Optimization
Vol. 00, No. 00, Month 200x, 1–21

RESEARCH ARTICLE

The Exact Penalty Map for Nonsmooth and Nonconvex Optimization

Regina S. Burachik[a], Alfredo N. Iusem[b] and Jefferson G. Melo[c,*]

[a] *School of Mathematics and Statistics, University of South Australia, Mawson Lakes, SA 5095, Australia*;

[b] *IMPA, Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro - RJ, CEP 22460-320, Brazil*;

[c] *Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus Samambaia, Goiânia - Go, CEP 74001-970, Brazil*

(*v4.4 released June 2012*)

Augmented Lagrangian duality provides zero duality gap and saddle point properties for nonconvex optimization. On the basis of this duality, subgradient-like methods can be applied to the (convex) dual of the original problem. These methods usually recover the optimal value of the problem, but may fail to provide a primal solution. We prove that the recovery of a primal solution by such methods can be characterized in terms of (i) the differentiability properties of the dual function, and (ii) the exact penalty properties of the primal-dual pair. We also connect the property of finite termination with exact penalty properties of the dual pair. In order to establish these facts, we associate the primal-dual pair to a penalty map. This map, which we introduce here, is a convex and globally Lipschitz function, and its epigraph encapsulates information on both primal and dual solution sets.

Keywords: Duality scheme, Exact Penalty, Augmented Lagrangians, Nonconvex optimization, Nonsmooth optimization, Banach spaces.

AMS Subject Classification: 49M29; 90C26; 65K10.

1. Introduction

A main tool for solving extended real-valued optimization problems is provided by augmented Lagrangians and their corresponding duality schemes. In Rockafellar and Wets [15], important duality results such as zero duality gap, saddle point properties, and exact penalty representations are established regardless of the convexity or smoothness of the primal problem. These results have been extended to infinite dimensional spaces, for instance, in [10, 13, 14, 16, 18]. Zero duality properties of augmented Lagrangians allow the use of solution methods which solve the (nonsmooth and convex) dual problem. The works [2, 3, 6, 7, 9, 12] use a deflected subgradient algorithm (DSG) for solving the dual problem induced by an augmented Lagrangian. DSG is an example of what we call a *primal-dual method*. Roughly speaking, a primal-dual method solves the dual problem induced by a suitable augmented Lagrangian. The search direction used at each step of the method requires the computation of a primal variable. In this way, a primal-dual sequence

*Corresponding author. Email: Jefferson@mat.ufg.br

ISSN: 0233-1934 print/ISSN 1029-4945 online
© 200x Taylor & Francis
DOI: 10.1080/0233193YYxxxxxxxx
http://www.informaworld.com

is automatically generated by the method, and a key question is whether or not every accummulation point of the primal sequence is a solution of the original (primal) problem. We call the latter property *primal convergence*. An even stronger property of a primal-dual method is *finite termination*, which means that primal convergence is achieved in a finite number of steps. Our aim is to establish a relationship between these types of primal convergence and exact penalty properties of the augmented Lagrangian. For establishing these relationships, a key tool turns out to be the *penalty map*, a globally Lipschitz convex function defined in the space of dual variables. This map is also instrumental for understanding the structure of the dual solution set.

The penalty map allows us to characterize primal convergence in terms of the differentiability of the dual function at a dual solution. More precisely, we prove that the differentiability of the dual function at a dual solution is in turn equivalent to *strong exactness* (see Definition 3.1) of the penalty map (see Section 3). As a concrete application of our theory, we use the penalty map to study primal convergence and finite termination of a wide family of primal-dual methods, which include DSG as a particular case (see Theorem 4.4). In the particular case of DSG, we fully characterize primal convergence in terms of the penalty map (see Corollary 5.3 and Theorem 6.1).

As far as we know, the first results relating penalty parameters and convergence properties of primal-dual methods are reported in [8], where the authors prove, in a finite dimensional setting, the equivalence between the differentiability of the dual function at the dual limit and primal convergence. The work [8] considers an optimization problem with one nonnegative equality constraint in finite dimensional spaces. For this setting, the authors showed that primal convergence of DSG is equivalent to the differentiability of the dual function at a dual solution and the existence of exact penalty parameters. These results were later on extended to more general problems, and more general primal-dual methods, in [1], where still a finite dimensional setting is used.

In spite of these results, a question remains as to which is the precise interplay between

(i) exact penalty properties,
(ii) primal convergence of solution methods, and
(iii) differentiability of the dual function at the dual limit.

These properties can be analyzed by means of the penalty function. The purpose of the present paper is to introduce the penalty map and use it to provide a unified setting for relating properties (i)-(iii). Our analysis is carried out in a very general setting, where the primal problem is formulated in a reflexive Banach space and the dual problem in a Hilbert space. The structure of the augmented Lagrangian function we use is similar to the one used in [1, 5–7, 10, 18], in which the augmenting function σ satisfies $\sigma(\cdot) \geq \|\cdot\|$, where $\|\cdot\|$ is the norm of the Hilbert space used in the definition of the duality parameterization.

The manuscript is organized as follows. In Section 2 we state our primal-dual problem and recall some preliminary material on augmented Lagrangians. Moreover, we present in this section a technical result (Proposition 2.6), which establishes main connections between primal and dual solution sets. In Section 3 we introduce the penalty map E, and define *exact* and *strongly exact* penalty maps. We illustrate these concepts in Examples 3.2 and 3.3. Moreover, we establish the equivalence between property (iii) above and strong exactness of E (see Theorem 3.6). Dual localization results are established in Section 3.2. In this section we also establish a connection between dual convergence of primal-dual methods and the graph of the penalty map. We prove in Section 3.3 that E is a globally Lipschitz convex function, which is proper if and only if the dual solution set is not empty. As a consequence of this fact, the well known property of a vector y supporting an exact penalty representation (see [15, Definition 11.60]) becomes equivalent to simply y belonging to dom (E) (see Remark **4**). Section 4 is devoted to the application of the new theory to establish primal and finite convergence of primal-dual methods. We relate the penalty map E with a stopping criterion for these methods, and establish minimal assumptions under which primal and finite convergence hold (see Theorem 4.4). In Section 5 we recall the deflected subgradient algorithm and relate its primal convergence with strong exactness of E. Finally, in Section 6 we apply our results to an optimization problem with equality constraints. For this problem, we prove that primal convergence of DSG is equivalent to exactness of E (see Theorem 6.1).

2. Preliminaries

Let X be a reflexive Banach space, and H a Hilbert space. We denote by $\|\cdot\|$ the norm in both X and H. The inner product of $z, y \in H$ is denoted by $\langle y, z\rangle$. We consider the optimization problem

$$\text{minimize } \varphi(x) \text{ subject to } x \text{ in } X, \tag{1}$$

where the function $\varphi : X \to \mathbb{R}_{+\infty} := \mathbb{R} \cup \{+\infty\}$ is a proper (i.e., dom $\varphi \neq \emptyset$ and $\varphi > -\infty$) and weakly lower semicontinuous (w-lsc) function. We consider a duality scheme by means of a *dualizing parameterization* for (1), which is a function $f : X \times H \to \bar{\mathbb{R}} := \mathbb{R}_{+\infty} \cup \{-\infty\}$ that verifies $f(x, 0) = \varphi(x)$ for all $x \in X$.

Definition 2.1: (see [4]) A function $\sigma : H \to \mathbb{R}_{+\infty}$ is said to be a *valley-at-zero* augmenting function if it is proper, w-lsc, and for every $t > 0$:

$$\sigma(0) = 0 \quad \textit{and} \inf_{\|z\| \geq t} \sigma(z) > 0.$$

The *augmented Lagrangian* function $\ell : X \times H \times \mathbb{R}_+ \to \bar{\mathbb{R}}$ corresponding to the dualizing parameterization function f and the augmenting function σ is given by

$$\ell(x, y, r) := \inf_{z \in H} \{f(x, z) - \langle z, y\rangle + r\sigma(z)\}. \tag{2}$$

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

$$\text{maximize } q(y, r) \text{ subject to } (y, r) \in H \times \mathbb{R}_+. \tag{3}$$

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

The next definition has been used in [10, Section 5] and [18].

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

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

Remark 1 : From now on, we only consider dualizing parameterization functions which are proper, weakly level-compact, and weakly lower semicontinuous.

Next we summarize some basic results for our primal-dual pair.

