Journal of Global Optimization (2006) 34: 55–78
DOI 10.1007/s10898-005-3270-5

On a Modified Subgradient Algorithm for Dual Problems via Sharp Augmented Lagrangian*

REGINA S. BURACHIK[1,†], RAFAIL N. GASIMOV[3], NERGIZ A. ISMAYILOVA[2] and C. YALÇIN KAYA[4]
[1]*Engenharia de Sistemas e Computação, COPPE-UFRJ, CP 68511, Rio de Janeiro-RJ, CEP 21941-972, Brazil (e-mail: regi@cos.ufrj.br)*
[2]*Department of Industrial Engineering, Osmangazi University, Bademlik 26030, Eskişehir, Turkey (e-mail: n.ismail@ogu.edu.tr)*
[3]*Department of Industrial Engineering, Osmangazi University, Bademlik 26030, Eskişehir, Turkey (e-mail: gasimovr@ogu.edu.tr)*
[4]*School of Mathematics and Statistics, University of South Australia, Mawson Lakes, S.A. 5095, Australia (e-mail: yalcin.kaya@unisa.edu.au)*

(Received 10 October 2003; accepted 1 March 2005)

Abstract. We study convergence properties of a modified subgradient algorithm, applied to the dual problem defined by the sharp augmented Lagrangian. The primal problem we consider is nonconvex and nondifferentiable, with equality constraints. We obtain primal and dual convergence results, as well as a condition for existence of a dual solution. Using a practical selection of the step-size parameters, we demonstrate the algorithm and its advantages on test problems, including an integer programming and an optimal control problem.

Mathematics Subject Classifications. 90C26, 49M29, 49M37, 90C90

Key words: Augmented Lagrangian, Nonconvex programming, Nonsmooth optimization, Sharp Lagrangian, Subgradient optimization

1. Introduction

Classical Lagrange and penalty functions and algorithms based on the corresponding duality framework can only be applied to some special classes of constrained optimization problems. This justifies the quest for other kinds of augmented Lagrangians, which are able to provide solution algorithms for a broader family of constrained optimization problems. The works [23–26, 31] study new kinds of Lagrangians, and their applications to different classes of constrained optimization problems. Specific applications of some of these new Lagrangians can be found in [8, 9]. The development of efficient methods for solving optimization problems depends on

*Partially Supported by 2003 UniSA ITEE Small Research Grant ER02.
†Supported by CAPES, Brazil, Grant No. 0664-02/2, during her visit to the School of Mathematics and Statistics, UniSA.

the duality relations provided by the augmented Lagrangian framework. The dual problem is typically a nonsmooth convex problem, and hence nonsmooth minimization techniques can be used for solving it. A typical example of these techniques is the subgradient method and its extensions. Subgradient methods were introduced in the middle 1960s, in the works of Demyanov [6], Poljak [16–18] and Shor [28, 29]. A detailed presentation of these methods can be found in [3, 5, 13] and the references therein.

In this paper we study nonconvex optimization problems with equality constraints, where the cost and constraints are only required to be continuous; namely we do not pose any differentiability conditions. Note that inequality constraints can be reformulated as (nonsmooth) equality constraints, without introducing extra (slack) variables (see the test problems in Section 5.2). In order to devise an efficient duality framework for solving nonconvex, equality constrained optimization problems, Gasimov proposed in [8] a *modified subgradient* (MSG) *algorithm*. The MSG algorithm solves the dual problem obtained with respect to the sharp Lagrangian. Note that sharp Lagrangians were introduced by Rockafellar and Wets in [20, Example 11.58]. This concept was also studied in [1, 2], where it was used for establishing conditions for the existence of solutions and zero duality gap properties for some classes of nonconvex optimization problems with inequality constraints.

One may consider versions of the MSG algorithm incorporating other forms of augmented Lagrangian functions (see [20, Chapter 11], where the authors study a family of augmented Lagrangians including the sharp as well as the classical quadratic augmented Lagrangian as special cases). In the present paper we focus only on the MSG using the sharp Lagrangian, as an extension of the approach taken by Gasimov [8]. A study of the convergence properties of the MSG algorithm using other augmented Lagrangians will be a focus of future study.

Our aim is to give a new convergence analysis for the MSG algorithm and propose new formulas for the step-size parameters. A theoretical and practical advantage of the MSG algorithm is that it generates a strictly increasing sequence of dual values. We point out that such a property is not possessed by the classical subgradient method. This important property, together with the new formulas for the step-size parameters, allows us to obtain additional convergence results. Namely, we establish convergence of the sequences of dual values and dual variables, and optimality of all accumulation points of a primal sequence related to the algorithm. To our knowledge, analogous convergence results do not exist for subgradient methods. Our analysis also allows us to obtain a condition (associated with the dual sequence generated by the MSG algorithm) which implies existence of dual solutions.

Our convergence and existence results are proved first for a general, and then for a specific, choice of the step-size parameters. For the specific choice, we are able to prove that boundedness of the sequence is equivalent to the existence of solutions. For a numerical implementation of the MSG algorithm we select practical step-size parameters. We demonstrate the advantages of the new step-sizes in some example applications, including optimal control and integer programming problems.

The paper is organized as follows. In Section 2, we review the theoretical background concerning the sharp Lagrangian duality framework. In Section 3, we state the MSG algorithm and provide motivations for the improvement of the convergence results given in [8]. In Section 4, we give our convergence and existence results. In Section 5, we define practical step-sizes for a numerical implementation, and demonstrate their use on test problems.

2. Preliminaries

We consider the nonlinear programming problem:

$$(P) \text{ minimize } f_0(x) \text{ over all } x \text{ in } X \text{ satisfying } f(x)=0,$$

where X is a compact subset of $\mathbb{R}^n$, and the functions $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $f : \mathbb{R}^n \to \mathbb{R}^m$ are continuous. Denote by $\mathbb{R}_+$, $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ the set of nonnegative numbers, the Euclidean norm and the Euclidean inner product on $\mathbb{R}^m$, respectively. The *sharp augmented Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}$ associated with (P) is defined as

$$L(x,u,c) := f_0(x) + c\|f(x)\| - \langle u, f(x)\rangle, \tag{1}$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ and $c \in \mathbb{R}_+$. The solution set of problem (P) is denoted by $S(P)$. We typically denote an element of $S(P)$ by $\bar{x}$. The dual function $H : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}$ is defined as

$$H(u,c) = \min_{x\in X}[f_0(x) + c\|f(x)\| - \langle u, f(x)\rangle]. \tag{2}$$

Then the dual problem of (P) is given by

$$(P^*) : \max_{(u,c)\in R^m \times R_+} H(u,c).$$

The solution set of problem (P^*) is denoted by $S(P^*)$. We typically denote an element in $S(P^*)$ by $\bar{z} = (\bar{u}, \bar{c})$. For convenience, we introduce the set

$$X(u,c) = \operatorname*{Argmin}_{x\in X}[f_0(x) + c\|f(x)\| - \langle u, f(x)\rangle]. \tag{3}$$

We assume in this work that we are able to solve the minimization problem given above. Let $d(z, w) := \|w - z\|^2$ for $w, z \in \mathbb{R}^m$.

Zero duality gap and saddle point properties are crucial in every duality framework. Theorems 1–3, quoted below, state that these properties hold for the sharp Lagrangian duality scheme, when applied to problem (P). Theorems 1 and 2 are proved (in a more general setting) in [20, Theorem 11.59], (see also [2]). [Note that Theorems 1–3 have been proved in [21] for the classical quadratic augmented Lagrangian. The proof of Theorem 3 is a straightforward modification of the one given in [21, Theorem 6.4].

THEOREM 1. *Suppose that* $\inf P$ *is finite and consider the augmented Lagrangian L given in* (1). *A pair of elements* $\bar{x} \in X$ *and* $(\bar{u}, \bar{c}) \in \mathbb{R}^m \times \mathbb{R}_+$ *furnishes a saddle point of L on* $X \times (\mathbb{R}^m \times \mathbb{R}_+)$ *if and only if* $\bar{x} \in S(P)$, $(\bar{u}, \bar{c}) \in S(P^*)$ *and* $\inf P = \sup P^*$.

THEOREM 2. *A pair of vectors* $\bar{x} \in X$ *and* $(\bar{u}, \bar{c}) \in \mathbb{R}^m \times \mathbb{R}_+$ *furnishes a saddle point of the augmented Lagrangian L on* $X \times (\mathbb{R}^m \times \mathbb{R}_+)$ *if and only if*

$$\bar{x} \in S(P) \quad \text{and} \quad h(v) \geqslant h(0) + \langle v, \bar{u} \rangle - \bar{c}\|v\| \quad \text{for all } v \in \mathbb{R}^m, \tag{4}$$

where $h(v) := \inf\{f_0(x) \mid f(x) \leqslant v, x \in X\}$ *is the* perturbation function *associated with problem* (P). *When this holds, any* $\hat{c} > \bar{c}$ *will have the property that*

$$\hat{x} \in S(P) \quad \text{if and only if } \hat{x} \in X(\bar{u}, \hat{c}).$$

REMARK 1. If $(\bar{u}, \bar{c})$ is a solution of the dual problem, then the pair $(\bar{u}, \hat{c})$ is also a solution for each $\hat{c} > \bar{c}$.

