A Primal Dual Modified Subgradient Algorithm with Sharp Lagrangian *

Regina S. Burachik† Alfredo N. Iusem ‡ Jefferson D. G. de Melo§

November 17, 2008

Abstract

We apply a modified subgradient algorithm (MSG) for solving the dual of a nonlinear and nonconvex optimization problem. The dual scheme we consider uses the sharp augmented Lagrangian. A desirable feature of this method is *primal convergence*, which means that every accumulation point of a primal sequence (which is automatically generated during the process), is a primal solution. This feature is not true in general for available variants of MSG. We propose here two new variants of MSG which enjoy both primal and dual convergence, as long as the dual optimal set is nonempty. These variants have a very simple choice for the stepsizes. Moreover, we also establish primal convergence when the dual optimal set is empty. Finally, our second variant of MSG converges in a finite number of steps.

Keywords. Nonsmooth optimization, nonconvex optimization, duality scheme, sharp Lagrangian, Modified subgradient algorithm.

1 Introduction

Duality is a very useful tool in optimization. The duality theory obtained through the ordinary (or classical) Lagrangian and its use for convex primal problems is well-known. However, when the primal problem is not convex, a duality gap may exist when the ordinary Lagrangian is used. This justifies the quest for other kinds of augmented Lagrangians, which are able to provide algorithms for solving a broader family of constrained optimization problems, including nonconvex ones. Recent literature has focused on dual problems

*Regina S. Burachik acknowledges support by the Australian Research Council Discovery Project Grant DP0556685 for this study; Jefferson D.G. de 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

constructed using *augmented* Lagrangian functions, where the augmenting functions are nonnegative and satisfy either coercivity assumptions ([11, Section 11], [7]) or peak-at-zero-type assumptions (see [13, 12, 4]).

A particular kind of such *augmented Lagrangian functions* is the sharp augmented Lagrangian, which has been recently studied for solving nonconvex and nonsmooth problems in [5, 1, 3].

We analyze the duality scheme generated by the sharp augmented Lagragian, when applied to nonsmooth and nonconvex problems. The dual problem is a nonsmoth convex problem (i.e. maximization of a concave function), and we can therefore solve it using the subgradient method or its variants. These methods (in general) converge slowly, but on the other hand we obtain a search direction in the calculation of the dual function with no extra cost, see Theorem 3.1(a). Gasimov in [5] proposed a modified subgradient algorithm for the same problem we are considering here. A deflection in the parameter ensures that the dual values are strictly increasing. This increasing property makes the modified subgradient algorithm very attractive, since (non-modified) subgradient methods in general do not have this property. Dual convergence results were obtained and numerical experiments were presented to illustrate the efficiency of the algorithm. In [1] the results of [5] were improved by relaxing the stepsize selection, and an example showing that the algorithm may fail to achieve primal convergence was presented. An auxiliary sequence, with an extra cost, was considered, and a primal convergence result was obtained for this sequence. In [3] an inexact version of the algorithm is proposed and analyzed. Similar results to those of the exact version were obtained. The applicability of these algorithms (exact and inexact versions) is better when we know the primal optimal value or at least a good estimate, see [1, Equations (16), (23) and Section 5.1] and also [3, Corollary 5.1]. In many problems even an approximate optimal value is both very hard to obtain and expensive. Therefore, it is desirable to look for a different stepsize selection rule, unrelated to the optimal value, and such that the convergence properties of the algorithm are preserved. It is also important to ensure convergence of the primal sequence generated by the algorithm.

In this paper we consider the same modified subgradient algorithm as [5, 1], but we propose a very simple stepsize selection rule. With this rule we get rid of the dependence on the optimal value. We obtain dual convergence results as in [5] and [1]. We also show that our algorithm has the property that all accumulation points of the primal sequence generated by the algorithm are primal solutions, and thus no auxiliary sequence (as required in [1]) is needed. This primal convergence is ensured even if the dual optimal set is empty. The latter result is very important, because, in general, it is impossible to know "a priori" whether the dual problem has optimal solutions. We also show that if there exists a dual solution, then it is possible to consider larger stepsizes, which ensure that after a finite number of iterations of the algorithm both primal and dual optimal solutions are reached. On the issue of finite convergence, see Remark 4.1 at the end of Section 4.

This paper is organized as follows. Section 2 contains some preliminary materials concerning on sharp Lagrangian and strong duality. In Section 3 we present the algorithm, as well as some basics results. In this section we also establish our main results, Theorem 3.3

and Theorem 3.4. In Section 4 we present a stepsize selection rule which ensures that the sequence generated by the algorithm reaches dual primal solutions after a finite number of iterations.

Our notation is the usual one: $\|\cdot\|$ is the Euclidean norm, $\langle\cdot,\cdot\rangle$ is the Euclidean scalar product, $\mathbb{R}_{++} := (0,\infty)$, $\mathbb{R}_{+} := [0,\infty)$, $\mathbb{R}_{+\infty} = \mathbb{R}\cup\{+\infty\}$, $\mathbb{R}_{-\infty} = \mathbb{R}\cup\{-\infty\}$, $\bar{\mathbb{R}} = \mathbb{R}_{-\infty}\cup\{\infty\}$.