Proposition 2.3:

i) *The dual function q is concave and weakly upper semicontinuous (*w-usc*);*
ii) *the function $q(y,\cdot)$ is nondecreasing for each fixed $y \in H$; in particular, if (y,c) is a dual solution then (y,r) is a dual solution for each $r \geq c$;*
iii) *Suppose that f is a proper,* w-lsc *dualizing parameterization function for problem* (1) *and consider the induced dual problem* (3)*. Assume that f is weakly level-compact and that* $\operatorname{dom} q \neq \emptyset$*. Then there is zero-duality gap between the primal and dual problems, i.e., $M_P = M_D$.*

Proof: Item (i) follows from the fact that q is the infimum of affine functions. Item (ii) follows from the fact that the augmenting function σ is nonnegative. Item (iii) is proved in [18, Theorem 3.1]. □

Remark 2 : The augmented Lagrangians considered in [18, Theorem 3.1] are generated by a function $g(\cdot,\cdot)$ instead of $\langle\cdot,\cdot\rangle$. The prototype is still $g(y,z) = \langle y,z\rangle$. Although [18, Theorem 3.1] assumes weak lower semicontinuity of g (which is not true in general for $g(y,z) = \langle y,z\rangle$), its proof requires only lower semicontinuity of $g(y,\cdot)$ with y fixed. This latter property holds for $g(y,z) = \langle y,z\rangle$ and then Proposition 2.3 (iii) provides the desired zero-duality gap property. Proposition 2.3(iii) can also be deduced from [4, Theorem 3.2], which holds in a more general setting.

Definition 2.4: Fix $(y,c) \in H \times \mathbb{R}_+$ and let $\Phi_{(y,c)} : X \times H \to \bar{\mathbb{R}}$ be defined by

$$\Phi_{(y,c)}(x,z) := f(x,z) - \langle y,z\rangle + c\sigma(z). \tag{4}$$

Define the point-to-set mapping $A : H \times \mathbb{R}_+ \rightrightarrows X \times H$ as the argmin of $\Phi_{(y,c)}$:

$$A(y,c) := \operatorname{argmin}_{(x,z)} \Phi_{(y,c)}(x,z) \subset X \times H. \tag{5}$$

Let $P_X(y,c)$ be the projection of $A(y,c)$ onto X, namely,

$$P_X(y,c) := \{x \in X : (x,z) \in A(y,c), \text{ for some } z \in H\}.$$

Analogously, we denote by $P_H(y,c)$ the projection of $A(y,c)$ in H.

The point-to-set mapping A induces a primal-dual method for problems (1),(3).

Definition 2.5: Let $A : H \times \mathbb{R}_+ \rightrightarrows X \times H$ be defined as in (5). A *primal-dual method* induced by A for problems (1),(3) is any sequence $\{(y_k,c_k),(x_k,z_k)\}$ in the graph of A (i.e., such that $(x_k,z_k) \in A(y_k,c_k)$ for all k).

DSG, as well as many other variants of subgradient methods applied to the dual problem (3), is a primal dual method in the sense of Definition 2.5.

For future reference, we note that

$$q(y,c) = \Phi_{(y,c)}(x,z) \text{ if and only if } (x,z) \in A(y,c). \tag{6}$$

We will also use that

$$\Phi_{(y,c)}(x,z) \geq \inf_{z' \in H, x' \in X} \Phi_{(y,c)}(x',z') = q(y,c). \tag{7}$$

The next proposition summarizes the connection between $P_X(y,c), P_H(y,c)$ and the primal solution set P_*.

Proposition 2.6: *Let P_* and D_* be as in Definition 2.2.*
a) For each $(y,c) \in D_*$, it holds that $P_* \subset P_X(y,c)$ and $0 \in P_H(y,c)$.
b) If $(y,c) \in D_*$ then $P_H(y,t) = \{0\}$ for all $t > c$.
c) If $P_H(y,c) = \{0\}$ then $(y,c) \in D_*$ and $P_X(y,c) = P_*$.

Proof: (a) Consider $x^* \in P_*$. By Proposition 2.3, $M_P = M_D$, and we can write

$$q(y,c) = M_D = M_P = \varphi(x^*) = f(x^*,0) - \langle y, 0\rangle + c\sigma(0) = \Phi_{(y,c)}(x^*,0). \tag{8}$$

Hence $(x^*,0) \in A(y,c)$, which in turn implies that $x^* \in P_X(y,c)$ and $0 \in P_H(y,c)$. Since $x_* \in P_*$ is arbitrary, the result follows.
(b) Take $t > c$. Since $(y,c) \in D_*$, it follows that $(y,t) \in D_*$, by Proposition 2.3 (ii). It follows from item (a) that $0 \in P_H(y,t)$. Take an arbitrary $z \in P_H(y,t)$. This means that there exists $x \in X$ such that $(x,z) \in A(y,t)$ and then $\Phi_{(y,t)}(x,z) = q(y,t) = M_D$. On the other hand, a simple manipulation shows that

$$\begin{aligned} M_D = \Phi_{(y,t)}(x,z) &= \Phi_{(y,c)}(x,z) + (t-c)\sigma(z) \\ &\geq q(y,c) + (t-c)\sigma(z) = M_D + (t-c)\sigma(z). \end{aligned}$$

Therefore $\sigma(z) = 0$, because $t > c$ and σ is nonnegative. As a consequence, $z = 0$ and then $P_H(y,t) = \{0\}$.
(c) If $P_H(y,c) = \{0\}$ then every element in $A(y,c)$ has the form $(x,0)$, with $x \in P_X(y,c)$. It is enough to show that, in this situation, $(y,c) \in D_*$ and $x \in P_*$. Indeed, $(x,0) \in A(y,c)$ implies that

$$q(y,c) = \Phi_{(y,c)}(x,0) = \varphi(x) \geq M_P = M_D \geq q(y,c).$$

The above expression readily yields $(y,c) \in D_*$ and $x \in P_*$. Since $x \in P_X(y,c)$ is arbitrary we conclude that $P_X(y,c) \subset P_*$. The reverse inclusion follows from (a), and therefore $P_X(y,c) = P_*$, concluding the proof. □

The next example shows that the inclusion $P_* \subset P_X(y,c)$ in Proposition 2.6(a) may be strict.

Example 2.7 Consider the problem

$$\min_{x \in \mathbb{R}} \varphi(x) := x + \delta_{[0,1]}(x),$$

with optimal value $M_P = 0$ and $P_* = \{0\}$. Take the parameterization function given by

$$f(x,z) = |x - z| + \delta_{[0,1]}(x) + \delta_{[0,1]}(z),$$

and the augmenting function $\sigma(z) = |z|$. It follows that

$$q(u,c) = \min_{x,z \in [0,1]} \{|x-z| + (c-u)z\} = \min\{0, c-u\}.$$

Direct calculations show that $D^* = \{(u,c) \in \mathbb{R} \times \mathbb{R}_+ \ : \ c \geq u\}$, so $(t,t) \in D_*$ for all $t \geq 0$. We claim that $P_X(t,t) = [0,1] \supsetneq P_*$. In order to show this, note that $x \in P_X(t,t)$ if and only if there exists $z \in \mathbb{R}$ such that $0 = M_D = q(t,t) = \Phi_{(t,t)}(x,z)$.

For this to hold, we must have

$$0 = q(t,t) = \Phi_{(t,t)}(x,z) = |x-z| + \delta_{[0,1]}(x) + \delta_{[0,1]}(z) + t|z| - tz = |x-z|,$$

which yields $x = z \in [0,1]$. Hence $P_X(t,t) = [0,1]$.

3. The Penalty Map

3.1. *Main Properties*

Consider the function $E : H \to \mathbb{R}_+ \cup \{+\infty\}$ defined by

$$E(y) := \inf\{c \geq 0 : (y,c) \in D_*\}. \tag{9}$$

We call E the *penalty map* for the primal problems (1)-(3). The infimum above is $+\infty$ when the argument of the infimum in (9) is empty. Hence,

$$\operatorname{dom} E := \{y \in H \ : \ D_* \neq \emptyset\}.$$

Remark 1: Since the dual function q is w-usc we obtain that $(y, E(y)) \in D_*$ if and only if $y \in \operatorname{dom} E$. Inasmuch D_* is a convex set, it is easy to conclude that the function E is convex. Moreover, the *epigraph* of E encapsulates information on both primal and dual solution sets. Denote the strict epigraph of E by $\operatorname{epi}_s(E) := \{(y,c) \ : \ c > E(y)\}$. We have that:

(i) $\operatorname{epi}_s(E) \subset D_*$,
(ii) If $(y,c) \in \operatorname{epi}_s(E)$ then $P_* = P_X(y,c)$,
(iii) $P_* \subset P_X(y, E(y))$.