THEOREM 3. *Suppose that* f_0 *and f are continuous, X is compact, and a feasible solution of* (P) *exists. Then* $\inf P = \sup P^*$ *and* $S(P) \neq \emptyset$. *Furthermore, the dual function H given in* (2) *is concave and finite everywhere on* $\mathbb{R}^m \times \mathbb{R}_+$. *Consequently, this maximization problem is effectively unconstrained.*

The following theorem was proved in [8] and will be used for defining a stopping criteria for the MSG algorithm.

THEOREM 4. *Let* $\inf P = \sup P^*$ *and suppose that for some* $(\bar{u}, \bar{c}) \in \mathbb{R}^m \times \mathbb{R}_+$, *and* $\bar{x} \in X$,

$$\min_{x \in X} L(x, \bar{u}, \bar{c}) = f_0(\bar{x}) + \bar{c}\|f(\bar{x})\| - \langle \bar{u}, f(\bar{x}) \rangle. \tag{5}$$

Then $\bar{x}$ is a solution to (P) and $(\bar{u}, \bar{c})$ is a solution to (P^*) if and only if

$$f(\bar{x}) = 0. \tag{6}$$

3. The MSG Algorithm and Motivation

The MSG algorithm is devised for solving the dual problem (P^*), which is described as follows:

$$(P^*) \quad \max_{(u,c)\in \mathbf{R}^m \times \mathbf{R}_+} \min_{x\in X} [f_0(x) + c\|f(x)\| - \langle u, f(x)\rangle].$$

Let us outline the MSG algorithm.
Step 0. Choose (u_0, c_0) with $c_0 \geqslant 0$. Set $k = 0$.
Step k. Given (u_k, c_k):
Step k.1. Solve the following subproblem:

$$\min_{x\in X} [f_0(x) + c_k\|f(x)\| - \langle u_k, f(x)\rangle]. \tag{7}$$

Let x_k be a solution, i.e., $x_k \in X(u_k, c_k)$. If $f(x_k) = 0$, then STOP; by Theorem 4 (u_k, c_k) is a solution of (P^*) and x_k is a solution of (P).
Step k.2. Set

$$u_{k+1} := u_k - s_k f(x_k), \tag{8}$$
$$c_{k+1} := c_k + (s_k + \varepsilon_k)\|f(x_k)\|, \tag{9}$$

where $s_k, \varepsilon_k > 0$. Set $k = k+1$ and repeat Step k.

Given a dual iterate (u_k, c_k), and $x_k \in X(u_k, c_k)$, we introduce the following notation for brevity.

$$\begin{aligned}
&z_k := (u_k, c_k),\\
&x_k \in X(u_k, c_k),\\
&f_k := f(x_k),\\
&\widetilde{x}_k \in X(u_k, c_k + \beta), \quad \text{where } \beta > 0,\\
&\widetilde{f}_k := f(\tilde{x}_k),\\
&H_k := H(u_k, c_k),\\
&\overline{H} := H(\bar{u}, \bar{c}),
\end{aligned}$$

We start by quoting the convergence results given in [8].

THEOREM 5 [8]. *Let $\{(u_k, c_k)\}$ be the sequence of dual variables generated by the MSG algorithm. Assume that (u_k, c_k) is not a solution of the dual problem for any k, that is, $f_k \neq 0$ for all k. Then*

(a) $0 < H_{k+1} - H_k \leqslant (2s_k + \varepsilon_k)\|f_k\|^2$ *for all* $s_k, \varepsilon_k > 0$.

(b) *Assume that there exists a dual solution* $\bar{z} = (\bar{u}, \bar{c})$ *and let* $d_k = (\bar{z}, z_k)$. *If*

$$0 < \varepsilon_k < s_k < \frac{2(\overline{H} - H_k)}{5\|f_k\|^2},$$

then $d_{k+1} - d_k < 0$.

(c) *Assume again that that there exists a dual solution and that all conditions of Theorem 3 hold. If*

$$0 < \varepsilon_k < s_k = \frac{(\overline{H} - H_k)}{5\|f_k\|^2}, \quad \text{then } H_k \to \overline{H}. \tag{10}$$

Our aim is to improve the convergence results given in Theorem 5. Instead of requiring the step-size s_k to satisfy (10), we establish convergence to the optimal dual value when s_k is chosen in a more general way. We also prove convergence of the whole sequence of dual variables for this general choice of s_k.

On the other hand, the sequence $\{x_k\}$ generated by the MSG algorithm may not converge to a primal solution. Indeed, Example 1 furnishes a case which satisfies the hypotheses of Theorem 5, but generates a sequence $\{x_k\}$ which does not converge to a solution of the original problem. For this reason, we introduce in Section 4 an auxiliary sequence $\{\tilde{x}_k\}$ such that $\tilde{x}_k \in X(u_k, c_k + \beta)$ for all k and for some $\beta > 0$, and establish optimality of all accumulation points of this sequence.

EXAMPLE 1. Let $f_0, f : \mathbb{R} \to \mathbb{R}$ be defined as $f_0(x) := |x|$ and $f(x) := 1 - x^2$, respectively. Assume also that $X = [-1, 1]$. It is easy to see that $S(P) = \{1, -1\}$ and $\inf P = \sup P^* = 1$. Let us show that $\{x_k\}$ calculated from the MSG algorithm does not converge to a solution of the original problem. It is straightforward to verify that, if $(u_k, c_k) \in \mathbb{R} \times \mathbb{R}_+$ is such that $(c_k - u_k) < 1$, then

$$X(u_k, c_k) = \{0\}, \qquad \text{and} \qquad H_k = (c_k - u_k). \tag{11}$$

In this situation, we must have $x_k = 0$ and so $f_k = |f_k| = 1$. Using the latter fact and that $\overline{H} = 1$, the rule (10) gives

$$s_k = \frac{1 - (c_k - u_k)}{5}. \tag{12}$$

Take $\varepsilon_k = 0.5 s_k$. Now we claim that, if $(c_k - u_k) < 1$, then $(c_{k+1} - u_{k+1}) < 1$. Indeed, using the definition of the MSG algorithm, (12) and the choice of ε_k, we get

$$c_{k+1} - u_{k+1} = c_k - u_k + \tfrac{5}{2} s_k = \frac{c_k - u_k}{2} + \tfrac{1}{2} < 1.$$

As a consequence, if we start with $(u_0, c_0) \in \mathbb{R} \times \mathbb{R}_+$ such that $c_0 - u_0 < 1$, then it must hold that $(c_k - u_k) < 1$ for all k. In this case, by (11) the sequence $x_k = 0$ for all k. Since $S(P) = \{-1, 1\}$, the sequence $\{x_k\}$ does not converge to a primal solution. However, convergence of the dual values is guaranteed by Theorem 5. Indeed, the last expression implies that

$$H_k = c_k - u_k = \frac{(c_0 - u_0) - 1}{2^k} + 1. \tag{13}$$

Since $c_0 - u_0 < 1$, the sequence $\{H_k\}_k$ converges increasingly to $\overline{H} = 1$. Moreover, it can also be shown that the dual variables converge to a dual solution. In order to achieve this, first note that

$$c_{k+1} - c_0 = \sum_{j=0}^{k} (s_j + \varepsilon_j) = \tfrac{3}{2} \sum_{j=0}^{k} s_j.$$

Then using (12) and (13), and letting $k \to \infty$, it is not difficult to derive

$$\bar{c} = c_0 + \tfrac{6}{10}[1 - (c_0 - u_0)] \tag{14}$$

for some given c_0 and u_0 such that $c_0 - u_0 < 1$. Combine this fact with (13) to get $\bar{u} = \bar{c} - 1$. It can be proved that $H(\bar{u}, \bar{c}) = \bar{c} - \bar{u} = 1$ and hence $(\bar{u}, \bar{c})$ is a dual solution.

Under the basic assumptions on problem (P), Theorem 3 asserts that S(P) is nonempty and there is no duality gap. However, the set of dual solutions $S(P^*)$ may in fact be empty, as the following example shows.