2 Preliminaries

We consider the nonlinear (primal) optimization problem

$$\text{minimize } f(x) \quad \text{s.t.} \quad x \ \text{in } K, \quad h(x) = 0, \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, is lower semicontinuous, $h : \mathbb{R}^n \to \mathbb{R}^m$ is continuous and $K \subset \mathbb{R}^n$ is compact. We consider the sharp Lagrangian (see [11, Section 11]):

$$L(x,y,c) := f(x) - \langle y, h(x)\rangle + c\|h(x)\|. \tag{2.1}$$

Associated with the sharp Lagrangian we consider the dual function $q : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}$ defined by

$$q(y,c) = \inf_{x\in K} L(x,y,c)$$

and the dual augmented problem given by

$$\text{maximize } \ q(y,c) \ \text{ s.t } \ (y,c) \text{ in } \ \mathbb{R}^m \times \mathbb{R}_+. \tag{D}$$

The sharp Lagrangian is a particular case of a more general family of Augmented Lagrangians proposed by Rockafellar and Wets in [11, Section 11]. Results on weak and strong duality, saddle point properties, and exact penalty parameter were established in [11, Section 11]. Further extensions of these augmented Lagrangians and their properties have been studied, e.g., in [11, 12, 7, 8, 9, 13, 10]. For further use, we recall in this section some of the existing results for augmented Lagrangians. The corresponding proofs can be found in [11, 12, 7] and references therein.

We consider the following primal problem.

$$\text{minimize } \varphi(x) \quad \text{s.t. } \ x \in \mathbb{R}^n, \tag{2.2}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}_{+\infty}$, a proper, lsc function. A *dualizing parameterization* for φ is a function $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$, such that $\phi(x,0) = \varphi(x)$ for all $x \in \mathbb{R}^n$. Following [11, Section 11], the augmented Lagrangian $\ell : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ is defined as

$$\ell(x,y,c) := \inf_{z\in\mathbb{R}^m} \{\phi(x,z) - \langle y,z\rangle + c\sigma(z)\},$$

where $\sigma : \mathbb{R}^m \to \mathbb{R}_{+\infty}$ is a lsc and convex function such that

$$\operatorname{argmin} \sigma(x) = 0 \quad \text{and} \quad \sigma(0) = 0.$$

The dual function $\psi : \mathbb{R}^m \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ induced by the augmented Lagrangian ℓ is defined by

$$\psi(y, c) = \inf_{x \in \mathbb{R}^n} \ell(x, y, c).$$

The dual problem is given by

$$\text{maximize } \psi(y, c) \quad \text{s.t.} \quad (y, c) \in \mathbb{R}^m \times \mathbb{R}_+. \tag{2.3}$$

Denote by $\bar{m}_p$ and $\bar{m}_d$ the optimal primal and dual values respectively.

The next definition was introduced in [11, Definition 1.16].

Definition 2.1 *A function $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$, is said to be* level-bounded in x locally uniform in z, *if for all $\bar{z} \in \mathbb{R}^m$ and for all $\alpha \in \mathbb{R}$, there exist an open neighborhood of $\bar{z}$, $V \subset \mathbb{R}^m$ and a bounded set $B \subset \mathbb{R}^n$, such that*

$$L_{V,f}(\alpha) := \{x \in \mathbb{R}^n : f(x, z) \leq \alpha\} \subset B, \quad \text{for all } z \in V.$$

The following proposition summarizes some basic results concerning the primal problem (2.2) and its dual (2.3).

Proposition 2.1 *Consider the primal problem* (2.2) *and the dual augmented problem* (2.3). *The following statements hold.*

(i) The dual function ψ is a concave and upper semicontinuous function (usc).

(ii) If $r > c$ then $\psi(y, r) \geq \psi(y, c)$ for all $y \in \mathbb{R}^m$. In particular, if (y, c) is a dual optimal solution, then also (y, r) is a dual optimal solution.

Proof. Item (i) follows from the fact that ψ is the infimum of affine functions. Item (ii) is a consequence of the fact that σ is nonnegative. □

¿From now on we use the following notation: P_*, D_* are the primal and dual optimal solution sets respectively, M_P and M_D are the optimal primal and dual values respectively.

The fact that in our approach the penalty parameter c is a dual variable, together with the use of a sharp Lagrangian, has some interesting consequences on the structure of the dual solution set D_*, improving upon the result of Proposition 2.1(ii).

Proposition 2.2 *Take $(y^*, c^*) \in D_*$, $\rho > 0$, and define $\Delta_\rho = \{(y, c) : \|y - y^*\| \leq \rho, c \geq c^* + \rho\}$. Then $\Delta_\rho \subset D_*$ for all $\rho > 0$.*

Proof. Take $(y, c) \in \Delta_\rho$. By assumption $q(y^*, c^*) = M_D$. Therefore,

$$\begin{aligned} q(y, c) &= \inf_{x \in K} \{f(x) - \langle h(x), y \rangle + c\|h(x)\|\} \\ &= \inf_{x \in K} \{f(x) - \langle h(x), y^* \rangle + c^*\|h(x)\| + (c - c^*)\|h(x)\| - \langle h(x), y - y^* \rangle\} \\ &\geq \inf_{x \in K} \{f(x) - \langle h(x), y^* \rangle + c^*\|h(x)\| + (c - c^* - \|y - y^*\|)\|h(x)\|\} \\ &\geq \inf_{x \in K} \{f(x) - \langle h(x), y^* \rangle + c^*\|h(x)\| + (c - c^* - \rho)\|h(x)\|\} \\ &\geq \inf_{x \in K} \{f(x) - \langle h(x), y^* \rangle + c^*\|h(x)\|\} = q(y^*, c^*) = M_D, \end{aligned} \tag{2.4}$$

using Cauchy-Schwarz inequality in the first inequality, the fact that $\|y - y^*\| \leq \rho$ in the second one, and the fact that $c \geq c^* + \rho$ in the third one. We conclude from (2.4) that $q(y,c) = M_D$ and so $(y,c) \in D_*$. □

Corollary 2.1 *If* $(y^*, c^*) \in D_*$ *then* $\{(0,c) : c \geq c^* + \|y^*\|\} \subset D_*$.

Proof. Follows from Proposition 2.2, taking $\rho = \|y^*\|$. □

The next theorem guarantees that there is no duality gap for the primal-dual pair (2.2)-(2.3).