All these statements follow directly from Proposition 2.6.

Definition 3.1: The penalty map E is *exact* at $y \in \operatorname{dom}(E)$ iff $P_X(y, E(y)) = P_*$. If for some $y \in \operatorname{dom}(E)$ it holds $P_H(y, E(y)) = \{0\}$, then E is *strongly exact* at y.

Proposition 2.6(c) shows that if E is strongly exact at y then it is exact at y. This motivates our terminology. Next we present an example in which E is exact at y but not strongly exact at y.

Example 3.2 Consider the following primal problem with $\phi : \mathbb{R} \to \mathbb{R}$:

$$\text{minimize } \phi(x) := \begin{cases} \ln(x+1), & \text{if } \ x \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

It is easy to see that $P_* = \{0\}$ and $M_P = 0$. Take $\sigma(z) = |z|$ and a parameterization function f given by

$$f(x,z) := \begin{cases} \ln(x+1) + z, & \text{if } \ x \geq 0, z \in [0,1], \\ +\infty, & \text{otherwise.} \end{cases}$$

From (2), the augmented Lagrangian is obtained as

$$l(x,y,c) = \ln(x+1) + \min\{c - y + 1, 0\} + \delta_{[0,\infty]}(x).$$

Hence the dual function is

$$q(y,c) = \begin{cases} c - y + 1, & \text{if } c < y - 1, \\ 0, & \text{if } c \geq y - 1. \end{cases}$$

Thus,

$$\begin{aligned} D_* &= \{(y,c) \,:\, q(y,c) = M_D = 0\} = \{(y,c) \,:\, \min\{c - y + 1, 0\} = 0\} \\ &= \{(y,c) \,:\, c \geq y - 1\}. \end{aligned}$$

Using now the definition of E we obtain $E(y) = \inf\{c \geq 0 \;:\; c \geq y - 1\} = \max\{0, y-1\}$. Fix $\hat{y} = 1 \in \operatorname{dom}(E)$. We will show that $P_H(\hat{y}, E(\hat{y})) = [0,1]$. We have that $E(\hat{y}) = 0$ and we can write

$$q(1, E(1)) = q(1,0) = \min_{(x,z)}\{f(x,z) - z\} = \min_{z \in [0,1]}\{z - z\} = 0,$$

Since the above optimal value is attained for every $z \in [0,1]$ and for $x = 0$, we conclude that $A(1, E(1)) = A(1,0) = \{0\} \times [0,1]$. In other words, $P_X(1, E(1)) = \{0\} = P_*$ and $P_H(1, E(1)) = [0,1]$. Therefore E is exact but not strongly exact at $y = 1$. Note that q is not differentiable at any point of the form $(y, E(y))$.

In the next example the penalty map E is strongly exact at every $y \in \operatorname{dom}(E)$.

Example 3.3 Consider the following primal problem

$$\text{minimize } \phi(x) := \begin{cases} \ln(x+1) + 1, & \text{if } x \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then the optimal set $P_* = \{0\}$ and $M_P = 1$. Let $\sigma(z) = |z|$ and a parameterization function f given by

$$f(x,z) := \begin{cases} \ln(x+1) + \exp(-z), & \text{if } x, z \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Using (2) and the definition of f, the Lagrangian becomes

$$l(x,y,c) = \ln(x+1) + \inf_{z \geq 0}\{\exp(-z) + (c-y)z\},$$

for $x \geq 0$. Denote by $\theta(y,c) := \inf_{z \geq 0}\{\exp(-z) + (c-y)z\}$. It is easy to check that

$$q(y,c) = \inf_{x \geq 0} l(x,y,c) = \theta(y,c) = \begin{cases} 1 & \text{if } c - y \geq 1, \\ (c-y)(1 - \ln(c-y)) & \text{if } 1 > c - y > 0, \\ 0 & \text{if } c = y, \\ -\infty & \text{if } c - y < 0. \end{cases}$$

Then $D_* = \{(y,c) \,:\, q(y,c) = 1\} = \{(y,c) \,:\, c \geq y+1\}$. Also $E(y) = \max\{y+1, 0\}$. It is not hard to check that and $A(y, E(y)) = \{(0,0)\}$ for all y. In particular, $P(y, E(y)) = P_*$ and $P_H(y, E(y)) = \{0_H\}$. Therefore E is strongly exact at every $y \in \operatorname{dom}(E)$. In this example q is differentiable at $(y, E(y))$.

Next, we recall [15, Definition 11.60] and relate it with our setting.

Definition 3.4: A vector $y \in H$ is said to *support an exact penalty representation* for Problem (1) if and only if for all c sufficiently large $q(y,c) = M_P$ and $\operatorname{argmin}_{x\in X} l(x,y,c) = P_*$.

Remark 2: The existence of y supporting an exact penalty representation is equivalent to $(y,c) \in D_*$ for some $c \geq 0$. Therefore, in our setting, we have that $y \in H$ supports an exact penalty representation if and only if $y \in \operatorname{dom}(E)$. Moreover, the value $E(y)$ is the threshold for y, that is, the infimum of c such that (y,c) satisfies Definition 3.4.

From now on we make the following assumption on the augmenting function:

$$(\mathbf{A_0}) \qquad \sigma(z) \geq \|z\| \quad \forall\, z \in H.$$

Remark 3: The analysis can easily be carried out for the case $\sigma(z) \geq b\|z\|$ for some $b > 0$. This would, however, change the Lispchitz constant obtained in Section 3.3 for E.

Examples of augmenting functions that satisfy (A_0) are listed below.

i) $\sigma_{p,q}(z) = \begin{cases} \|z\|^p, & \text{if } \|z\| \leq 1, \\ \|z\|^q, & \text{otherwise,} \end{cases}$ with $0 < p \leq 1 \leq q$.

ii) $H = \mathbb{R}^n$, $\sigma_k(z) = (\sum_{i=1}^n |z_i|^{\frac{1}{k}})^k$, with $k \in \mathbb{N}$.

iii) $\sigma_{q,p}(z) = \|z\|^q + \|z\|^p$ for each $z \in H$, with $0 < p \leq 1 \leq q$.

The following technical lemma has a crucial role in relating strong exactness of E with differentiability of the dual function. Denote by $B(y,\varepsilon) := \{z \in H : \|z - y\| < \varepsilon\}$ the open ball of center y and radius ε in H.

Lemma 3.5: *Let $(y, c_y) \in H \times \mathbb{R}_{++}$ be such that (y, c_y) belongs to* $\operatorname{int}(\operatorname{dom} q)$. *Then for each $b \geq M_P$ there exist a weak compact set $B \subset X \times H$ and $\varepsilon > 0$ such that*

$$\emptyset \neq L_b(w,c) := \{(x,z) \in X \times H : \Phi_{(w,c)}(x,z) \leq b\} \subset B \tag{10}$$

for all $(w,c) \in B(y,\varepsilon) \times (c_y - \varepsilon, c_y + \varepsilon)$. In particular, $A(w,c) \neq \emptyset$ for each $(w,c) \in B(y,\varepsilon) \times (c_y - \varepsilon, c_y + \varepsilon)$, that is, there exists some $(\tilde{x},\tilde{z}) \in X \times H$ such that

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

Proof: Take $(y,c_y) \in \operatorname{int}(\operatorname{dom} q)$ and suppose that $B(y,r) \times [c_y - r, c_y + r] \subset \operatorname{dom} q$ for some $r \in (0, c_y)$. Take $\varepsilon := r/2$. Let us show that there exists a bounded set $B \subset X \times H$ such that $L_b(w,c) \subset B$ for all $(w,c) \in B(y,\varepsilon/2) \times [c_y - \varepsilon, c_y + \varepsilon]$. Suppose by contradiction that this is not true. Thus, there exist $\{(y_k,c_k)\} \subset B(y,\varepsilon/2) \times [c_y - \varepsilon, c_y + \varepsilon]$ and a sequence $\{(x_k,z_k)\}$ satisfying $(x_k,z_k) \in L_b(y_k,c_k)$ for all $k \in \mathbb{N}$ and $\lim_k \|(x_k,z_k)\| = \infty$. Let $d := c_y - r > 0$, and recall that $\Phi_{(w,c)}(x,z) := f(x,z) - \langle w,z\rangle + c\sigma(z)$, for $(x,c) \in X \times \mathbb{R}_+$ and $z, w \in H$. It follows that

$$\Phi_{(y_k,c_k)}(x_k,z_k) = f(x_k,z_k) - \langle y_k, z_k\rangle + c_k\sigma(z_k) \leq b \text{ for each } k, \tag{11}$$