EXAMPLE 2. Let $f_0, f : \mathbb{R} \to \mathbb{R}$ be defined as $f_0(x) := -|x|$ and $f(x) := (1/2)x^2$, respectively. Assume also that $X = [-1, 1]$. It is easy to see that $S(P) = \{0\}$ and $\inf P = \sup P^* = 0$. It can be checked that

$$\begin{aligned} &\operatorname*{Argmin}_{x \in [-1,1]} f_0(x) + c\|f(x)\| - uf(x) = \operatorname*{Argmin}_{x \in [-1,1]} -|x| + \frac{c-u}{2} x^2 \\ &= \begin{cases} \left\{\frac{1}{c-u}, -\frac{1}{c-u}\right\}, & \text{if } c-u > 1, \\ \{1, -1\}, & \text{if } c-u \leqslant 1. \end{cases} \end{aligned}$$

As a consequence, the dual function H is given by

$$H(u, c) = \begin{cases} \frac{-1}{2(c-u)}, & \text{if } c-u > 1, \\ -1 + \frac{c-u}{2}, & \text{if } c-u \leqslant 1. \end{cases}$$

It is clear that $\overline{H} = \sup_{(u,c) \in \mathbf{R} \times \mathbf{R}_+} H(u, c) = 0$, and that this supremum is never attained, so $S(P^*) = \emptyset$.

4. Existence and Convergence Results

Throughout this section, we assume that the hypotheses of Theorem 3 hold. Lemma 1 sets conditions which will be instrumental in the first part of our analysis, using a general choice of the step-size parameter s_k. Later on, these conditions will be relaxed, under a specific choice of s_k.

4.1. ANALYSIS WITH A GENERAL CHOICE OF s_k

LEMMA 1. *Consider the notation and definitions of the MSG algorithm. The following statements are equivalent.*

(a) $\sum_{k=0}^{\infty}(s_k+\varepsilon_k)\|f_k\|<\infty$.
(b) *The sequence* $\{z_k\}$ *is bounded.*

Proof. From the MSG algorithm

$$c_{m+1}-c_0=\sum_{k=0}^{m}(s_k+\varepsilon_k)\|f_k\| \quad \text{and} \quad \|u_{m+1}-u_0\|\leqslant\sum_{k=0}^{m}s_k\|f_k\|. \tag{15}$$

Then we readily obtain that conditions (a) and (b) are equivalent. □

Conditions which guarantee the existence of dual solutions are often related to properties of the perturbation function (see (4)) associated with the problem. Unfortunately, the calculation of the perturbation function is very difficult, and hence it makes sense to establish alternative ways of guaranteeing existence of dual solutions. One of these alternative ways was given in [1] by Azimov and Gasimov, who presented conditions which depend on the objective and constraint functions. Other conditions of this kind can be found in references [9, 22, 24–26]. We give below a new existence condition by using the dual sequence generated by the MSG algorithm using a general step-size s_k. In what follows, $\{z_k\}=\{(u_k,c_k)\}$ is the dual sequence generated by the MSG algorithm. In view of Theorem 4, we see that the MSG algorithm either stops at Step k.1, yielding a primal-dual solution; or it generates an infinite dual sequence. So in our analysis we only study the case in which the algorithm generates an infinite dual sequence. This assumption will hold until the end of this section and it is equivalent to require $f_k\neq 0$ for all k.

THEOREM 6. *Let* $\overline{H}$ *be the optimal dual value* (*i.e.* $\overline{H}=\sup P^*$). *Assume that* $\{z_k\}$ *is bounded and that the step-size* s_k *satisfies*

$$s_k\geqslant\eta\frac{(\overline{H}-H_k)}{\|f_k\|^2} \tag{16}$$

for some fixed $\eta > 0$. *Then every accumulation point of* $\{z_k\}$ *is a dual solution. In particular,* $S(P^*) \neq \emptyset$.

Proof. By (15), boundedness of the sequence $\{(u_k, c_k)\}$ implies that

$$\sum_{k=0}^{\infty} s_k \|f_k\| < +\infty. \tag{17}$$

Let $(\bar{u}, \bar{c})$ be an accumulation point of the sequence $\{(u_k, c_k)\}$, and denote by $\mathcal{K}$ the infinite set of indices such that

$$\lim_{\substack{k \in \mathcal{K}, \\ k \to \infty}} (u_k, c_k) = (\bar{u}, \bar{c}).$$

We will prove that $(\bar{u}, \bar{c}) \in S(P^*)$. By boundedness of $\{x_k\}$, we can also assume that the whole sequence $\{x_k\}_{k \in \mathcal{K}}$ converges to some $\overline{x}$. If $f(\overline{x}) = 0$, we claim that $\overline{x} \in X(\overline{u}, \overline{c})$. In this case, Theorem 4 implies that $(\overline{u}, \overline{c}) \in S(P^*)$. Indeed, by definition of x_k, we have that

$$f_0(x_k) + c_k \|f(x_k)\| - \langle u_k, f(x_k)\rangle \leqslant f_0(x) + c_k \|f(x)\| - \langle u_k, f(x)\rangle,$$

for all $x \in X$ and for all k. Taking limits for $k \in \mathcal{K}, k \to \infty$ in the above expression we get

$$f_0(\bar{x}) + \bar{c}\|f(\bar{x})\| - \langle \bar{u}, f(\bar{x})\rangle \leqslant f_0(x) + \bar{c}\|f(x)\| - \langle \bar{u}, f(x)\rangle,$$

for all $x \in X$. Hence $\bar{x} \in X(\bar{u}, \bar{c})$ and thus $(\bar{u}, \bar{c}) \in S(P^*)$. Assume now that $f(\bar{x}) \neq 0$. This fact, together with (17), implies that the sequence $\{s_k\}_{k \in \mathcal{K}}$ converges to zero. Using also (16) for $k \in \mathcal{K}$, we conclude that the subsequence of dual values $\{H_k\}_{k \in \mathcal{K}}$ converges to $\overline{H}$. By upper- semicontinuity of H we have that

$$H(\overline{u}, \overline{c}) \geqslant \limsup_{\substack{k \in \mathcal{K}, \\ k \to \infty.}} H(u_k, c_k) = \overline{H}.$$

This shows that $H(\overline{u}, \overline{c})$ has optimal functional value $\overline{H}$ and hence $(\overline{u}, \overline{c}) \in S(P^*)$. The proof is complete. □

REMARK 2. Consider again Example 2, with $(c_0, u_0) = (1, 0)$, $\varepsilon_k = (1/2)s_k$ and s_k as in (10). With these choices, the dual sequence generated by the MSG algorithm is given by

$$c_k = 1 + (3/5)[(3/2)^k - 1],$$
$$u_k = (2/5)[1 - (3/2)^k]$$

and hence the dual sequence is unbounded. This fact follows from Theorem 6. Indeed, since (10) satisfies (16), if $\{z_k\}$ were bounded, then we would have had $S(P^*) \neq \emptyset$. However in this case $S(P^*) = \emptyset$, and hence $\{z_k\}$ is unbounded.

We now give a useful and simple estimate.

LEMMA 2. *Fix* $z = (u, c) \in \mathbb{R}^m \times \mathbb{R}_+$. *Let* $d_k = d(z, z_k)$ *Then*

$$d_{k+1} - d_k \leqslant (s_k^2 + (s_k + \varepsilon_k)^2)\|f_k\|^2 - 2s_k(H(u,c) - H_k) - 2\varepsilon_k\|f_k\|(c - c_k).$$

Proof. Note that

$$\begin{aligned} d_{k+1} - d_k &= \|z - z_{k+1}\|^2 - \|z - z_k\|^2 \\ &= \|z_k - z_{k+1}\|^2 + 2\langle z - z_k, z_k - z_{k+1}\rangle. \end{aligned}$$

Let $A_k := \|z_k - z_{k+1}\|^2$ and $B_k := \langle z - z_k, z_k - z_{k+1}\rangle$. Using the definition of the MSG algorithm, we can write

$$A_k = \|u_k - u_{k+1}\|^2 + |c_k - c_{k+1}|^2 = \|f_k\|^2(s_k^2 + (s_k + \varepsilon_k)^2).$$

Combining the two previous expressions, we get

$$d_{k+1} - d_k = \|f_k\|^2(s_k^2 + (s_k + \varepsilon_k)^2) + 2B_k. \tag{18}$$

The term B_k is written out as follows.

$$\begin{aligned} B_k &= \langle u - u_k, u_k - u_{k+1}\rangle + (c - c_k)(c_k - c_{k+1}) \\ &= \langle u - u_k, s_k f_k\rangle - (c - c_k)(s_k + \varepsilon_k)\|f_k\| \\ &= s_k[\langle u - u_k, f_k\rangle - (c - c_k)\|f_k\|] - \varepsilon_k(c - c_k)\|f_k\|. \end{aligned} \tag{19}$$

In order to estimate the expression between brackets, we use the subgradient inequality:

$$H(u,c) \leqslant H(u_k, c_k) + \langle(-f_k, \|f_k\|), (u - u_k, c - c_k)\rangle$$

or

$$[\langle u - u_k, f_k\rangle - (c - c_k)\|f_k\|] \leqslant H_k - H(u,c).$$

Using this in (19), we obtain

$$B_k \leqslant s_k(H_k - H(u,c)) - \varepsilon_k(c - c_k)\|f_k\|.$$

Equation (18) now yields

$$d_{k+1}-d_k \leqslant \|f_k\|^2(s_k^2+(s_k+\varepsilon_k)^2)-2[s_k(H(u,c)-H_k) \\ +\varepsilon_k(c-c_k)\|f_k\|], \tag{20}$$

which completes the proof. □

Lemma 2 allows us to prove that the dual sequence is convergent. For proving that the limit is in fact optimal we will need extra assumptions on the step-sizes (see Theorem 8).