Theorem 2.1 *Consider the primal problem* (2.2) *and its dual augmented problem* (2.3). *Assume that the dualizing parameterization function* $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ *for the primal function* φ *is proper, lsc and level bounded in* x *locally uniform in* z. *Suppose that there exists some* $(y,r) \in \mathbb{R}^m \times \mathbb{R}_+$, *such that* $\psi(y,r) > -\infty$. *Then zero duality gap holds, i.e.* $\bar{m}_p = \bar{m}_d$.

Proof. See for instance, [11, Theorem 11.59]. □

It is not difficult to see that the sharp Lagrangian is an augmented Lagrangian. For $z \in \mathbb{R}^m$, define $\Omega(z) := K \cap \{x : h(x) = z\}$, where K and h are as in problem (P). Given $A \subset \mathbb{R}^n$, $\delta_A : \mathbb{R}^n \to \mathbb{R}_{+\infty}$ is defined as $\delta_A(x) = 0$, if $x \in A$; and $\delta_A(x) = \infty$ otherwise. Let

$$\varphi(x) := \begin{cases} f(x) & x \in \Omega(0), \\ \infty & \text{otherwise,} \end{cases}$$

where f is as in problem (P). Consider now a dualizing parameterization function given by $\phi(x,z) = f(x) + \delta_{\Omega(z)}(x)$. It is easy to see that $\phi(x,0) = \varphi(x)$ for all $x \in \mathbb{R}^n$. Using also (2.1) and $\sigma := \|\cdot\|$, we have that $\ell(x,y,c) = \delta_K(x) + L(x,y,c)$ and $q(y,c) = \psi(y,c)$. Thus, since K is compact, it is easy to see that the hypotheses of Theorem 2.1 are verified. In particular there is no duality gap between primal problem (P) and dual problem (D). We give next some definitions.

Definition 2.2 *Consider a concave function* $q : \mathbb{R}^p \to \mathbb{R}_{-\infty}$. *The superdifferential of* q *at* $y_0 \in \mathrm{dom}(q) := \{y \in \mathbb{R}^p \ : \ q(y) > -\infty\}$ *is the set* $\partial q(y_0)$ *defined by*

$$\partial q(y_0) := \{z \in \mathbb{R}^p : q(y) \leq q(y_0) + \langle z, y - y_0 \rangle \ \ \forall y \in \mathbb{R}^p\}.$$

We mention that the set $\partial q(y_0)$ is called *subdifferential* in [1, 3, 5]. Since $q(\cdot)$ is concave, we prefer our notation in order to avoid any confusion between the set above and the subdifferential of a convex function, where the inequality is reversed.

¿From now on we use the following notation: P_*, D_* are the primal and dual optimal solution sets respectively, M_P and M_D are the optimal primal and dual values respectively, Consider the following set

$$A(y,c) = \{x \in K \subset \mathbb{R}^n : L(x,y,c) = q(y,c)\}. \tag{2.5}$$

Note that $A(y,c) = \text{argmin}_{x \in K} L(x,y,c)$. Since K is compact, f is a lsc function and h is a continuous function, we have that $L(\cdot, y, c)$ is a lsc function for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$, and $A(y,c)$ is nonempty for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$. Thus, we have that $q(y,c) > -\infty$, for all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_+$, and also $M_P > -\infty$. In particular, since by Proposition 2.1(ii) the dual function q is concave, we conclude also that q is continuous (note that q can be extended in a natural way to $\mathbb{R}^m \times \mathbb{R}$, preserving its concavity).

3 Algorithm 1

We state next our first version of the Modified Subgradient Algorithm (MSG-1).

Step 0. Choose $(y_0, c_0) \in \mathbb{R}^m \times \mathbb{R}_+$, and exogenous parameters, $\{\epsilon_k\} \subset \mathbb{R}_{++}$. Also fix $\beta \geq \eta > 0$. Set $k := 0$.

Step 1 (**Subproblem and Stopping Criterion)**

a) Find $x_k \in A(y_k, c_k)$,

b) if $h(x_k) = 0$ stop,

c) if $h(x_k) \neq 0$, go to **Step** 2.

Step 2 (**Stepsizes selection and Update of Dual Variable)**

$\eta_k := \min\{\eta, \|h(x_k)\|\}$, $\beta_k := \max\{\beta, \|h(x_k)\|\}$, and choose s_k in $[\eta_k, \beta_k]$,

$y_{k+1} := y_k - s_k h(x_k)$,

$c_{k+1} := c_k + (\epsilon_k + s_k)\|h(x_k)\|$.

Set $k = k+1$ and go to **Step** 1.

Note that $[\eta, \beta] \subset [\eta_k, \beta_k]$. In particular, if we consider $\eta = \beta$ then we see that constant stepsizes ($s_k = \eta$, for all k) can be considered. The parameter ϵ_k (which "modifies" the classical subgradient step) was proposed by Gasimov in [5]. It ensures that the dual values are strictly increasing. It is well known that *pure* subgradient methods (i.e., when $\epsilon_k \equiv 0$) in general do not have this property. This is a special characteristic of this modified subgradient algorithm. The stepsize selection rule given above has not been considered in [5, 1, 6, 3]. In all these references, some knowledge of the optimal value is required (see, for instance [3, Corollary 5.1]). The next theorem establishes the relation between the minimization implicit in $A(y,c)$ and the superdifferential $\partial q(y,c)$. This result can be found in [3]. However, we prove it here for the sake of completeness.