which implies

$$
\begin{aligned}
b \geq \Phi_{(y_k,c_k)}(x_k,z_k) &= \Phi_{(y,d)}(x_k,z_k) + \langle y-y_k, z_k\rangle + (c_k-d)\sigma(z_k) \\
&\geq q(y,d) - \|y_k-y\|\|z_k\| + \varepsilon\sigma(z_k) \\
&\geq q(y,d) + \tfrac{\varepsilon}{2}\|z_k\|,
\end{aligned} \tag{12}
$$

where we used the fact that $c_k \geq c_y - \varepsilon = d + \varepsilon$ in the second inequality, and $\sigma(\cdot) \geq \|\cdot\|$ and $y_k \in B(y, \frac{\varepsilon}{2})$ in the last inequality. We obtain from (12) that $\|z_k\| \leq 2[b - q(y,d)]/\varepsilon := a$. Therefore there exists a subsequence $\{z_{k_j}\}$ weakly convergent to some z. Take $\alpha := b + a(\|y\| + \varepsilon)$. By the compactness assumption on the sublevel of f (see Definition 2.2), there exists a weak neighborhood W of z and a weak compact set B such that $L_{f,W}(\alpha) = \{x : f(x,u) \leq \alpha\} \subset B$ for all $u \in W$. In particular, since $\{z_{k_j}\}$ is weakly convergent to z, we obtain $z_{k_j} \in W$ for all j sufficiently large. Using the estimates above in (11) we conclude that $x_{k_j} \in L_{f,W}(\alpha)$ for all j sufficiently large and therefore $\{x_{k_j}\}$ is bounded. Hence $\{(x_{k_j}, z_{k_j})\}$ is bounded, which is a contradiction with the fact that $\lim_k \|(x_k, z_k)\| = \infty$. This completes the proof of (10). Since each function in the expression of $\Phi_{(w,c)}(\cdot,\cdot)$ is w-lsc, the last statement of the lemma follows from (10) and the fact that every w-lsc function attains its minimum on a weakly compact set. □

In the next theorem we establish the equivalence between the differentiability of the dual function q at $(y, E(y))$ and the strong exactness of E. We will use the following fact:

$$
\text{If } z \in P_H(y,c), \text{ then } (-z, \sigma(z)) \in \partial q(y,c). \tag{13}
$$

Recall that the right-most inclusion can be equivalently written as

$$
q(u,d) \leq q(y,c) - \langle u-y, z\rangle + (d-c)\sigma(z),
$$

for every $(u,d) \in H \times \mathbb{R}_+$. We denote by 0_H the null element of the Hilbert space H.

Theorem 3.6: *Let $y \in \operatorname{dom} E$ be such that $(y, E(y)) \in \operatorname{int}(\operatorname{dom} q)$. Then the dual function q is differentiable at $(y, E(y))$ iff E is strongly exact at y.*

Proof: Suppose that q is differentiable at $(y, E(y))$. Let us prove that $P_H(y, E(y)) = \{0\}$. Since $(y, E(y)) \in D_*$ we obtain from Proposition 2.6 (a) that $0_H \in P_H(y, E(y))$. In order to show that 0_H is the only element in $P_H(y, E(y))$, take an arbitrary $z \in P_H(y, E(y))$. Thus there exists $x \in X$ such that $(x,z) \in A(y, E(y))$. By (13) $(-z, \sigma(z)) \in \partial q(y, E(y))$. On the other hand, $(0_H, 0) \in \partial q(y, E(y))$ because $(y, E(y))$ maximizes the concave function q. Since q is differentiable at $(y, E(y))$, we conclude $\partial q(y, E(y)) = \{(0_H, 0)\} \subset H \times \mathbb{R}$. Therefore $z = 0_H$ and then $P_H(y, E(y)) = \{0_H\}$. In order to prove the converse statement, suppose that E is strongly exact at y. The differentiability of q at $(y, E(y))$ will follow from Lemma 3.5 and a well known representation formula for the subdifferential of a convex function given by the maximum of convex functions. Indeed, by Lemma 3.5, there exists a weak compact set B such that

$$
\emptyset \neq \{(x,z) : \Phi_{(w,c)}(x,z) \leq M_P\} \subset B,
$$

for all $(w,c) \in B(y,\varepsilon) \times (E(y) - \varepsilon, E(y) + \varepsilon)$ and some $0 < \varepsilon < E(y)$. It follows that

$$q(w,c) = \min\{\Phi_{(w,c)}(x,z) : (x,z) \in B\}.$$

We observe that for each $(x,z) \in B$ the function $\Phi_{(w,c)}(x,z)$ (as function of (w,c)) is an affine function and its derivative at (w,c) is $(-z, \sigma(z))$. Therefore, we obtain from [17, Proposition 4.5.2] that the superdifferential of q at $(y, E(y))$ is given by

$$\partial q(y, E(y)) = \overline{\text{co}}^w\{(-z, \sigma(z)) : (x,z) \in A(y, E(y)) \text{ for some } x \in X\},$$

where $\overline{\text{co}}^w(Q)$ denotes the weak closure of the set $Q \subset H \times \mathbb{R}$. Since $\{0_H\} = P_H(y, E(y)) = \{z : (x,z) \in A(y, E(y)) \quad \text{for some} \quad x \in X\}$, we must have $(-z, \sigma(z)) = (0_H, 0)$ for every $(x,z) \in A(y, E(y))$. This fact, together with the above expression now yields

$$\partial q(y, E(y)) = \overline{\text{co}}^w\{(0_H, 0)\} = \{(0,0)\}.$$

Therefore q is differentiable at $(y, E(y))$. This completes the proof. □

3.2. *Dual Localization results*

The function $\Phi_{(y,c)}(x,z)$ (defined in (4)) is crucial for analyzing the map E and the dual solution set. In particular, the following estimate will be useful. For each $y, w \in H$ and $c \in \mathbb{R}_+$,

$$\Phi_{(y,c)}(x,z) \leq \Phi_{(w,c+\|w-y\|)}(x,z), \text{ for all } \quad (x,z) \in X \times H. \tag{14}$$

Indeed,

$$\begin{aligned}\Phi_{(y,c)}(x,z) &= f(x,z) - \langle y, z\rangle + c\sigma(z) \\ &= f(x,z) - \langle w, z\rangle + (c + \|w-y\|)\sigma(z) + \langle w - y, z\rangle \\ &\quad -\|w-y\|\sigma(z),\end{aligned}$$

now using the definition of Φ and Cauchy-Schwarz inequality we obtain

$$\Phi_{(y,c)}(x,z) \leq \Phi_{(w,c+\|w-y\|)}(x,z) + \|w-y\|(\|z\| - \sigma(z)),$$

and the result follows using (A_0).

Proposition 3.7 and its corollary can be seen as a "localization" result for the dual solution set.

Proposition 3.7: *Take $(y,c) \in D_*$. Then $(y+h, c+\|h\|) \in D_*$ for all $h \in H$.*

Proof: Fix $(y,c) \in D_*$, and take $h = w - y$, then we have

$$M_D = q(y,c) = \inf_{(x,z)} \Phi_{(y,c)}(x,z) \leq \inf_{(x,z)} \Phi_{(w,c+\|w-y\|)}(x,z) = q(w, c + \|w-y\|) \leq M_D,$$

in view of (14). Use $(w, c+\|w-y\|) = (y+h, c+\|h\|)$ for completing the proof. □

Corollary 3.8: *Take $(y,c) \in D_*$, and $\rho > 0$. Then*

$$\Delta_\rho := \{(y+h, c+\rho) \in H \times \mathbb{R}_+ \ : \ \|h\| \leq \rho\} \subset D_*.$$

Proof: The result follows directly from Proposition 3.7 and Proposition 2.3 (ii). □

We can use our localization results for connecting properties of dual sequences with the graph of E.