THEOREM 7. *If the sequence* $\{z_k\}$ *is bounded, then it is convergent.*

Proof. Let $\hat{z}$ be an accumulation point of $\{z_k\}$, and $\{z_{k_j}\}_j$ a subsequence converging to $\hat{z}$. Using Lemma 2 for the choice $z:=\hat{z}=(\hat{u},\hat{c})$, we conclude that the sequence $\{d(\hat{z},z_k)\}$ verifies

$$d(\hat{z},z_{k+1})-d(\hat{z},z_k)\leqslant(s_k^2+(s_k+\varepsilon_k)^2)\|f_k\|^2 \\ -2s_k(H(\hat{z})-H_k)-2\varepsilon_k\|f_k\|(\hat{c}-c_k).$$

By Theorem 5(b), $\{H_k\}$ is a strictly increasing sequence for all $s_k,\varepsilon_k>0$, and hence $\lim_k H_k=\sup_k H_k$. Using now upper semi-continuity of H, we get,

$$H(\hat{z})\geqslant\lim_j\sup H_{k_j}=\lim_j H_{k_j}=\sup_i H_i\geqslant H_k,\quad \text{for all } k.$$

So $(H(\hat{z})-H_k)\geqslant 0$ for all k. Using also that $\{c_k\}$ is a strictly increasing sequence, we have that $(\hat{c}-c_k)\geqslant 0$ for all k. Hence,

$$d(\hat{z},z_{k+1})-d(\hat{z},z_k)\leqslant(s_k^2+(s_k+\varepsilon_k)^2)\|f_k\|^2.$$

Since $\{z_k\}$ is bounded, by Lemma 1, the series with $a_k:=(s_k^2+(s_k+\varepsilon_k)^2)\|f_k\|^2$ is finite, and this implies (by [19, Lemma 2.2.2]) that the sequence $\{d(\hat{z},z_k)\}_k$ is convergent. But the subsequence $\{d(\hat{z},z_{k_j})\}_j$ of this sequence converges to zero, and so the whole sequence converges to zero, yielding the uniqueness of the accumulation point.

Fix $\beta>0$. Consider a sequence $\{\tilde{x}_k\}$ such that $\tilde{x}_k\in X(u_k,c_k+\beta)$ for all k. We call such a sequence a *primal sequence*. We prove below that all accumulation points of this sequence are primal solutions.

THEOREM 8 (Primal–Dual Convergence). *Assume that the sequence* $\{z_k\}$ *generated by the MSG algorithm is bounded. Assume also that for some* $\eta>0$ *the step-size* s_k *satisfies*

$$s_k \geqslant \eta \frac{(\overline{H} - H_k)}{\|f_k\|^2}. \tag{21}$$

Then $H_k \to \overline{H}$ and the limit of the sequence $\{z_k\}$ is a dual solution. Additionally, all accumulation points of $\{\tilde{x}_k\}$ are solutions of (P).

Proof. By Theorem 5(b) the sequence $\{H_k\}$ is strictly increasing. Theorem 6 implies that every accumulation point of $\{z_k\}$ is a dual solution. Since $\{z_k\}$ is bounded, Theorem 7 allows us to conclude that $\{z_k\}$ is convergent. Combining these two facts, we get that $\{z_k\}$ converges to a dual solution. From (21), we have that

$$s_k \|f_k\| \geqslant \eta \frac{(\overline{H} - H_k)}{\|f_k\|}.$$

By Lemma 1, boundedness of the dual sequence is equivalent to the fact that $\sum_{k=0}^{\infty}\{(s_k + \varepsilon_k)\|f_k\|\}$ is finite. In particular, the sequence $\{s_k\|f_k\|\}$ tends to zero. Note also that the sequence $\{\|f_k\|\}$ is bounded. These facts, combined with the expression above, yield $H_k \to \overline{H}$. Now we will show that all accumulation points of the primal sequence $\{\tilde{x}_k\}$ are solutions of Problem (P). In order to prove this fact, we will show that the numerical sequence $\{\|\tilde{f}_k\|\}$ has zero as its unique accumulation point. Fix $\beta > 0$ and take $\tilde{x}_k \in X(u_k, c_k + \beta)$ for all k. Take $a \geqslant 0$ as an accumulation point of the sequence $\{\|\tilde{f}_k\|\}$. So there exists a subsequence $\{\|\tilde{f}_{k_j}\|\}$ such that $a = \lim_{j\to\infty}\|\tilde{f}_{k_j}\|$. Then

$$\begin{aligned} H(u_{k_j}, c_{k_j}) = H_{k_j} &\leqslant H(u_{k_j}, c_{k_j} + \beta) + \langle(-\tilde{f}_{k_j}, \|\tilde{f}_{k_j}\|), (0, -\beta)\rangle \\ &\leqslant H(u_{k_j}, c_{k_j} + \beta) - \beta\|\tilde{f}_{k_j}\| \end{aligned}$$

We can rewrite this as

$$\beta\|\tilde{f}_{k_j}\| \leqslant H(u_{k_j}, c_{k_j} + \beta\|f_{k_j}\|) - H_{k_j} \leqslant \overline{H} - H_{k_j}.$$

Using the fact that $\lim_j \overline{H} - H_{k_j} = 0$ we get

$$a = \lim_{j\to\infty}\|\tilde{f}_{k_j}\| = 0. \tag{22}$$

Thus the sequence $\{\|\tilde{f}_k\|\}$ converges to zero. Take now $\tilde{x}$ as an accumulation point of $\{\tilde{x}_k\}$. Since zero is the limit of $\{\|\tilde{f}_k\|\}$, we must have $f(\tilde{x}) = 0$. Without loss of generality, assume the whole sequence $\{\tilde{x}_k\}$ converges to $\tilde{x}$.

Then

$$\begin{aligned}\inf P \leqslant f_0(\tilde{x}) &= \lim_k f_0(\tilde{x}_k) + (c_k+\beta)\|f(\tilde{x}_k)\| - \langle u_k, f(\tilde{x}_k)\rangle \\ &= \lim_k \min_{x\in X} f_0(x) + (c_k+\beta)\|f(x)\| - \langle u_k, f(x)\rangle \\ &= \lim_k H(u_k, c_k+\beta) \leqslant \overline{H} = \sup P^*,\end{aligned}$$

where we have used the definition of $\tilde{x}_k$ in the second equality. By weak duality, we must have $f_0(\tilde{x}) = \inf P$ and since $f(\tilde{x}) = 0, \tilde{x}$ is a primal solution. □

4.2. ON A SPECIAL CHOICE OF s_k

In this subsection we study a special choice of the parameter s_k for which nonemptiness of $S(P^*)$ is equivalent to the boundedness of $\{z_k\}$. The step-size we consider is as follows:

$$\eta \frac{\overline{H} - H_k}{\|f_k\|^2} \leqslant s_k \leqslant 2\frac{\overline{H} - H_k}{\|f_k\|^2} \tag{23}$$

with $\eta \in (0, 2)$.

For establishing the announced fact, we need an auxiliary result.

LEMMA 3. *Assume that $S(P^*) \neq \emptyset$ and let $\{z_k\}$ be the sequence generated by the MSG algorithm with step-size $\{s_k\}$ satisfying*

$$\liminf_k \left[2\frac{\overline{H} - H_k}{\|f_k\|^2} - s_k \right] > -\infty. \tag{24}$$

Then $\{z_k\}$ is bounded.