Theorem 3.1 *The following results hold for* MSG-1.

a) If $\hat{x} \in A(\hat{y}, \hat{c})$, *then* $(-h(\hat{x}), \|h(\hat{x})\|) \in \partial q(\hat{y}, \hat{c})$.

b) If $\epsilon_k = \alpha_k s_k$, $(\alpha_k > 0)$, *with* $\{\alpha_k\}$ *bounded, then* MSG-1 *generates a dual bounded sequence* $\{(y_k, c_k)\}$ *if and only if* $\sum_k s_k \|h(x_k)\| < +\infty$.

c) If MSG-1 *stops at iteration* k, *then* x_k *is an optimal primal solution, and* (y_k, c_k) *is an* *optimal dual solution.*

Proof.

a) For all $(y,c) \in \mathbb{R}^m \times \mathbb{R}_{++}$ we have

$$\begin{aligned} q(y,c) &= \min_{x\in K}\{f(x) - \langle h(x), y\rangle + c\|h(x)\|\} \\ &\leq f(\hat{x}) - \langle h(\hat{x}), y\rangle + c\|h(\hat{x})\| \qquad (3.1) \\ &= f(\hat{x}) - \langle h(\hat{x}), \hat{y}\rangle + \hat{c}\|h(\hat{x})\| + \langle -h(\hat{x}), y-\hat{y}\rangle + (c-\hat{c})\|h(\hat{x})\|. \end{aligned}$$

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

$$\begin{aligned} q(y,c) &\leq q(\hat{y},\hat{c}) + \langle -h(\hat{x}), y-\hat{y}\rangle + (c-\hat{c})\ \|h(\hat{x})\| \\ &= q(\hat{y},\hat{c}) + \langle (-h(\hat{x}), \|h(\hat{x})\|), (y,c) - (\hat{y},\hat{c})\rangle. \end{aligned}$$

That is, $(-h(\hat{x}), \|h(\hat{x})\|) \in \partial q(\hat{y},\hat{c})$.

b) Since $\{\alpha_k\}$ is bounded, the equivalence follows from the expressions:

$$\|y_{k+1} - y_0\| \leq \sum_{j=0}^{k} \|y_{j+1} - y_j\| = \sum_{j=0}^{k} s_j\|h(x_j)\|, \qquad (3.2)$$

$$c_{k+1} - c_0 = \sum_{j=0}^{k} c_{j+1} - c_j = \sum_{j=0}^{k} (\alpha_j + 1)s_j\|h(x_j)\|. \qquad (3.3)$$

c) If MSG-1 stops at iteration k, then $h(x_k) = 0$. Therefore by Theorem 2.1 we have that

$$M_D = M_P \leq f(x_k) = f(x_k) - \langle y_k, 0\rangle + c_k\|0\| = q(y_k, c_k) \leq M_P,$$

which implies that $M_D = q(y_k, c_k)$, and $f(x_k) = M_P$. That is to say, x_k is an optimal primal solution, and (y_k, c_k) is an optimal dual solution. The theorem is proved. □

The following theorem describes an advantage of MSG-1 over the classical subgradient method, namely the monotonic improvement of the dual values.

Theorem 3.2 *Let $\{(y_k, c_k)\}$ be the sequence generated by* MSG-1*. If (y_k, c_k) is not a dual solution, then $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$.*

Proof. The proof of this result is given in [5, Theorem 7]. Even though the assumptions in [5] include continuity of f, only lower semicontinuity is required in the proof. □

Remark 3.1 We mention that the results in Theorems 3.1, 3.2 do not depend on the choice of the stepsize s_k.

3.1 Convergence results

¿From now on we assume that $h(x_k) \neq 0$ for all k, which means that the algorithm produces an infinite primal-dual sequence.

The next result provides an estimate which is essential for proving our main result. We will use in the sequel the following notation: $q_k := q(y_k, c_k)$, $\bar{q} := M_D$.

Lemma 3.1 *The following estimate is satisfied for all $k \geq 1$,*

$$\max\{q_0 + (\sum_{j=0}^{k-1} \alpha_j s_j \|h(x_j)\|)\|h(x_k)\|, f(x_k) - \langle y_0, h(x_k)\rangle\} \leq q_k. \tag{3.4}$$

Proof. It is easy to see that $y_k = y_0 - \sum_{j=0}^{k-1} s_j h(x_j)$. Therefore we have

$$\langle y_k, h(x_k)\rangle = \langle y_0, h(x_k)\rangle - \sum_{j=0}^{k-1} s_j \langle h(x_j), h(x_k)\rangle \leq \langle y_0, h(x_k)\rangle + \sum_{j=0}^{k-1} s_j \|h(x_j)\|\|h(x_k)\|,$$

using Cauchy-Schwarz inequality. Hence

$$q_k = f(x_k) - \langle y_k, h(x_k)\rangle + c_k\|h(x_k)\| \geq f(x_k) - \langle y_0, h(x_k)\rangle + (c_k - \sum_{j=0}^{k-1} s_j \|h(x_j)\|)\|h(x_k)\|. \tag{3.5}$$

On the other hand, a simple manipulation in (3.3) gives

$$c_k - \sum_{j=0}^{k-1} s_j \|h(x_j)\| = c_0 + \sum_{j=0}^{k-1} \alpha_j s_j \|h(x_j)\|. \tag{3.6}$$

Using (3.6) in (3.5) we obtain

$$\begin{aligned} q_k &\geq f(x_k) - \langle y_0, h(x_k)\rangle + c_0\|h(x_k)\| + \sum_{j=0}^{k-1} \alpha_j s_j \|h(x_j)\|\|h(x_k)\| \\ &\geq q_0 + \sum_{j=0}^{k-1} \alpha_j s_j \|h(x_j)\|\|h(x_k)\|. \end{aligned}$$

The result follows easily from these two last inequalities. □

Lemma 3.2 *Consider the sequence $\{(y_k, c_k)\}$ generated by* MSG-1. *If $\{c_k\}$ is bounded then $\{y_k\}$ is bounded. If the dual optimal set is nonempty, the converse of the last statement also holds.*