Corollary 3.9: *Consider a sequence $\{(u_k, t_k)\}$ converging strongly to some $(\bar{u}, \bar{t}) \in D_*$. Assume that $P_H(u_k, t_k) \neq \{0\}$ for all k. Then $\bar{t} = E(\bar{u})$.*

Proof: Take $d := \bar{t} - E(\bar{u})$. Since $(\bar{u}, \bar{t}) \in D_*$, $d \geq 0$ by definition of E. We must show that $d = 0$. Suppose for the sake of contradiction that $d > 0$. By Proposition 3.7 applied to $(y, c) := (\bar{u}, E(\bar{u})) \in D_*$ and $h := u - \bar{u}$ for each $u \in H$, we have that $(u, t) \in D_*$ for $t \geq E(\bar{u}) + \|u - \bar{u}\|$. Take $\bar{k}$ such that $\|u_k - \bar{u}\| \leq d/3$ and $t_k \geq \bar{t} - d/3$ for all $k \geq \bar{k}$. It follows that

$$t_k \geq \bar{t} - \frac{d}{3} = E(\bar{u}) + \frac{2d}{3} \geq E(\bar{u}) + \|u_k - \bar{u}\| + \frac{d}{3}.$$

Therefore, Propositions 3.7 and 2.6(b) imply that $(u_k, t_k) \in D_*$ and $P_H(u_k, t_k) = \{0\}$ for all $k \geq \bar{k}$, which is a contradiction with the assumption. Thus $d = 0$ as claimed. □

3.3. *Properties of E: more connections with D_**

Define the map $T : H \times H \to \mathbb{R}_+ \cup \{+\infty\}$ as

$$T(y, w) := E(y) + \|w - y\|.$$

We already mentioned the fact that E is a convex function. It follows that $T(\cdot, w)$ is a convex function with $\operatorname{dom}(T) = \operatorname{dom}(E) \times H$ and $T(y, y) = E(y)$ for all $y \in H$. The function T is useful for proving that E is a Lipschitz continuous mapping. First we use T to characterize the set of $t \geq 0$ such that $(0, t) \in D_*$.

Corollary 3.10: *The following statements ae equivalent.*

(i) D_* *is nonempty,*
(ii) $\operatorname{dom}(E) = H$,
(iii) $\operatorname{dom}(E) \neq \emptyset$.

In this situation, for each $y \in H$ we have that $(0, t) \in D_*$ for all $t \geq T(y, 0)$.

Proof: (ii) implies (iii) and (iii) implies (i) are trivial from the definitions. We proceed to prove (i) implies (ii). Fix $(y', c') \in D_*$ and $y \in H$. Using Proposition 3.7 with reference point $(y', c') \in D_*$ and $h := y - y'$ we conclude that $(y' + h, c' + \|h\|) = (y, c' + \|h\|) \in D_*$. This yields $E(y) \leq c' + \|h\|$ and therefore $y \in \operatorname{dom} E$. Since y is arbitrary, $\operatorname{dom} E = H$. We have proved that (i) implies (ii). In order to prove the last statement, take any $0 \neq y \in H = \operatorname{dom}(E)$. Inasmuch $(y, E(y)) \in D_*$, Corollary 3.8 for $h = -y$ yields

$$(y + h, E(y) + \|h\|) = (0, T(y, 0)) \in D_*,$$

using the definition of T. This fact and Proposition 2.3(iii) complete the proof. □

Remark 4: In view of Corollary 3.10 and Definition 3.4, a point $y \in H$ supports an exact penalty representation if and only if $D_* \neq \emptyset$.

We are now in conditions to show that whenever E is proper, it is globally Lipschitz with constant at most 1. Examples 3.2 and 3.3 show that the maximum constant value 1 may be attained.

Corollary 3.11: *If $y \in \operatorname{dom} E$ then $(w, T(y,w)) \in D_*$, $\forall\, w \in H$. Consequently, $E(w) \le T(y,w)$ for all $y, w \in H$. In this situation, E is a Lipschitz continuous mapping in H with Lipschitz constant at most 1.*

Proof: Fix $y \in \operatorname{dom} E$ and $w \in H$. Then $(y, E(y)) \in D_*$, and using Proposition 3.7 for $h = w - y$ and reference point $(y, E(y)) \in D_*$, we obtain

$$(w, T(y,w)) = (w, E(y) + \|w - y\|) \in D_*.$$

Hence $E(w) \le T(y,w)$ and therefore $w \in \operatorname{dom}(E)$. This yields $\operatorname{dom}(E) = H$. The latter inequality and the definition of T yield

$$E(w) - E(y) \le \|w - y\|.$$

Interchanging the roles of y by w we obtain $E(y) - E(w) \le \|y - w\|$. Therefore $|E(y) - E(w)| \le \|y - w\|$ for any $y, w \in \operatorname{dom}(E) = H$, completing the proof. □

The following result can be seen as a characterization of exact penalty properties in terms of the mapping T. More precisely, E verifies an exact penalty property at some $y \in H$ if and only if $T(\cdot, y)$ verifies an exact penalty property at every element of H.

Proposition 3.12: *Assume that $D_* \neq \emptyset$. Then there exists $y \in H$ such that $P_X(y, E(y)) = P_*$ if and only if $P_X(w, T(w,y)) = P_*$ for every $w \in H$,.*

Proof: Note that Corollary 3.10 yields $\operatorname{dom}(E) = H$. Assume first that $P_X(w, T(y,w)) = P_*$ for every $w \in H$. Choosing $w := y$ and using that $T(y,y) = E(y)$ we obtain $P_X(y, E(y)) = P_*$. In order to prove the converse statement, assume that for some $y \in H$ it holds that $P_X(y, E(y)) = P_*$. Corollary 3.11 yields $(w, T(y,w)) \in D_*$ for every $w \in H$. In particular, we obtain that $P_* \subset P_X(w, T(y,w))$, by Proposition 2.6 (a). It remains to prove that $P_X(w, T(y,w)) \subset P_*$. Let $d := T(y,w)$ and $x_w \in P_X(w,d)$. Hence, there exists $z_w \in H$ such that $(x_w, z_w) \in A(w,d)$. Let $c := E(y)$ and $x_y \in P_*$. By strong duality (Proposition 2.3 (iii)) and inequality (14) we obtain

$$M_D = \varphi(x_y) = q(y,c) \le \Phi_{(y,c)}(x_w, z_w) \le \Phi_{(w, c+\|w-y\|)}(x_w, z_w) = q(w,d) \le M_D, \tag{15}$$

where the last equality follows from the fact that $(x_w, z_w) \in A(w,d)$ and $d = c + \|w - y\|$. Therefore all the inequalities in (15) are equalities, which implies that

$$q(y,c) = \Phi_{(y,c)}(x_w, z_w) = f(x_w, z_w) - \langle z_w, y \rangle + c\sigma(z_w),$$

that is, $(x_w, z_w) \in A(y,c)$ and thus $x_w \in P_X(y,c)$. By assumption $P_X(y,c) = P_X(y, E(y)) = P_*$, so that $x_w \in P_*$. Since x_w is an arbitrary element in $P_X(w,d)$, we obtain that $P_X(w,d) \subset P_*$, concluding the proof. □

4. The Penalty Map and Primal-Dual methods

In this section we relate the previous results with convergence properties of primal-dual methods. We mentioned in the introduction that key properties of primal-dual methods are *primal convergence* and *finite termination*, which we define formally below.

Definition 4.1: A primal-dual method induced by A has *primal convergence* if and only if for every sequence $\{(x_k, z_k) \in A(y_k, c_k)\}$, it holds that

(i) $\{x_k\} \subset X$ is bounded, and
(ii) every weak accummulation point of $\{x_k\}$ belongs to P_*.

We say that a primal-dual method induced by A has *finite termination* if and only if it has *primal convergence* and the sequence $\{x_k\}$ stops (i.e., it becomes constant) after a finite number of iterations.

The results of this section are devoted to establish conditions under which a primal-dual method induced by A has primal convergence or finite termination. We also relate this properties with the penaly map E whenever possible.

Remark 1: The condition of strong exactness provides a stopping criterion for the primal-dual method A. Indeed, if for some k we have that $P_H(y_k, c_k) = \{0_H\}$, then for every $(x_k, z_k) \in A(y_k, c_k)$ we will have $x_k \in P_*$ (by Proposition 2.6(c)). Therefore, the method may stop here. With this in mind, it may be reasonable to stop the method A at a point $(x_k, z_k) \in A(y_k, c_k)$ such that $z_k = 0$. This is the situation for DSG, as we see in the next section. If A has this stopping criteron and generates an infinite sequence $\{(y_k, c_k)\}$ converging to some $(y, c) \in D_*$, then Corollary 3.9 yields $c = E(y)$. If E happens to be *exact* at y, Proposition 2.6(c) yields primal convergence of A. This is why exactness of E is a crucial property for primal-dual methods.

The next proposition establishes assumptions under which A has primal convergence.