Proof. Fix a dual solution $(\bar{u}, \bar{c}) \in S(P^*)$. For contradiction purposes, assume that $\{z_k\}$ is unbounded. This means that either $\{u_k\}$ or $\{c_k\}$ are unbounded. If $\{u_k\}$ is unbounded, then by (15) $\{c_k\}$ must be unbounded. Hence in either case we must have $\{c_k\}$ unbounded. Since it is a strictly increasing sequence, it tends to infinity. On the other hand, by definition of the MSG algorithm,

$$\begin{aligned}\|\bar{u} - u_{k+1}\|^2 &= \|\bar{u} - u_k + s_k f_k\|^2 \\ &= \|\bar{u} - u_k\|^2 + 2s_k\langle \bar{u} - u_k, f_k\rangle + s_k^2\|f_k\|^2.\end{aligned} \tag{25}$$

In order to estimate the middle term of the expression above we use the subgradient inequality,

$$\overline{H} - H_k \leqslant \langle \bar{u} - u_k, -f_k\rangle + (\bar{c} - c_k)\|f_k\|. \tag{26}$$

Multiply both sides by $2s_k$ and rearrange the resulting expression, to get

$$2s_k\langle \bar{u}-u_k, f_k\rangle \leqslant -2s_k(\overline{H}-H_k)+2s_k(\bar{c}-c_k)\|f_k\|.$$

Combine this fact with (25) to obtain

$$\begin{aligned}\|\bar{u}-u_{k+1}\|^2 &\leqslant \|\bar{u}-u_k\|^2-2s_k(\overline{H}-H_k)+2s_k(\bar{c}-c_k)\|f_k\|+s_k^2\|f_k\|^2\\ &=\|\bar{u}-u_k\|^2+s_k\|f_k\|^2\left[s_k-\frac{2(\overline{H}-H_k)}{\|f_k\|^2}+\frac{2(\bar{c}-c_k)}{\|f_k\|}\right].\end{aligned} \tag{27}$$

Assumption (24) means that there exist a constant $\rho\in\mathbb{R}$ and an index k_0 such that

$$s_k-\frac{2(\overline{H}-H_k)}{\|f_k\|^2}\leqslant\rho$$

for all $k\geqslant k_0$. As pointed out above, $\{c_k\}$ tends to infinity and hence there exists an index $k_1\geqslant k_0$ such that

$$\rho\leqslant\frac{2(c_k-\bar{c})}{\|f_k\|}\quad\text{for all } k\geqslant k_1,$$

where we are also using the fact that the sequence $\{\|f_k\|\}$ is bounded. Altogether, we conclude that for all $k\geqslant k_1$,

$$s_k-\frac{2(\overline{H}-H_k)}{\|f_k\|^2}+\frac{2(\bar{c}-c_k)}{\|f_k\|}\leqslant 0.$$

This fact, combined with (27), yields $\|\bar{u}-u_{k+1}\|\leqslant\|\bar{u}-u_k\|$ for all $k\geqslant k_1$ and this implies that $\{u_k\}$ is bounded. Using Cauchy–Schartz inequality in (26), we get

$$(c_k-\bar{c})\|f_k\|\leqslant-(\overline{H}-H_k)+\|f_k\|\|\bar{u}-u_k\|\leqslant\|f_k\|\|\bar{u}-u_k\|.$$

Since we are assuming that $f_k\neq 0$ for all k, we get

$$(c_k-\bar{c})\leqslant\|\bar{u}-u_k\|$$

and hence $\{c_k\}$ must be bounded, a contradiction. This implies that the sequence $\{z_k\}$ must be bounded. □

Condition (24) is not practical from an algorithmic point of view, because it cannot be verified during the process. For this reason, and also for simplicity of exposition, we replace it by the right-hand side inequality in (23), which can be effectively checked at each iteration. The latter condition readily implies (24).

THEOREM 9. *Assume the step-size in the MSG algorithm is chosen according to* (23), *then the following statements are equivalent.*

(a) *The sequence* $\{z_k\}$ *is bounded.*
(b) $S(P^*) \neq \emptyset$.

Proof. The fact that (a) implies (b) is a consequence of Theorem 6 and the left-hand side inequality in (23). Indeed, Theorem 6 implies that every accumulation point of $\{z_k\}$ is a dual solution. In particular, $S(P^*)$ is non-empty. In order to show that (b) implies (a), observe that this follows from Lemma 3 and the fact that the right-hand side inequality in (23) implies (24). □

In the theorem below, we consider again the sequence $\{\tilde{x}_k\}$ such that $\tilde{x}_k \in X(u_k, c_k + \beta)$, where $\beta > 0$. We recover the same convergence results as the ones reported in Theorem 8, but without the assumption of boundedness of $\{z_k\}$.

THEOREM 10. *Assume the step-size in the MSG algorithm is chosen according to* (23). *Suppose also that* $S(P^*) \neq \emptyset$. *Then,*

(i) *The dual sequence* $\{z_k\}$ *converges to a dual solution.*
(ii) *The sequence of dual values* $\{H_k\}$ *converges to an optimal dual value.*
(iii) *All accumulation points of the primal sequence* $\{\tilde{x}_k\}$ *are solutions of Problem* (P).

Proof. By Theorem 9 and the fact that $S(P^*) \neq \emptyset$, we conclude that $\{z_k\}$ is bounded. Using now the left-hand side of (23) and Theorem 8, we conclude that statements (i)–(iii) hold. □

The following simple result will be useful in the next section.

PROPOSITION 1. *Assume that one of the following conditions holds:*

(i) *The stepsize* s_k *satisfies* (16) *and* $\{z_k\}$ *is bounded.*
(ii) *The stepsize* s_k *satisfies* (23) *and* $S(P^*) \neq \emptyset$.
Then there exists a dual solution $(\bar{u}, \bar{c})$ *such that* $\bar{c} > c_k$ *for all* k.

Proof. Under assumption (i), by Theorem 6 it holds that $S(P^*) \neq \emptyset$. So under either assumption, we must have $S(P^*) \neq \emptyset$. Now fix a dual solution $(\bar{u}, \bar{c})$. Under assumption (ii), and using Theorem 9, we conclude that the sequence $\{z_k\}$ is bounded. So again under either assumption, we must have $\{z_k\}$ bounded. Thus there exists $\hat{c} \geqslant c_k$ for all k. Using Remark 1, we have that $(\bar{u}, \hat{c} + \bar{c})$ is also a dual solution, and this dual solution is as in the statement of the proposition. □

5. Numerical Implementation

In Section 5.1, we select practical step-size parameters for a numerical implementation of the MSG algorithm. In Section 5.2, we demonstrate these step-sizes on test problems.

5.1. STEP-SIZE SELECTION

We assume that the dual sequence $\{z_k\}$ generated by the MSG algorithm is bounded. The step-size s_k given by (21) is quite general in the sense that the constant η can be chosen to be any positive number. Although a wide range of step-sizes can be chosen using (21), we will rather restrict our attention to the estimate given in Lemma 2 in deriving a step-size, because it reflects the structure of the problem.

For simplicity let s_k and ε_k be related through $s_k = \alpha\varepsilon_k$, where $\alpha > 0$. Since the hypotheses of Proposition 1(i) are satisfied, there exists a dual solution $(\bar{u}, \bar{c})$ such that $\bar{c} \geqslant c_k$ for all k. Now, Lemma 2 for $z := (\bar{u}, \bar{c})$ and $\varepsilon_k = \alpha s_k$ yields, after some trivial manipulations

$$s_k\left[\frac{2[(\overline{H} - H_k) + \alpha(\bar{c} - c_k)\|f_k\|]}{q(\alpha)\|f_k\|^2} - s_k\right] \leqslant \frac{d_k - d_{k+1}}{q(\alpha)\|f_k\|^2}, \tag{28}$$

where $q(\alpha) = 1 + (1+\alpha)^2$. Taking now s_k such that

$$s_k = \delta\frac{(\overline{H} - H_k) + \alpha(\bar{c} - c_k)\|f_k\|}{q(\alpha)\|f_k\|^2}, \quad 0 < \delta < 2, \tag{29}$$

we see that the left-hand side of the expression in (28) is nonnegative, and therefore the sequence $\{d_k\}$ is nonincreasing, and hence convergent. This choice of s_k forces convergence of the dual values towards the optimal value $\overline{H}$.

The most commonly used step-sizes for the classical subgradient method are the *infinite series rule* and the *known optimal value rule* (see, e.g., [12]). As can be seen, the step-size we propose above for the MSG algorithm also requires the knowledge of the optimal value. Choosing a value for the unknown $\overline{H}$ in the computation of s_k is an important issue in subgradient methods. Note that any feasible solution of (P) provides an upper bound for $\overline{H}$. According to Bazaraa and Sherali [4] $\overline{H}$ can be chosen as a convex combination of a fixed upper bound and the current best dual value. On the other hand, Sherali et al. [27] recently proposed the so-called variable target value method which assumes no a priori knowledge regarding $\overline{H}$. In the numerical experiments of the present paper we simply use an upper bound as an estimate $\widehat{H}$ for $\overline{H}$ to illustrate the behaviour of the MSG algorithm. However, the above-mentioned schemes for updating the estimate $\widehat{H}$

in each iteration can as well be utilized to increase the efficiency of the algorithm.

Recall that any $\hat{c} \geqslant \bar{c}$ is also a dual solution. So we can replace $\bar{c}$ in (29) by an upper bound $\hat{c}$ for $\bar{c}$. We set the step-size s_k as

$$s_k = \delta \frac{(\widehat{H} - H_k) + \alpha(\hat{c} - c_k)\|f_k\|}{q(\alpha)\|f_k\|^2} \tag{30}$$

where $q(\alpha) = 1 + (1+\alpha)^2$.

It must be noted that when $\overline{H}$ is replaced by $\widehat{H} \geqslant \overline{H}$ the right-hand inequality in (23) does not hold. Therefore Theorems 9 and 10 cannot be stated. However, the main convergence theorems 6–8 hold.

5.2. COMPUTATIONAL TEST PROBLEMS

In this section we apply the MSG algorithm with the new and the earlier [8] step-size parameters (Cases (i) and (iii) below, respectively) to three test problems. We make comparisons among these two cases and the classical subgradient method (Case (ii) below).