Proof. The first statement follows directly from (3.2) and (3.3). For proving the second statement, suppose that $\{y_k\}$ is bounded and take a dual solution $(\bar{y}, \bar{c})$. The supergradient inequality yields

$$q(\bar{y}, \bar{c}) \leq q(y_k, c_k) - \langle h(x_k), \bar{y} - y_k \rangle + (\bar{c} - c_k)\|h(x_k)\|,$$

and therefore

$$c_k \leq \frac{q(y_k, c_k) - q(\bar{y}, \bar{c})}{\|h(x_k)\|} + \|y_k - \bar{y}\| + \bar{c},$$

using Cauchy Schwarz inequality. Since $(\bar{y}, \bar{c})$ is a dual solution we get that

$$c_k \leq \|y_k - \bar{y}\| + \bar{c}.$$

Therefore, since $\{y_k\}$ is bounded, we conclude that $\{c_k\}$ is bounded. The result follows. □

Next, we establish convergence for a stepsize which is more general than the one used in Step 2 of MSG-1. Indeed, we prove convergence for $s_k \in [\eta_k, \bar{\beta}_k]$, where $\bar{\beta}_k \geq \beta_k$. More precisely, we make the following assumption.

$(\mathbf{A_1})$: There exist $\hat{k} > 0$ such that

$$\eta_k \leq s_k \leq \beta_k + \frac{2(\bar{q} - q_k)}{\|h(x_k)\|^2} =: \bar{\beta}_k \quad \text{for all } k > \hat{k}.$$

Remark 3.2 At least from the theoretical point of view, the step (A_1) is an improvement over the stepsizes used in [5, 1]. Indeed, the step (A_1) ensures primal and dual convergence, while in [1, 5] only dual convergence results hold, and primal convergence is proved only for an auxiliary primal sequence in [1]. In fact in [1, Example 1], the authors presented a problem for which the MSG-1 with their stepsize selection rule produces a primal sequence which does not converge to a primal solution. We will see later on that the step (A_1) forces the primal sequence to converge to a feasible point (see Theorem 3.3). The dual optimal value considered in Assumption (A_1) is just for enlarging the interval where the stepsizes can be chosen. It is clear that the interval $[\eta_k, \beta_k]$ considered at iteration k in Step 2 of MSG-1, is contained in the interval $[\eta_k, \bar{\beta}_k]$, and s_k can be chosen in $[\eta_k, \beta_k]$ without the knowledge of the dual optimal value.

¿From now on we take $\epsilon_k = \alpha_k s_k$ for all k, where $\{\alpha_k\} \subset (0, \alpha)$ for some $\alpha > 0$. Next we establish our main convergence results.

Theorem 3.3 *If the dual optimal set is nonempty then the following statements hold.*

i) Algorithm MSG-1 *generates a bounded dual sequence.*

ii) $\{h(x_k)\}$ *converges to zero and* $\{q_k\}$ *converges to* $\bar{q}$.

iii) All accumulation points of $\{(y_k, c_k)\}$ *are dual solutions.*

iv) All accumulation points of $\{x_k\}$ *are primal solutions.*

Proof. For proving (i), note that if $\{c_k\}$ is bounded then $\{(y_k, c_k)\}$ is bounded by Lemma 3.2. Thus it suffices to prove that $\{c_k\}$ is bounded. Suppose, for the sake of contradiction, that $\{c_k\}$ is unbounded. By monotonicity of $\{c_k\}$ we have that

$$\lim_{k\to\infty} c_k = \infty. \tag{3.7}$$

Observe that by continuity of h and compactness of K, we have that $\sup_k \|h(x_k)\| := b < \infty$, in particular $\{\beta_k\}$ is bounded. Consider $\hat{\beta}$ such that $\beta_k \leq \hat{\beta}$ for all k. Take $\hat{k}$ as in Assumption (A_1). In view of (3.7), there exists $k_0 > \hat{k}$ such that $c_k \geq \bar{c} + \frac{\hat{\beta} b}{2}$, for all $k \geq k_0$. Take $(\bar{y}, \bar{c}) \in D_*$. For all $k \geq k_0$ we can write

$$\begin{aligned} \|\bar{y} - y_{k+1}\|^2 &= \|\bar{y} - (y_k - s_k h(x_k))\|^2 \\ &= \|\bar{y} - y_k\|^2 + s_k^2 \|h(x_k)\|^2 + 2 s_k \langle \bar{y} - y_k, h(x_k) \rangle \\ &\leq \|\bar{y} - y_k\|^2 + s_k^2 \|h(x_k)\|^2 + 2 s_k \left[q_k - \bar{q} + \|h(x_k)\| (\bar{c} - c_k) \right], \end{aligned} \tag{3.8}$$

using the update of the dual variables in the first equality, and the supergradient inequality in the inequality. Rearranging the right-hand side of the expression above, and using Assumption (A_1), we obtain

$$\begin{aligned} \|\bar{y} - y_{k+1}\|^2 &\leq \|\bar{y} - y_k\|^2 + s_k [s_k \|h(x_k)\|^2 + 2(q_k - \bar{q})] + 2 s_k \|h(x_k)\| (\bar{c} - c_k) \\ &\leq \|\bar{y} - y_k\|^2 + s_k \beta_k \|h(x_k)\|^2 + 2 s_k \|h(x_k)\| (\bar{c} - c_k) \\ &= \|\bar{y} - y_k\|^2 + s_k \|h(x_k)\| (\beta_k \|h(x_k)\| + 2\bar{c} - 2c_k) \\ &\leq \|\bar{y} - y_k\|^2 + s_k \|h(x_k)\| (\hat{\beta} b + 2\bar{c} - 2c_k), \end{aligned} \tag{3.9}$$

using the definition of b in the last inequality. Therefore,

$$\|\bar{y} - y_{k+1}\|^2 \leq \|\bar{y} - y_k\|^2 + s_k \|h(x_k)\| (\hat{\beta} b + 2\bar{c} - 2c_k). \tag{3.10}$$