Proposition 4.2: *Consider a primal-dual method induced by A, i.e., a sequence such that $\{(x_k, z_k) \in A(y_k, c_k)\}$. Assume that:*

(i) $\{(y_k, c_k)\}$ *converges strongly to* $(\bar{y}, \bar{c}) \in D_*$,
(ii) $\{z_k\}$ *has weak accummulation points.*

In this situation, the following hold:

(a) If $P_X(\bar{y}, \bar{c}) = P_*$ then every weak accumulation point (if any) of $\{x_k\}$ is a primal solution.
(b) Every weak accummulation point of $\{z_k\}$ belongs to $P_H(\bar{y}, \bar{c})$.

Proof: (a) Let z be a weak accummulation point of $\{z_k\}$ and $\bar{x}$ a weak accummulation point of $\{x_k\}$. Assume that $\{z_{k_j}\}$ and $\{x_{k_j}\}$ are the corresponding weakly convergent subsequences. Under the assumptions of the proposition, it follows that $\langle y_{k_j}, z_{k_j}\rangle$ converges to $\langle \bar{y}, z\rangle$. Using now the lower semicontinuity of both f (see Remark **1**) and σ (see Definition 2.1), we obtain

$$\begin{aligned} M_D = q(\bar{y}, \bar{c}) &\le f(\bar{x}, z) - \langle \bar{y}, z\rangle + \bar{c}\sigma(z) \\ &\le \liminf_j f(x_{k_j}, z_{k_j}) - \langle y_{k_j}, z_{k_j}\rangle + c_{k_j}\sigma(z_{k_j}) \\ &= \liminf_j q(y_{k_j}, c_{k_j}) \le M_D, \end{aligned}$$

using the fact that $(x_{k_j}, z_{k_j}) \in A(y_{k_j}, c_{k_j})$ in the second equality. The above expression yields $(\bar{x}, z) \in A(\bar{y}, \bar{c})$ and hence $\bar{x} \in P_X(\bar{y}, \bar{c})$ and $z \in P_H(\bar{y}, \bar{c})$. Thus, (b) is proved. If $P_* = P_X(\bar{y}, \bar{c})$, then $\bar{x} \in P_*$, which completes the proof of (a). □

Remark 2: The conclusions of Proposition 4.2 hold under any assumption ensuring that $\langle y_k, z_k \rangle$ converges to some $\langle y, z \rangle$. We will show in Section 5 that the deflected subgradient algorithm satisfies assumption (i) of Proposition 4.2.

Some primal-dual methods, like the one considered in [2], may not have primal convergence. To overcome this, the authors of [2] introduced an auxiliary sequence $\{\tilde{x}_k\}$, and proved that it converges to a primal solution. We consider the sequence $\{\tilde{x}_k\}$ in the next proposition, and prove that it has the stronger property of finite convergence, in a much more general setting than DSG. The finite convergence holds because for k sufficiently large, $E(y_k)$ is strictly less than the "penalty parameter" $c_k + \beta$, $\beta > 0$.

Proposition 4.3: *Suppose that D_* is nonempty, and assume that A generates an infinite dual sequence $\{(y_k, c_k)\}$ strongly convergent to a dual solution $(\bar{y}, \bar{c})$. Fix $\beta > 0$ and consider an auxiliary sequence $\{\tilde{x}_k\}$ such that $\tilde{x}_k \in P_X(y_k, c_k + \beta)$ for each $k \geq 0$. Then there exists $\bar{k}$ such that $P_H(y_{\bar{k}}, c_{\bar{k}} + \beta) = \{0_H\}$. In particular $\tilde{x}_{\bar{k}}$ is a primal solution.*

Proof: Suppose by contradiction that the conclusion is not true, that is, $P_H(y_k, c_k + \beta) \neq \{0_H\}$ for each $k \geq 0$. Let $\{(u_k, t_k)\}$ be defined by $(u_k, t_k) := (y_k, c_k + \beta)$, which converges to $(\bar{y}, \bar{c} + \beta) \in D_*$. By Corollary 3.9 we obtain

$$E(\bar{y}) = \bar{c} + \beta > \bar{c}. \tag{16}$$

On the other hand, since $(\bar{y}, \bar{c}) \in D_*$ we have $E(\bar{y}) \leq \bar{c}$ by definition of E, which is a contradiction with (16). Therefore there exists $\bar{k}$ such that $P_H(y_{\bar{k}}, c_{\bar{k}} + \beta) = \{0_H\}$. The last statement follows from the latter fact and Proposition 2.6(c). □

The next proposition establishes primal convergence and finite termination when $\{\sigma(z_{k_j})\}$ converges to zero.

Theorem 4.4: *Let $\{(x_k, z_k) \in A(y_k, c_k)\}$ be a sequence generated by a primal-dual method A which stops when $z_k = 0_H$.*

(a) Assume that the algorithm never stops and the following statements hold:
 (i) $\{(y_k, c_k)\}$ converges strongly to $(\bar{y}, \bar{c}) \in D_*$, and
 (ii) there exists a subsequence $\{\sigma(z_{k_j})\}$ converging to zero.
 Then A has primal convergence and $\{z_{k_j}\}$ converges strongly to 0_H.

(b) If the algorithm stops, i.e., there exists $\hat{k}$ such that $z_{\hat{k}} = 0_H$, then $x_{\hat{k}} \in P_*$ and $(y_{\hat{k}}, c_{\hat{k}}) \in D_*$. In other words, A has finite termination.

Proof: (a) First, note that $\bar{c} = E(\bar{y})$. Indeed, this follows from Corollary 3.9, the stopping criterion, and the fact that A generates an infinite sequence. Second, note that (ii) implies that $\{z_{k_j}\}$ is bounded. Indeed, if the latter is not true, then there exists a subsequence of $\{z_{k_j}\}$ (which we still call $\{z_{k_j}\}$ for convenience) such that $\|z_{k_j}\| \geq 1$. By Definition 2.1 there exists $\delta_1 > 0$ such that

$$0 < \delta_1 \leq \inf_{\|z\| \geq 1} \sigma(z) \leq \inf_j \sigma(z_{k_j}) = 0,$$

using (ii) in the equality. Since the above expression entails a contradiction, $\{z_{k_j}\}$ is bounded. We claim now that $\{z_{k_j}\}$ converges strongly to 0_H. Indeed, let $\hat{z}$ be a weak accummulation point of $\{z_{k_j}\}$. For convenience we still denote by $\{z_{k_j}\}$ a weakly convergent subsequence to $\hat{z}$. From (ii) and assumption (A_0), we obtain

$$0 = \lim_{j \to \infty} \sigma(z_{k_j}) \geq \lim_{j \to \infty} \|z_{k_j}\| \geq 0.$$

Use now the weak lower semicontinuity of σ and (A_0) to obtain

$$0 = \lim_{j\to\infty} \sigma(z_{k_j}) \geq \sigma(\hat{z}) \geq \|\hat{z}\| \geq 0.$$

and hence $\hat{z} = 0$. The combination of these two facts implies that $\{z_{k_j}\}$ converges strongly to 0_H. By definition of A, we have that for every j the following holds:

$$\begin{aligned} f(x_{k_j}, z_{k_j}) &= q(y_{k_j}, c_{k_j}) + \langle y_{k_j}, z_{k_j}\rangle - c_{k_j}\sigma(z_{k_j}) \\ &\leq q(y_{k_j}, c_{k_j}) + \|y_{k_j}\|\,\|z_{k_j}\|, \end{aligned}$$

Taking limsup for $j \to \infty$ and using upper semicontinuity of q, (i) and the fact that $\{z_{k_j}\}$ converges strongly to 0_H, we obtain

$$\limsup_{j\to\infty} f(x_{k_j}, z_{k_j}) \leq M_D. \tag{17}$$

We claim now that $\{x_{k_j}\}$ is has weak accummulation points. Indeed, by Definition 2.2 there exists a weak neighborhood V of 0_H such that the set $L_{V,f}(M_D+1) := \{x \in X : f(x,z) \leq M_D + 1\}$ is weakly compact for every $z \in V$. Inasmuch $\{z_{k_j}\}$ converges strongly to 0_H (in fact, weak convergence would suffice here), there exists j_0 such that $z_{k_{j_0}} \in V$ for every $j \geq j_0$. Equation (17) now yields

$$\{x_{k_j}\}_{j\geq j_0} \subset L_{V,f}(M_D + 1).$$

The weak compactness implies the existence of a weak accummulation point $\bar{x} \in X$ of the sequence $\{x_{k_j}\}$. For simplicity we still denote by $\{x_{k_j}\}$ a subsequence weakly convergent to $\bar{x}$. Combine the latter fact with the w-lsc of f, the fact that $\{z_{k_j}\}$ converges weakly to 0_H, and (17), for concluding that

$$M_P = M_D \geq \limsup_{j\to\infty} f(x_{k_j}, z_{k_j}) \geq \liminf_{j\to\infty} f(x_{k_j}, z_{k_j}) \geq f(\bar{x}, 0) = \varphi(\bar{x}) \geq M_P, \tag{18}$$

which implies that A has primal convergence. This completes the proof of part (a). In order to prove (b), note that A stops at iteration $\hat{k}$ if and only if $z_{\hat{k}} = 0$. For $k = \hat{k}$ we have, by definition of A,

$$(x_{\hat{k}}, z_{\hat{k}}) = (x_{\hat{k}}, 0) \in A(y_{\hat{k}}, c_{\hat{k}}),$$

which, as in (a), yields

$$M_D \geq q(y_{\hat{k}}, c_{\hat{k}}) = f(x_{\hat{k}}, 0) = \varphi(x_{\hat{k}}) \geq M_P.$$

