov, *On augmented Lagrangians for optimization problems with a single constraint,* J. Global Optim., 28 (2004), pp. 153-173.

[11] X.X. Huang and X.Q. Yang, *A unified augmented Lagrangian approach to duality and exact penalization,* Math. Oper. Res., 28 (2003), pp. 533-552.

[12] X.X. Huang and X.Q. Yang, *Further study on augmented Lagrangian duality theory,* J. Global Optim., 31 (2005), pp. 193-210.

[13] K. Ito and S. Kunisch *The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces* Math. Program., 46 (1990). pp. 341-360.

[14] K.Ito and S.Kunisch *Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces,* Nonlinear Anal., 41 (2000), pp. 591-616.

[15] A. Nedić and A. Ozdaglar, *Approximate Primal Solutions and Rate Analysis for Dual Subgradient Methods,* SIAM J. Optim., 19 (2009), pp. 1757-1780.

[16] A. Nedić and A. Ozdaglar, *Subgradient Methods for Saddle-Point Problems,* J. Optim. Theory Appl. (2009), DOI 10.1007/s10957-009-9522-7.

[17] J.P. Penot. *Augmented Lagrangians, duality and growth conditions,* J. Nonlinear Convex Anal., 3 (2002), pp. 283-302.

[18] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[19] A.M. Rubinov and X.Q. Yang, *Lagrange-type Functions in Constrained Non-convex Optimization,* Kluwer Academic, Dordrecht, The Netherlands, 2003.

[20] A.M. Rubinov, X.X. Huang and X.Q. Yang, *The zero duality gap property and lower semicontinuity of the perturbation function,* Math. Oper. Res., 27 (2002), pp. 775-791.

[21] S. Volkwein, *Mesh-independence for an augmented Lagrangian-SQP method in Hilbert spaces,* SIAM J. Control Optim., 38 (2000), pp. 767-785.

[22] Y.Y. Zhou and X.Q. Yang, *Augmented Lagrangian function, non-quadratic growth condition and exact penalization,* Oper. Res. Lett., 34 (2006), pp. 127-134.

[23] Y.Y. Zhou and X.Q. Yang, *Duality and Penalization in Optimization via an Augmented Lagrangian Function with Applications,* J. Optim. Theory Appl., 140 (2009), pp. 171-188.