The expression between parentheses in (3.10) is negative by definition of k_0. Hence, we obtain

$$\|y_{k+1} - \bar{y}\| \leq \|y_k - \bar{y}\| \leq \|y_{k_0} - \bar{y}\|, \quad \text{for all } \; k \geq k_0. \tag{3.11}$$

Thus, $\{y_k\}$ is bounded, and by Lemma 3.2 we conclude that $\{(y_k, c_k)\}$ is bounded, in contradiction with (3.7), and hence (i) holds. Moreover, we have that $\sum_k s_k \|h(x_k)\| < \infty$, by Theorem 3.1$(b)$. In particular $\{s_k \|h(x_k)\|\}$ goes to zero. On the other hand, by the first inequality in Assumption (A_1) we have that

$$s_k \|h(x_k)\| \geq \eta_k \|h(x_k)\| > 0 \quad \text{for all } k \geq \hat{k}, \tag{3.12}$$

where $\eta_k = \min\{\eta, \|h(x_k)\|\}$. Hence we obtain from (3.12) that $\{h(x_k)\}$ converges to zero. We are going to prove (ii) and (iv) simultaneously. Since $\{x_k\} \subset K$ and K is compact, $\{x_k\}$ is bounded. Take an accumulation point $\bar{x}$ of $\{x_k\}$. Suppose that $\{x_{k_j}\}$ converges to $\bar{x}$. By lower semi-continuity of f and Lemma 3.1, we obtain

$$f(\bar{x}) \leq \liminf_j \left(f(x_{k_j}) - \langle y_0, h(x_{k_j}) \rangle \right) \leq \liminf_j q_{k_j} \leq \bar{q} = M_P, \tag{3.13}$$

using also the fact that $\{h(x_{k_j})\}$ converges to zero. On the other hand, by continuity of h we have that $h(\bar{x}) = 0$. Therefore, we conclude from (3.13) that $\bar{x}$ is a primal solution. In particular, all inequalities in (3.13) are equalities and then $\liminf q_{k_j} = \bar{q}$. Since $\{q_k\}$ is increasing by Theorem 3.2, we get that $\{q_k\}$ converges to $\bar{q}$, and we have thus proved (ii) and (iv). For proving (iii), take a subsequence $\{(y_{k_j}, c_{k_j})\}_j$ converging for some $(\hat{y}, \hat{c})$. By upper-semicontinuity of q (Proposition 2.1(i)), we get

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

using the fact that $\{q_k\}$ converges to $\bar{q}$ by (ii). Hence we have that $(\hat{y}, \hat{c})$ is a dual solution. This proves (iii), and the theorem follows. □

Theorem 3.3 presents convergence results for the primal and dual sequences generated by Algorithm MSG-1 assuming the existence of an optimal dual solution. The next theorem ensures convergence of the primal sequence generated by MSG-1 even when the dual solution set is empty. This is very important, because in general, it is not possible to know "a priori" whether the dual solution set is nomempty. Also, in our dual formulation, which includes the penalty parameter c as a dual variable, optimal dual solutions exist only when the problem admits exact penalization (cf. Remark 4.1 in Section 4), and many problems of interest fail to enjoy this property.

Theorem 3.4 *Assume that $\epsilon_k = \alpha_k s_k$ and $\hat{\alpha} := \inf_k \alpha_k > 0$. Then $\{h(x_k)\}$ converges to zero and $\{q_k\}$ converges to $\bar{q}$. Moreover, all accumulation points of the primal sequence $\{x_k\}$ are primal solutions.*

Proof. By monotonicity of the sequence $\{c_k\}$, either it goes to infinite, or it converges to some $\hat{c}$. In the second case, we have that $\{c_k\}$ is bounded, therefore by Lemma 3.2 we get that $\{(y_k, c_k)\}$ is also bounded. Hence repeating the proof of Theorem 3.3 $(ii), (iii)$ and (iv) we get that the dual solution set is nonempty (observe that in Theorem 3.3 we use the nonemptiness of the dual solution set just for ensuring the boundedness of the dual sequence). Thus, in this case (i.e., when $\{c_k\}$ is bounded) the theorem is proved. So we just need to consider the case in which $\{c_k\}$ goes to infinite. In this case, by Theorem 3.1(b), $\sum_j s_j \|h(x_j)\| = \infty$. On the other hand, by Lemma 3.1 we obtain that

$$\hat{\alpha} \left(\sum_{j=0}^{k-1} s_j \|h(x_j)\| \right) \|h(x_k)\| \leq \left(\sum_{j=0}^{k-1} \alpha_j s_j \|h(x_j)\| \right) \|h(x_k)\| \leq q_k - q_0 \leq \bar{q} - q_0. \quad (3.14)$$

Note that $\bar{q} < \infty$ and $\sum_j s_j \|h(x_j)\| = \infty$. Therefore we conclude that $\{h(x_k)\}$ converges to zero. The proof of the remaining statements follows the same steps as in (ii) and (iv) of Theorem 3.3. □

Remark 3.3 Theorem 3.4 ensures that MSG-1 generates a primal sequence converging to the primal solution set, in the sense that all its accumulation points are primal solutions. Note that our results hold without differentiability or convexity assumptions.

The following corollary establishes the equivalence between the boundedness of the dual sequence and the existence of dual solutions. A similar result was obtained in [1, 3].