Use now strong duality to conclude that $x_{\hat{k}} \in P_*$ and $(y_{\hat{k}}, c_{\hat{k}}) \in D_*$. This completes the proof of (b). □

Remark 3 : From the proof of Theorem 4.4(a) we conclude that if $\{z_k\}$ is weakly convergent to zero and $\{\sigma(z_k)\}$ has 0 as an accummulation point, then $\{z_k\}$ is strongly convergent to zero.

5. Application to Deflected Subgradient Methods

We mentioned before that DSG is an example of a primal-dual method. DSG is defined as follows: given a current dual iterate $(u_k, c_k) \in H \times \mathbb{R}_+$, find $(x_k, z_k) \in$

$A(u_k, c_k)$; the next the dual iterate is defined as

$$\begin{aligned} y_{k+1} &:= y_k - s_k z_k, \\ c_{k+1} &:= c_k + (\epsilon_k + s_k)\sigma(z_k), \end{aligned} \tag{19}$$

where $\epsilon_k, s_k > 0$ are positive stepsizes. Note that the above update indicates that DSG stops when $z_k = 0_H$. Therefore the results of the previous section, namely, Theorem 4.4 can be applied to analyze DSG.

The DSG (studied, e.g., in [2, 3, 5–9, 12]) generates a sequence of dual values $\{q(u_k, c_k)\}$ which is strictly increasing and converges to the dual optimal value. According to the choice of s_k, DSG may have primal convergence or finite termination (see [5]). As in the convergence analysis of [5], we take the sequence of parameters $\{\epsilon_k\}$ in (19) as

$$\epsilon_k := \alpha_k s_k, \text{ where } \alpha := \inf_k \alpha_k > 0, \tag{20}$$

In [5], the authors prove that the dual sequence converges weakly to a dual solution. This sequence, in fact, converges strongly, as we prove next.

Theorem 5.1: *If $D_* \neq \emptyset$, then the sequence $\{(y_k, c_k)\}$ generated by DSG is strongly convergent to a dual solution.*

Proof: Using the update formulas $y_{k+1} = y_k - s_k z_k$ and $c_{k+1} = c_k + (1 + \alpha_k)s_k\sigma(z_k)$, it follows that

$$\|y_{k+j} - y_k\| \le \sum_{l=k}^{k+j-1} \|y_{l+1} - y_l\| = \sum_{l=k}^{k+j-1} s_l\|z_l\| \le \sum_{l=k}^{k+j-1} s_l\sigma(z_l).$$

$$c_{k+j} - c_k = \sum_{l=k}^{k+j-1} c_{l+1} - c_l = \sum_{l=k}^{k+j-1} (\alpha_l + 1)s_l\sigma(z_l).$$

As a consequence of theses estimates we obtain for all $k, j \in \mathbb{N}$,

$$c_{k+j} - c_k \ge (1 + \alpha)\|y_{k+j} - y_k\|. \tag{21}$$

By [5, Theorem 3.15], $\{(y_k, c_k)\}$ is weakly convergent to a dual solution. Therefore $\{c_k\} \subset \mathbb{R}_+$ is convergent. In particular $\{c_k\}$ is a Cauchy sequence. The estimate (21) implies that $\{y_k\}$ is also a Cauchy sequence and hence strongly convergent, because H is a Hilbert space. Therefore $\{(y_k, c_k)\}$ is strongly convergent to a dual solution. □

The next proposition relates the exactness of E and the primal convergence of DSG.

Proposition 5.2: *Let $\{(y_k, c_k)\}$ be an infinite sequence generated by DSG (i.e., according to* (19)-(20)*). Assume that*

(i) $\{(y_k, c_k)\}$ *converges strongly to* $(\bar{y}, \bar{c}) \in D_*$, *with* $\bar{c} \ge c_k$ *for all* k, *and*
(ii) there exists a subsequence $\{\sigma(z_{k_j})\}$ *converging to zero.*

Then, the exact penalty map E is strongly exact at $\bar{y}$.

Proof: From Theorem 4.4 and Corollary 3.9 we know that $\{z_{k_j}\}$ converges strongly to zero and $E(\bar{y}) = \bar{c}$. Assume that E is not strongly exact at $\bar{y}$, so that there exists $0 \neq \bar{z} \in P_H(\bar{y}, E(\bar{y}))$. In particular, $\sigma(\bar{z}) > 0$. By antimonotonicity of ∂q we have

$$\langle(-\bar{z}, \sigma(\bar{z})) - (-z_{k_j}, \sigma(z_{k_j})), (\bar{y}, \bar{c}) - (y_{k_j}, c_{k_j})\rangle \leq 0,$$

which is equivalent to

$$\langle z_{k_j} - \bar{z}, \bar{y} - y_{k_j}\rangle + (\sigma(\bar{z}) - \sigma(z_{k_j}))(\bar{c} - c_{k_j}) \leq 0.$$

Using Cauchy Schwarz inequality and re-arranging the resulting expression we obtain

$$(\sigma(\bar{z}) - \sigma(z_{k_j}))(\bar{c} - c_{k_j}) \leq \|\bar{z} - z_{k_j}\|\|\bar{y} - y_{k_j}\|.$$

The estimate (21) implies that $\|\bar{y} - y_{k_j}\| \leq (1+\alpha)^{-1}(\bar{c} - c_{k_j})$. Therefore

$$(\sigma(\bar{z}) - \sigma(z_{k_j}))(\bar{c} - c_{k_j}) \leq \|\bar{z} - z_{k_j}\|(1+\alpha)^{-1}(\bar{c} - c_{k_j}).$$

Since DSG generates an infinite sequence and $z_{k_j} \neq 0$ for all j, we must have $0 \leq c_{k_j} < \bar{c}$ for all j. Hence, the above expression simplifies to

$$(\sigma(\bar{z}) - \sigma(z_{k_j})) \leq \|\bar{z} - z_{k_j}\|(1+\alpha)^{-1}.$$

Letting now $j \to \infty$ and using (A_0) we obtain

$$\|\bar{z}\| \leq \sigma(\bar{z}) \leq \|\bar{z}\|(1+\alpha)^{-1} < \|\bar{z}\|,$$

which is a contradiction. The proof is complete. □

The following result characterizes convergence of DSG in terms of the map E.

Corollary 5.3: *Let $\{(x_k, z_k)\}$ and $\{(y_k, c_k)\}$ be bounded sequences generated by* DSG. *Suppose that $\{(y_k, c_k)\}$ converges strongly to some $(\bar{y}, \bar{c}) \in D_* \cap \operatorname{int}(\operatorname{dom} q)$. The following statements are equivalent:*

a) There sequence $\{\sigma(z_k)\}$ converges to 0;
b) the dual function q is differentiable at $(\bar{y}, \bar{c})$;
c) the penalty map E is strongly exact at $\bar{y}$.

Moreover, under any of these statements, the sequence $\{z_k\}$ converges strongly to 0 and every accumulation point of $\{x_k\}$ is a primal solution.