Case (i). $s_k = \delta \dfrac{(\widehat{H} - H_k) + \alpha(\hat{c} - c_k)\|f_k\|}{[1 + (1+\alpha)^2]\|f_k\|^2}, \quad \varepsilon_k = \alpha s_k, \;\; 0 < \delta < 2, \;\; \alpha > 0.$

Case (ii). $s_k = \delta \dfrac{\widehat{H} - H_k}{2\|f_k\|^2}, \quad (\varepsilon_k = 0), \;\; 0 < \delta < 2.$

Case (iii). $s_k = \delta \dfrac{\widehat{H} - H_k}{5\|f_k\|^2}, \quad \varepsilon_k = 0.95 s_k, \;\; 0 < \delta < 2.$

For the solution of the subproblem in Step k.1 we used the MATLAB function m-file `fminsearch`. In this section, x_i^k, u_i^k will stand for the ith coordinates of the iterates x_k, u_k, and c^k will stand for c_k. In the tables reporting the results, x^{-1} denotes the initial guess used for the subproblem when $k = 0$.

Problem 1. The following nonconvex problem has become a common test problem (see [7]). It was originally studied by Murtagh and Saunders[15].

$$P_1 : \begin{cases} \min & f_0(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^3 + (x_3 - x_4)^4 \\ & \qquad + (x_4 - x_5)^4, \\ \text{subject to} & f_1(x) = x_1 + x_2^2 + x_3^2 - 3\sqrt{2} - 2 = 0, \\ & f_2(x) = x_2 - x_3^2 + x_4 - 2\sqrt{2} + 2 = 0, \\ & f_3(x) = x_1 x_5 - 2 = 0. \end{cases}$$

Table 1. Problem 1 – Case (i) with $\delta = 0.5, \alpha = 5, \widehat{H} = 0.1, \hat{c} = 2$

k	x_1^k	x_2^k	x_3^k	x_4^k	x_5^k	c^k	u_1^k	u_2^k	u_3^k	$\Vert f(x^k)\Vert$	$f_0(x^k)$	$H(u^k, c^k)$
−1	0.00000	0.00000	0.00000	0.00000	0.00000					6.61	1.00000	
0	1.40745	1.57585	1.31868	1.87329	2.28086	1.000	0.00000	1.000	1.000	1.50	0.33359	−0.25987
1	1.11664	1.22044	1.53779	1.97277	1.79109	1.425	0.00278	0.958	0.943	3.9×10^{-5}	0.02931	0.02931

Table 2. Problem 1 – Case (ii) with $\delta = 0.5, \widehat{H} = 0.1$

k	x_1^k	x_2^k	x_3^k	x_4^k	x_5^k	c^k	u_1^k	u_2^k	u_3^k	$\Vert f(x^k)\Vert$	$f_0(x^k)$	$H(u^k, c^k)$
−1	0.00000	0.00000	0.00000	0.00000	0.00000					6.61	1.00000	
0	1.40745	1.57585	1.31868	1.87329	2.28086	1.000	0.00000	1.000	1.000	1.50	0.33359	−0.25987
1	1.34254	1.53502	1.35489	1.84465	2.18073	1.060	0.00236	0.965	0.952	1.17	0.23052	−0.09894
2	1.28971	1.49004	1.39045	1.82401	2.08792	1.102	0.00440	0.939	0.918	8.9×10^{-1}	0.16524	−0.01104
3	1.23202	1.42169	1.43625	1.81843	1.97425	1.134	0.00597	0.919	0.894	5.5×10^{-1}	0.11173	0.03541
4	1.11657	1.22035	1.53783	1.97299	1.79119	1.163	0.00737	0.901	0.894	3.3×10^{-5}	0.02931	0.02931

Table 3. Problem 1 – Case (iii) with $\delta = 0.5, \widehat{H} = 0.1$

k	x_1^k	x_2^k	x_3^k	x_4^k	x_5^k	c^k	u_1^k	u_2^k	u_3^k	$\Vert f(x^k)\Vert$	$f_0(x^k)$	$H(u^k, c^k)$
−1	0.00000	0.00000	0.00000	0.00000	0.00000					6.61	1.00000	
0	1.40745	1.57585	1.31868	1.87329	2.28086	1.000	0.00000	1.000	1.000	1.50	0.33359	−0.25987
1	1.36990	1.55374	1.33872	1.85665	2.22469	1.060	0.00094	0.986	0.981	1.31	0.27087	−0.16018
2	1.33728	1.53096	1.35833	1.84250	2.17223	1.085	0.00184	0.974	0.965	1.15	0.22319	−0.08825
3	1.30787	1.50681	1.37776	1.83058	2.12152	1.118	0.00263	0.964	0.952	9.8×10^{-1}	0.18572	−0.03645
4	1.27974	1.47974	1.39795	1.82119	2.06969	1.144	0.00332	0.955	0.941	8.3×10^{-1}	0.15470	0.00072
5	1.24903	1.44446	1.42197	1.81638	2.00949	1.168	0.00391	0.948	0.932	6.5×10^{-1}	0.12581	0.02709
6	1.11663	1.22044	1.53779	1.97277	1.79110	1.189	0.00445	0.941	0.923	3.0×10^{-5}	0.02931	0.02931

The reported solution [7] is

$$\bar{x} = (1.1166, 1.2204, 1.5378, 1.9728, 1.7911)$$

which results in

$$\bar{f_0} = 0.0293,$$

satisfying the constraints with

$$\Vert f(\bar{x})\Vert = 8.6 \times 10^{-5}.$$

Our solution coincides with the reported one up to four digits of accuracy after the decimal point. The iterations using Cases (i)–(iii) are depicted in Tables 1–3. The classical subgradient algorithm (Case (ii)) and the MSG algorithm with the step-sizes in Case (iii) reach the solution in 5 and 7 iterations, respectively. With the new step-sizes in Case (i), the MSG algorithm yields the solution in just two iterations.

Problem 2. Consider the following quadratic integer programming problem [7].

$$\begin{cases} \min & f_0(x) = a^T x + \frac{1}{2} x^T Q x \\ \text{subject to} & -1 \leqslant x_1 x_2 + x_3 x_4 \leqslant 1 \\ & -3 \leqslant x_1 + x_2 + x_3 + x_4 \leqslant 2 \\ & x_i \in \{-1, 1\}, \end{cases}$$

where

$$a^T = [6 \ 8 \ 4 \ -2], \qquad Q = \begin{bmatrix} -1 & 2 & 0 & 0 \\ 2 & -1 & 2 & 0 \\ 0 & 2 & -1 & 2 \\ 0 & 0 & 2 & -1 \end{bmatrix}$$

Let $g_1(x) := x_1 x_2 + x_3 x_4$, $g_2(x) := x_1 + x_2 + x_3 + x_4$. Then we can re-write the above problem with continuous equality constraints as

$$P_2 : \begin{cases} \min & f_0(x) = a^T x + \frac{1}{2} x^T Q x \\ \text{subject to} & f_1(x) = \max(0, g_1(x)) - 1) = 0 \\ & f_2(x) = \max(0, -(g_1(x) + 1)) = 0 \\ & f_3(x) = \max(0, g_2(x) - 2) = 0 \\ & f_4(x) = \max(0, -(g_2(x) + 3)) = 0 \\ & f_5(x) = \sum_{i=1}^{4} |(x_i - 1)(x_i + 1)| = 0 \end{cases}$$

The solutions obtained by the MSG and classical subgradient algorithms agree with the global solution reported in [7]. The iterations for Cases (i)–(iii) are shown in Tables 4–6. While the classical subgradient algorithm (Case (ii)) takes 10 iterations to achieve the solution, the MSG algorithm with the step-sizes in Case (iii) requires 15 iterations. The MSG algorithm with the new step-sizes in Case (i) obtains the solution in only two iterations.