Corollary 3.1 *Consider a dual sequence $\{(y_k, c_k)\}$ generated by Algorithm* MSG-1. *This sequence is bounded if and only if the dual optimal set is nonempty.*

Proof. If the dual optimal set is nonempty, then Theorem 3.3 (i) ensures that $\{(y_k, c_k)\}$ is bounded. For proving the converse statement, we just note that in the proof of Theorem 3.3, we only use the existence of a dual solution for ensuring boundedness of the dual sequence. Thus, if we assume that the dual sequence is bounded, we can repeat the proof of Theorem 3.3(*ii*) and (*iii*), and prove that the dual optimal set is nonempty. The result follows. □

In Theorem 3.3 we proved that all accumulation points of the dual sequence $\{(y_k, c_k)\}$ generated by MSG-1 are optimal solutions. Since the dual problem is convex and we are applying a subgradient method, we should expect convergence of the whole sequence. The next proposition establishes this result.

Proposition 3.1 *Consider the dual sequence $\{(y_k, c_k)\}$ generated by* MSG-1, *and assume that $\epsilon_k = \alpha_k s_k$ and $0 < \alpha_k < \alpha < \infty$. If D_* is nonempty, then $\{(y_k, c_k)\}$ converges to a dual solution.*

Proof. Since D_* is nonempty, it follows from Theorem 3.3 that $\{(y_k, c_k)\}$ is bounded. In particular $\{y_k\}$ and $\{c_k\}$ are bounded. Take an accumulation point $(\bar{y}, \bar{c})$ of $\{(y_k, c_k)\}$. It follows that $(\bar{y}, \bar{c})$ belongs to D_* (Theorem 3.3(*iii*)). Since $\{c_k\}$ is increasing and bounded, it converges to $\bar{c}$. Therefore we just need to prove that $\{y_k\}$ converges to $\bar{y}$. Consider a subsequence $\{y_{k_j}\}_j$ converging to $\bar{y}$. Using the same calculations as in (3.10), we obtain

$$\|\bar{y} - y_{k+1}\|^2 \leq \|\bar{y} - y_k\|^2 + \tilde{\beta} s_k \|h(x_k)\|, \tag{3.15}$$

where $\tilde{\beta} := \hat{\beta} b + 2\bar{c}$, with b and $\hat{\beta}$ as in the proof of Theorem 3.3. On the other hand, since $\{(y_k, c_k)\}$ is bounded, we have that $\sum s_k \|h(x_k)\| < \infty$, by Theorem 3.1. Therefore, given an arbitrary $\epsilon > 0$, there exists k_0 sufficiently large such that

$$\sum_{k > k_0} s_k \|h(x_k)\| < \frac{\epsilon}{2\tilde{\beta}}.$$

Since $\{y_{k_j}\}_j$ converges to $\bar{y}$, there exists j_0 such that $k_{j_0} > k_0$ and $\|y_{k_j} - \bar{y}\|^2 < \frac{\epsilon}{2}$ for all $j \geq j_0$. Using (3.15) we obtain, for all $k > k_{j_0}$,

$$\|\bar{y} - y_k\|^2 \leq \|\bar{y} - y_{k_{j_0}}\|^2 + \tilde{\beta} \sum_{l = k_{j_0}}^{k-1} s_l \|h(x_l)\| < \epsilon. \tag{3.16}$$

Since ϵ is arbitrary, we conclude that $\{y_k\}$ converges to $\bar{y}$. Therefore $\{(y_k, c_k)\}$ converges to $(\bar{y}, \bar{c})$, and the proposition follows. □

4 Algorithm 2

In this section we present and analyze algorithm MSG-2. This algorithm has a stepsize which ensures finite termination, as long as there exist dual solutions.

Step 0. Choose $(y_0, c_0) \in \mathbb{R}^m \times \mathbb{R}_{++}$, and exogenous parameters $\{\epsilon_k\} \subset \mathbb{R}_{++}$. Also fix $\bar{\beta} \geq \bar{\eta} > 0$. Set $k := 0$.

Step 1 **(Subproblem and Stopping Criterion)**

a) Find $x_k \in A(y_k, c_k)$,

b) if $h(x_k) = 0$ stop,

c) if $h(x_k) \neq 0$, go to **Step** 2.

Step 2 **(Step-size Choice and Update of Dual Variables)**

$\eta_k := \frac{\bar{\eta}}{\|h(x_k)\|}$, $\beta_k := \frac{\bar{\beta}}{\|h(x_k)\|}$, and choose $s_k \in [\eta_k, \beta_k]$,

$y_{k+1} := y_k - s_k h(x_k)$,

$c_{k+1} := c_k + (\epsilon_k + s_k)\|h(x_k)\|$.

Set $k = k + 1$ and go to **Step** 1.

Observe that the only difference between MSG-1 and MSG-2 is the choice of β_k and η_k.

Theorem 4.1 *Suppose that the dual solution set is nonempty. Consider the* MSG-2 *algorithm. Choose $\epsilon_k = \alpha_k s_k$, with $\{\alpha_k\} \subset (0, \alpha)$ for some $\alpha > 0$. Then there exists $\bar{k} > 0$ such that $h(x_{\bar{k}}) = 0$, i.e.* MSG-2 *stops at the $\bar{k}$-th iteration. In particular $x_{\bar{k}}$ and $(y_{\bar{k}}, c_{\bar{k}})$ are primal and dual optimal solutions respectively.*

Proof. We prove first that the dual sequence is bounded. If $c_k \leq \frac{\bar{\beta}}{2} + \bar{c}$ for all k, then $\{y_k\}$ is also bounded by Lemma 3.2. Assuming, for the sake of contradiction, that $\{c_k\}$ is unbounded, we can repeat the same calculations as in (3.8), (observing that $q_k \leq \bar{q}$), to get

$$\begin{aligned} \|\bar{y} - y_{k+1}\|^2 &\leq \|\bar{y} - y_k\|^2 + s_k\|h(x_k)\|(\beta_k\|h(x_k)\| + 2\bar{c} - 2c_k) \\ &= \|\bar{y} - y_k\|^2 + s_k\|h(x_k)\|(\bar{\beta} + 2\bar{c} - 2c_k). \end{aligned} \tag{4.1}$$

Since $\{c_k\}$ is increasing, there exists k_0 such that $c_k > \frac{\bar{\beta}}{2} + \bar{c}$ for all $k > k_0$. Using this estimate in (4.1), we obtain, for all $k > k_0$,

$$\|\bar{y} - y_{k+1}\|^2 \leq \|\bar{y} - y_k\|^2. \tag{4.2}$$

¿From (4.2) we obtain that $\{y_k\}$ is bounded. Thus, $\{(y_k, c_k)\}$ is bounded by Lemma 3.2, contradicting the supposed unboundedness of $\{c_k\}$. Hence, the dual sequence is bounded. Let us prove now that the algorithm has finite termination. If this is not true, we must have $h(x_k) \neq 0$ for all k (note that the algorithm stops at k if and only if $h(x_k) = 0$).

Using the fact that the sequence $\{c_k\}$ is bounded and the definition of the stepsize, we can write

$$\infty > \lim_{k\to\infty} c_k - c_0 = \sum_{k=0}^{\infty}(c_{k+1} - c_k) = \sum_{k=0}^{\infty}(\epsilon_k + s_k)\|h(x_k)\| \geq \sum_{k=0}^{\infty} s_k\|h(x_k)\| \geq \sum_{k=0}^{\infty} \bar{\eta} = \infty,$$

which entails a contradiction. Thus there exists $\bar{k}$ such that $h(x_{\bar{k}}) = 0$. In view of Remark 3.1, the result follows by Theorem 3.1(c). □

Remark 4.1 A finitely convergent algorithm for nonsmooth and nonconvex problems might seem too good to be true, but the point here is that the assumption of existence of optimal dual solutions is stronger than it looks at first sight. Observe that we have included the penalty parameter c among the dual variables, and hence the existence of optimal dual solutions implies in particular the existence of an optimal penalty parameter c^*. It is easy to verify that any c larger than such a c^* turns out to be an exact penalty parameter, in the sense of [11, Section 11]. Thus, in our formulation, if optimal dual solutions exist then the problem admits exact penalization. In such a setting, for achieving finite convergence it is enough to have a stepsize selection rule wich allows c_k to attain arbitrarily large values. In fact, after establishing that the sequence $\{y_k\}$ is bounded, as is the case for both Algorithm 1 and 2, Proposition 2.2 provides an alternative argument for the finite convergence of Algorithm 2, assuming existence of a dual solution (y^*, c^*): if ρ is such that $\|y_k - y^*\| \leq \rho$ for all k, then any pair (y, c) with $\|y - y^*\| \leq \rho$, $c \geq c^* + \rho$ belongs to D_* by Proposition 2.2, and hence we get $(y_k, c_k) \in D_*$ as soon as $c_k > c^* + \rho$. Once such a value of k is reached, x_k will be an optimal primal solution, because, as commented above, the fact that x_k belongs to $A(y_k, c_k)$, as prescribed in Step 1(a) of Algorithm 2, is equivalent to saying that x_k is an exact minimizer of $L(\cdot, y_k, c_k)$ on K. It should be emphasized, however, that attempting to circunvent the dual updating by guessing the "right" values of c^* and ρ (assuming that it is known in advance that the problem admits exact penalization), does not seem to be in general a good strategy: quite likely one will overshoot the value of the parameters, and then suffer the consequences, in terms of numerical instability, of a too large penalty parameter (of course, this comment applies to any penalty method in the presence of exact penalization; not just to ours). A sensible gradual increase of the penalty parameter, like the updating of c_k in Algorithm 2, is likely to give rise to a better numerical behavior. See also the discussion in [2] on the comparison of the actual numerical behavior of a dual updating similar to ours with a classical penalty method.

References

[1] 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. *Journal of Global Optimization*, 34(1):55–78, 2006.

[2] R. S. Burachik and C. Y. Kaya. An update rule and a convergence result for a penalty function method. *Journal of Industrial and Management Optimization*, 3(2):381–398, 2007.

[3] R. S. Burachik, C. Y. Kaya and M. Mammadov. An inexact modified subgradient algorithm for nonconvex optimization. *Computational Optimization and Applications*, Published Online on 29/01/2008, DOI 10.1007/s10589-008-9168-7.

[4] R. S. Burachik and A. M. Rubinov. Abstract convexity and Augmented Lagrangians. *SIAM Journal on Optimization*, 18: 413–436, 2007

[5] R. N. Gasimov. Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming. *Journal of Global Optimization*, 24(2):187–203, 2002.

[6] R. N. Gasimov and A. M. Rubinov. On augmented Lagrangians for optimization problems with a single constraint. *Journal of Global Optimization*, 28(2):153–173, 2004.

[7] X. X. Huang and X. Q. Yang. A unified augmented Lagrangian approach to duality and exact penalization. *Mathematics of Operations Research*, 28(3):533–552, 2003.

[8] X. X. Huang and X. Q. Yang. Further study on augmented Lagrangian duality theory. *Journal of Global Optimation*, 31(2):193–210, 2005.

[9] A. Nedić, A. Ozdaglar and A. Rubinov. Abstract convexity for nonconvex optimization duality. *Optimization*, 56(5-6):655–674, 2007.

[10] Jean-Paul Penot. Augmented Lagrangians, duality and growth conditions. *Journal of Nonlinear and Convex Analysis*, 3(3):283–302, 2002.

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

[12] A. M. Rubinov and X. Q. Yang. *Lagrange-type Functions in Constrained Non-convex Optimization*, Kluwer, Boston, 2003.

[13] A. M. Rubinov, X. X. Huang and X. Q. Yang. The zero duality gap property and lower semicontinuity of the perturbation function. *Mathematics of Operations Research*, 27(4):775–791, 2002.