Problem 3. The following problem concerns finding a bang–bang constrained time-optimal control of the van der Pol system, which is also

Table 4. Problem 2 – Case (i) with $\delta = 0.1, \alpha = 3, \widehat{H} = -19.00, \hat{c} = 20$

k	x_1^k	x_2^k	x_3^k	x_4^k	c^k	u_1^k	u_2^k	u_3^k	u_4^k	u_5^k	$\|f(x^k)\|$	$f_0(x^k)$	$H(u^k, c^k)$
−1	−2.0000	−2.0000	−2.0000	−2.0000							14.76	−16.000	
0	−1.4279	−1.0000	−1.5582	1.5582	1.00	−1.00	−1.00	−1.00	−1.00	−1.00	3.89	−28.747	−20.958
1	−1.0000	−1.0000	−1.0000	1.0000	2.35	−1.00	−1.00	−1.00	−1.00	−1.34	0	−20.000	−20.000

Table 5. Problem 2 – Case (ii) with $\delta = 0.1$, and $\widehat{H} = -19.00$

k	x_1^k	x_2^k	x_3^k	x_4^k	c^k	u_1^k	u_2^k	u_3^k	u_4^k	u_5^k	$\lVert f(x^k)\rVert$	$f_0(x^k)$	$H(u^k, c^k)$
−1	−2.0000	−2.0000	−2.0000	−2.0000							14.76	−16.000	
0	−1.4279	−1.0000	−1.5582	1.5582	1.00	−1.00	−1.00	−1.00	−1.00	−1.00	3.89	−28.747	−20.958
1	−1.3554	−1.0000	−1.5347	1.5347	1.03	−1.00	−1.00	−1.00	−1.00	−1.03	3.55	−28.046	−20.771
2	−1.2873	−1.0000	−1.5125	1.5122	1.05	−1.00	−1.00	−1.00	−1.00	−1.05	3.23	−27.349	−20.602
3	−1.2248	−1.0000	−1.4916	1.4916	1.07	−1.00	−1.00	−1.00	−1.00	−1.07	2.95	−26.790	−20.449
4	−1.1671	−1.0004	−1.4720	1.4725	1.10	−1.00	−1.00	−1.00	−1.00	−1.10	2.70	−26.243	−20.310
5	−1.1445	−1.0000	−1.3379	1.3379	1.12	−1.00	−1.00	−1.00	−1.00	−1.12	1.89	−24.455	−20.207
6	−1.1042	−1.0000	−1.2326	1.2326	1.16	−1.00	−1.00	−1.00	−1.00	−1.16	1.26	−23.014	−20.107
7	−1.0530	−1.0000	−1.1119	1.1119	1.20	−1.00	−1.00	−1.00	−1.00	−1.20	5.8×10^{-1}	−21.423	−20.028
8	−1.0000	−1.0000	−1.0000	1.0007	1.29	−1.00	−1.00	−1.00	−1.00	−1.29	1.5×10^{-3}	−20.003	−20.000
9	−1.0000	−1.0000	−1.0000	1.0000	35.3	−1.00	−1.00	−1.00	−1.00	−35.3	0	−20.000	−20.000

Table 6. Problem 2 – Case (iii) with $\widehat{H} = -19.00, \delta = 0.1$

k	x_1^k	x_2^k	x_3^k	x_4^k	c^k	u_1^k	u_2^k	u_3^k	u_4^k	u_5^k	$\lVert f(x^k)\rVert$	$f_0(x^k)$	$H(u^k, c^k)$
−1	−2.0000	−2.0000	−2.0000	−2.0000							14.76	−16.000	
0	−1.4279	−1.0000	−1.5582	1.5582	1.00	−1.00	−1.00	−1.00	−1.00	−1.00	3.89	−28.747	−20.958
1	−1.3846	−1.0000	−1.5442	1.5442	1.01	−1.00	−1.00	−1.00	−1.00	−1.01	3.69	−28.327	−20.845
2	−1.3429	−1.0000	−1.5304	1.5310	1.04	−1.00	−1.00	−1.00	−1.00	−1.02	3.49	−27.925	−20.740
3	−1.3032	−1.0000	−1.5176	1.5176	1.06	−1.00	−1.00	−1.00	−1.00	−1.03	3.30	−27.542	−20.640
4	−1.2650	−1.0000	−1.5050	1.5050	1.08	−1.00	−1.00	−1.00	−1.00	−1.04	3.13	−27.175	−20.546
5	−1.2304	−1.0002	−1.4934	1.4934	1.10	−1.00	−1.00	−1.00	−1.00	−1.05	2.97	−26.834	−20.457
6	−1.2011	−1.0000	−1.4680	1.4694	1.12	−1.00	−1.00	−1.00	−1.00	−1.06	2.76	−26.372	−20.373
7	−1.1720	−1.0000	−1.4684	1.4676	1.14	−1.00	−1.00	−1.00	−1.00	−1.07	2.68	−26.212	−20.294
8	−1.1498	−1.0073	−1.3705	1.3723	1.15	−1.00	−1.00	−1.00	−1.00	−1.08	2.10	−24.917	−20.230
9	−1.1310	−1.0000	−1.3015	1.3015	1.18	−1.00	−1.00	−1.00	−1.00	−1.09	1.67	−23.951	−20.170
10	−1.1051	−1.0000	−1.2349	1.2349	1.20	−1.00	−1.00	−1.00	−1.00	−1.11	1.27	−23.045	−20.109
11	−1.0745	−1.0000	−1.1611	1.1611	1.24	−1.00	−1.00	−1.00	−1.00	−1.12	8.5×10^{-1}	−22.064	−20.055
12	−1.0343	−1.0000	−1.0712	1.0712	1.29	−1.00	−1.00	−1.00	−1.00	−1.12	3.7×10^{-1}	−20.900	−20.011
13	−1.0001	−1.0086	−1.0000	1.0001	1.40	−1.00	−1.00	−1.00	−1.00	−1.12	1.8×10^{-2}	−20.044	−19.998
14	−1.0000	−1.0000	−1.0000	1.0000	3.57	−1.00	−1.00	−1.00	−1.00	−2.32	0	−20.000	−20.000

studied in [10, 11, 14, 30]. The dynamics of the van der Pol system are given by the ordinary differential equations

$$\begin{aligned} \dot{z}_1(t) &= z_2(t), \\ \dot{z}_2(t) &= -z_1(t) - (z_1^2(t) - 1)z_2(t) + v(t), \end{aligned} \tag{31}$$

where $\dot{z}_i := \mathrm{d}z_i/\mathrm{d}t, i = 1, 2$, the state vector $z : [0, t_f] \to \mathbb{R}^2$ is continuous, the control function $v : [0, t_f] \to \{-1, 1\}$ is bang–bang, i.e. $v(t)$ switches between the values -1 and 1, and t_f is the terminal time. The aim is to get from the initial state $z(0) = (1, 1)$ to the target state $z(t_f) = (0, 0)$ in minimum (terminal) time t_f.

We consider three switchings such that

$$v(t) = (-1)^{k+1} \quad \text{for} \quad t_{k-1} \leqslant t < t_k, \tag{32}$$

where $k = 1, \cdots, 4, t_0 = 0$ and $t_4 = t_f$. So the sequence of control values we use is $\{1, -1, 1, -1\}$. Here t_1, t_2, and t_3 are called the switching times. For convenience, the duration of arcs, namely arc-times, are defined as

$$\xi_k := t_k - t_{k-1} \geqslant 0.$$

Then $t_f = \xi_1 + \xi_2 + \xi_3 + \xi_4$. Let $\xi := (\xi_1, \xi_2, \xi_3, \xi_4)$. The problem of minimizing t_f while getting from $z(0) = (1, 1)$ to $z(t_f) = (0, 0)$ can be stated as

$$\text{P}_3: \begin{cases} \min\limits_{\xi} & f_0(\xi) = \xi_1 + \xi_2 + \xi_3 + \xi_4 \\ \text{subject to} & f_i(\xi) = z_i(\xi_1 + \xi_2 + \xi_3 + \xi_4) = 0, \quad i = 1, 2, \\ & f_3(\xi) = \sum\limits_{k=1}^{4} \min\{0, \xi_k\} = 0, \end{cases}$$

where $f_i(\xi), i = 1, 2$, are obtained by solving the system equations in (31) with (32). Note that the third constraint in Problem (P_3) represents the nonnegativity conditions $\xi_k \geqslant 0$.

Iterations for Cases (i)–(iii) are depicted in Tables 7–9. Figure 1 illustrates the initial guess trajectory of the van der Pol system with a dashed curve in the phase plane. The solution trajectory is shown by a solid curve. The switching points are marked by asterisks. As can be seen, the initial trajectory generated by three switchings, or four arcs, with $\xi^0 = (1, 1, 1, 1)$, is far from getting to the target $z(t_f) = (0, 0)$. The solution trajectory gets to the target in the minimum time t_f=3.09520 with only one switching, or

Table 7. Problem 3 – Case (i) with $\delta = 0.1, \alpha = 5\widehat{H} = 4, \hat{c} = 5$

k	ξ_1^k	ξ_2^k	ξ_3^k	ξ_4^k	c^k	u_1^k	u_2^k	u_3^k	$\lVert x(t_f)\rVert$	t_f	$H(u^k, c^k)$
−1	1.00000	1.00000	1.00000	1.00000					1.94	4.00000	
0	0.00000	0.85420	1.02920	0.00000	2.00000	−1.00000	−1.00000	−5.00000	9.8×10^{-1}	1.88340	2.99265
1	0.00000	0.72300	2.37220	0.00000	2.16121	−1.01019	−0.92004	−5.00000	2.0×10^{-5}	3.09520	3.09520

Table 8. Problem 3 – Case (ii) with $\delta = 0.1, \widehat{H} = 4$

k	ξ_1^k	ξ_2^k	ξ_3^k	ξ_4^k	c^k	u_1^k	u_2^k	u_3^k	$\lVert x(t_f)\rVert$	t_f	$H(u^k, c^k)$
−1	1.00000	1.00000	1.00000	1.00000					1.94	4.00000	
0	0.00000	0.85420	1.02920	0.00000	2.00000	−1.00000	−1.00000	−5.00000	9.8×10^{-1}	1.88340	2.99265
1	0.00000	0.79071	1.53069	0.00000	2.05151	−1.00651	−0.94890	−5.00000	7.0×10^{-1}	2.32140	3.05605
2	0.00000	0.72300	2.37220	0.00000	2.11897	−1.00289	−0.88154	−5.00000	2.0×10^{-5}	3.09520	3.09520

Table 9. Problem 3 – Case (iii) with $\delta = 0.1$, $\widehat{H} = 4$

k	ξ_1^k	ξ_2^k	ξ_3^k	ξ_4^k	c^k	u_1^k	u_2^k	u_3^k	$\|x(t_f)\|$	t_f	$H(u^k, c^k)$
−1	1.00000	1.00000	1.00000	1.00000					1.94	4.00000	
0	0.00000	0.85420	1.02920	0.00000	2.00000	−1.00000	−1.00000	−5.00000	9.8×10^{-1}	1.88340	2.99265
1	0.00000	0.80353	1.42160	0.00000	2.04018	−1.00261	−0.97956	−5.00000	7.7×10^{-1}	2.22513	3.02523
2	0.00000	0.77882	1.64210	0.00000	2.08957	−1.00207	−0.95424	−5.00000	6.2×10^{-1}	2.42092	3.07802
3	0.00000	0.72300	2.37220	0.00000	2.14715	−0.99951	−0.92482	−5.00000	2.0×10^{-5}	3.09520	3.09520

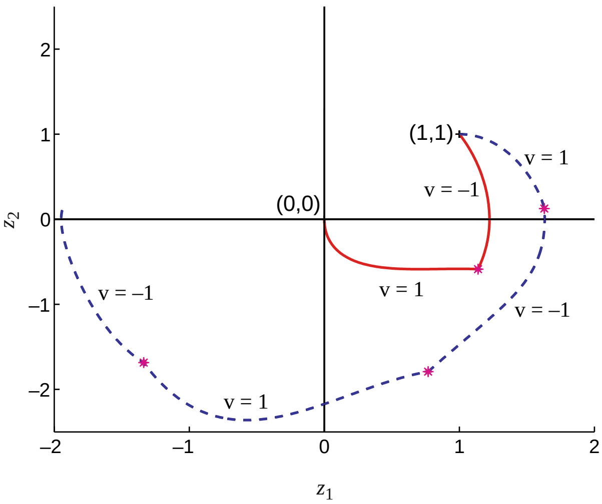

Figure 1. Initial guess and solution trajectories of the van der Pol system.

two arcs, durations of which are given by (0.72300, 2.37220). This result agrees with the solution presented in [11].

5.3. COMMENTS ON THE TEST PROBLEMS

In all the three problems, the MSG algorithm with the new stepsizes given in Case (i) provides a primal solution in fewer iterations than the classical subgradient method (Case (ii)). The classical subgradient method in turn generates a solution in fewer iterations than the MSG algorithm with the stepsizes of Case (iii). In the reported experiments the parameter $\delta \in (0, 2)$ has been chosen small enough to observe the differences between the performances of Cases (i) and (iii).

Acknowledgements

The authors thank Alex Rubinov and Adil Bagirov for useful discussions which improved the quality of the manuscript. Helpful comments by the anonymous referees are acknowledged. During most of this research, Regina Burachik was spending part of her sabbatical leave in the School of Mathematics and Statistics at UniSA, whom she would like to thank for providing a motivating environment and good infrastructure.

References

1. Azimov, A.Y. and Gasimov, R.N. (2002), Stability and duality of nonconvex problems via augmented lagrangian, *Cybernetics and Systems Analysis* 38, 412–421.
2. Azimov, A.Y. and Gasimov, R.N. (1999), On weak conjugacy, weak subdifferentials and duality with zero gap in nonconvex optimization, *International Journal of Applied Mathematics,* 1(4), 171–192.
3. Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (1993), *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons Inc., New York.
4. Bazaraa, M.S. and Sherali, H.D. (1981), On the choice of step sizes in subgradient optimization, *European Journal of Operations Research* 7, 380–388.
5. Bertsekas, D.P. (1995), *Nonlinear Programming,* Athena Scientific, Belmont, Massachusetts.
6. Demyanov, V.F. (1968), Algorithms for some minimax problems, *Journal of Computer and Systems Sciences* 2, 342–380.
7. Floudas, A.C., Pardalos, P.M., Adjiman, C.S., Esposito, W.R., Gümüş, Z.H., Harding, S.T., Klepeis, J.L., Meyer, C.A. and Schweiger, C.A. (1999), *Handbook of Test Problems in Local and Global Optimization*, Kluwer Academic Publishers, the Netherlands.
8. Gasimov, R.N. (2002), Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming, *Journal of Global Optimization* 24, 187–203.
9. Gasimov, R.N. and Rubinov, A.M. (2004), On augmented Lagrangians for optimization problems with a single constraint, *Journal of Global Optimization* 28(2), 153–173.
10. Kaya, C.Y. and Noakes, J.L. (1996), Computations and time-optimal controls, *Optimal Control Applications and Methods* 17, 171–185.
11. Kaya, C.Y. and Noakes, J.L. (2003), Computational method for time-optimal switching control, *Journal of Optimization Theory and Applications* 117(1), 69–92.
12. Kim, S. and Hyunsil, A. (1991), Convergence of a generalized subgradient method for nondifferentiable convex optimization, *Mathematical Programming* 50, 75–80.
13. Konnov, I.V. (2003), On convergence properties of a subgradient method, *Optimization Methods and Software* 18(1), 53–62.
14. Maurer, H. and Osmolovskii, N.P. (2004), Second order sufficient conditions for time-optimal bang–bang control problems, *SIAM Journal on Optimization and Control* 42(6), 2239–2263.
15. Murtagh, B.A. and Saunders, M.A. (1983), *MINOS 5.4 User's Guide*. Systems Optimization Laboratory, Department of Operations Research, Stanford University, Technical Report SOL 83-20R.
16. Polyak, B.T. (1969), The conjugate gradient method in extremal problems, *Z. Vychislitelnoy Matematiki i Matematicheskoy Fiziki* 9, 94–112.
17. Polyak, B.T. (1969), Minimization of unsmooth functionals, *Z. Vychislitelnoy Matematiki i Matematicheskoy Fiziki* 9, 14–29.
18. Polyak, B.T. (1969), Iterative methods using Lagrange multipliers for solving extremal problems with constraints of the equation type, *Z. Vychislitelnoy Matematiki i Matematicheskoy Fiziki* 10, 1098–1106.
19. Polyak, B.T. (1987), *Introduction to Optimization*, Optimization Software Inc., Publications Division, New York.

20. Rockafellar, R.T. and Wets, R.J.-B. (1998), *Variational Analysis*, Springer, Berlin.
21. Rockafellar, R.T. (1993), Lagrange multipliers and optimality, *SIAM Review* 35, 183–238.
22. Rubinov, A.M. (2000), *Abstract Convexity and Global Optimization*, Kluwer Academic Publishers, Dordrecht.
23. Rubinov, A.M., Yang, X.Q., Bagirov, A.M. and Gasimov, R.N. (2003), Lagrange-type functions in constrained optimization, *Journal of Mathematical Sciences* 115(4), 2437–2505.
24. Rubinov, A.M. and Gasimov, R.N. (2002), The nonlinear and augmented Lagrangians for nonconvex optimizations problems with a single constraint, *Applied and Computational Mathematics* 1(2), 142–157.
25. Rubinov, A.M. and Gasimov, R.N. (2003), Strictly increasing positively homogeneous functions with applications to exact penalization, *Optimization* 52(1), 1–28.
26. Rubinov, A.M., Glover, B.M. and Yang, X.Q. (1999), Decreasing functions with applications to penalization, *SIAM Journal on Optimization*, 10, 289–313.
27. Sherali, H.D., Choi, G. and Tuncbilek, C.H. (2000), A variable target value method for nondifferentiable optimization, *Operations Research Letters* 26, 1–8.
28. Shor, N.Z. (1985), *Minimization Methods For Nondifferentiable Functions*, Springer Verlag, Berlin.
29. Shor, N.Z. (1995), Dual estimates in multiextremal problems, *Journal of Global Optimization* 7, 75–91.
30. Simakov, S.T., Kaya, C.Y. and Lucas, S.K. (2002), Computations for time-optimal bang-bang control using a Lagrangian formulation, *Preprints of IFAC 15th Triennial World Congress*, Barcelona, Spain.
31. Yang, X.Q. and Hung, X.X. (2001), A nonlinear Lagrangian approach to constrained optimization problems, *SIAM Journal on Optimization* 14, 1119–1144.