Proof: The last statement follows from (a), assumption (A_0), or from Theorem 4.4(a). Note also that the assumption $(\bar{y}, \bar{c}) \in \operatorname{int}(\operatorname{dom} q)$ implies $\bar{c} > 0$, since $\operatorname{dom} q \subset H \times \mathbb{R}_+$. We proceed to prove the equivalences. Theorem 3.6 entails the equivalences between (b) and (c). Proposition 5.2 shows that (a) implies (c). Hence, it is enough to show that (c) implies (a). From Proposition 4.2, we know that every weak accumulation point of $\{z_k\}$ belongs to $P_H(\bar{y}, \bar{c})$. Since we are assuming (c), we have that $P_H(\bar{y}, \bar{c}) = \{0\}$ and thus we obtain that $\{z_k\}$ converges weakly to 0. We observe that since σ is just w-lsc (our prototype of augmenting function is $\sigma(\cdot) = \|\cdot\|$, which is w-lsc but not weakly continuous), we cannot conclude yet that (a) holds. Consider the following sequences

$$m_k := f(x_k, z_k) - \langle y_k, z_k\rangle \text{ and } q_k := m_k + c_k\sigma(z_k), \quad \text{for all } k.$$

Since $\{y_k\}$ is strongly convergent to $\bar{y}$ and $\{z_k\}$ is weakly convergent to 0, we have $\{\langle y_k, z_k\rangle\}$ converges to 0. We know that $\{q_k\}$ converges to M_D ($= M_P$, the optimal value). We claim that $\{m_k\}$ also converges to M_D. Fix a subsequence $\{m_{k_j}\}$ convergent to $m := \liminf_k m_k = \sup_n \inf_{k\geq n} m_k$. We claim that m is finite. Indeed,

$$\begin{aligned} m &= \liminf_k m_k \leq \limsup_k m_k = \limsup_k f(x_k, z_k) - \langle y_k, z_k\rangle \\ &= \limsup_k q_k - c_k\sigma(z_k) \leq \limsup_k q_k \leq M_P, \end{aligned}$$

where we used usc of q and strong duality in the last inequality. Take a subsequence $\{x_{k_{j_n}}\}$ weakly convergent to some $\hat{x}$. In particular $\{m_{k_{j_n}}\}$ converges to m. From weak lower semicontinuity of f and the fact that $f(\cdot, 0) = \varphi(\cdot)$ we have

$$\begin{aligned} M_P \leq \varphi(\hat{x}) &\leq \liminf_n f(x_{k_{j_n}}, z_{k_{j_n}}) - \langle y_{k_{j_n}}, z_{k_{j_n}}\rangle = \lim_n m_{k_{j_n}} \\ &= \liminf_k m_k \leq \limsup_k m_k \leq \limsup_k q_k = M_P. \end{aligned}$$

Therefore $\{m_k\}$ converges to $m = M_P$ and hence our claim holds. Since $\{q_k\}$ also converges to M_P and $\bar{c} > 0$, we obtain

$$0 \leq \lim_{k\to\infty} \sigma(z_k) = \lim_{k\to\infty} \frac{q_k - m_k}{c_k} = \frac{M_P - M_P}{\bar{c}} = 0.$$

That is, $\{\sigma(z_k)\}$ converges to 0, and then (a) holds. The proof is complete. □

6. Equality Constrained Problems

Consider the following equality constrained problem

$$\min\ \psi(x)\ \text{ s.t. } x \in K, h(x) = 0, \tag{22}$$

where $h : X \to H$ is weak-to-weak continuous, $\psi : X \to \mathbb{R}$ is w-lsc, and $K \subset X$ is weakly compact. Consider the following equivalent extended real-valued problem:

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

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

$$f(x,z) = \begin{cases} \psi(x) & \text{if } x \in K \text{ and } z = h(x), \\ \infty, & \text{otherwise.} \end{cases} \tag{23}$$

In this case, we obtain

$$\Phi_{(y,c)}(x,z) = \begin{cases} \psi(x) - \langle y, h(x)\rangle + c\sigma(h(x)) & \text{if } x \in K \text{ and } z = h(x) \\ \infty & \text{otherwise,} \end{cases} \tag{24}$$

which yields the following explicit formula for the augmented Lagrangian function

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

The dual function induced by this augmented Lagrangian is

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

and the dual problem is

$$\max \; q(y,c) \; \text{ s.t. } \; (y,c) \in H \times \mathbb{R}_+. \tag{25}$$

It is not difficult to see that, under the assumptions of this section, the canonical dualizing parameterization function defined in (23) is weakly level-compact. Defining $M(y,c)$ as $M(y,c) = \{x \in K : \ell(x,y,c) = q(y,c)\}$, it follows from (24) that the set $A(y,c)$ defined in Section 2 becomes

$$A(y,c) = \{(x, h(x)) : x \in M(y,c)\}.$$

Hence,

$$P_X(y,c) = M(y,c) \text{ and } P_H(y,c) = \{h(x) : x \in M(y,c)\}. \tag{26}$$

Remark 1 : In this setting, $P_H(y,c) = \{0\} \Leftrightarrow P_X(y,c) = P_*$, that is, the dual penalty map E is exact if and only if it is strongly exact. Indeed, in view of Proposition 2.6(c), if E is strongly exact then E is exact. Assume now that E is exact, that is, $P_X(y, E(y)) = P_*$. Let $z \in P_H(y, E(y))$. Using (26) we have that $z = h(x)$, for some $x \in M(y, E(y)) = P_X(y, E(y)) = P_*$. Since $x \in P_*$, it follows that $h(x) = 0$, which implies that $z = 0$. Since z is an arbitrary element in $P_H(y, E(y))$, it follows that $P_H(y, E(y)) = \{0\}$, that is, E is strongly exact.

Combining Remark **1** with Corollary 5.3 and other previous results we obtain the following theorem.

Theorem 6.1 : *Consider the sequences $\{(y_k, c_k)\}$ and $\{x_k\}$ generated by* DSG *applied to the dual problem* (25)*. Suppose that $\{(y_k, c_k)\}$ converges strongly to some $(\bar{y}, \bar{c}) \in D_* \cap \operatorname{int}(\operatorname{dom} q)$. Then the following statements are equivalent:*

a) The dual function q is differentiable at $(\bar{y}, \bar{c})$;
b) The dual penalty map E is exact at $\bar{y}$;
c) The sequence $\{h(x_k)\}$ converges strongly to 0*;*
d) Every weak accumulation point of $\{x_k\}$ is a primal solution.

Proof: Under the hypotheses of the theorem, Corollary 3.9 yields that $E(\bar{y}) = \bar{c} > 0$. The sequence $\{x_k\}$ has weak accummulation points because it is contained in the weakly compact set K. From Remark **1** and Theorem 3.6 we conclude the equivalence between (a) and (b). Let us prove now that (b) implies (c). From Remark **1** and part "(c) implies (a)" of Corollary 5.3, we have that $\{\sigma(z_k) = \sigma(h(x_k))\}$ converges to 0. Combining this fact with assumption (A_0) we obtain (c). Let us prove now that (c) implies (d). Let $\hat{x}$ be a weak accummulation of $\{x_k\}$ and still denote by $\{x_k\}$ a subsequence weakly convergent to $\hat{x}$. Use the fact that

$\{h(x_k)\}$ converges strongly to zero to conclude that

$$\begin{aligned} M_P &\geq \limsup_k q(y_k, c_k) = \limsup_k f(x_k, h(x_k)) - \langle y_k, h(x_k)\rangle + c_k\sigma(h(x_k)) \\ &\geq \liminf_k f(x_k, h(x_k)) - \langle y_k, h(x_k)\rangle + c_k\sigma(h(x_k)) \\ &\geq \liminf_k f(x_k, h(x_k)) - \langle y_k, h(x_k)\rangle + \liminf_k c_k\sigma(h(x_k)) \\ &\geq f(\hat{x}, 0) = \varphi(\hat{x}) \geq M_P, \end{aligned}$$

and hence $\hat{x} \in P_*$, as wanted. Now we only need to prove (d) implies (a). Assume that (d) holds. We show first that (d) implies that $\{h(x_k)\}$ converges weakly to zero. Indeed, since $\{x_k\}$ has weak accummulation points and h is weak-to-weak continuous, the sequence $\{h(x_k)\}$ has weak accummulation points. We claim that 0_H is the only weak accummulation point of $\{h(x_k)\}$. To prove the claim, let z be a weak accummulation point of $\{h(x_k)\}$, and take a subsequence $\{h(x_{k_l})\}$ weakly converging to z. Since $\{x_{k_l}\}$ is bounded, it has a weak accummulation point $\tilde{x}$. Denote still by $\{x_{k_l}\}$ the subsequence weakly converging to $\tilde{x}$. The weak-to-weak continuity of h yields

$$z = weak - \lim_l h(x_{k_l}) = h(\tilde{x}) = 0,$$

because $\tilde{x} \in P_*$ by (d). Hence $z = 0$ and hence $\{h(x_k)\}$ converges weakly to 0_H, as claimed. We claim now that $\liminf_k f(x_k, h(x_k)) \geq M_P$. Indeed, denote by $L := \liminf_k f(x_k, h(x_k))$ and assume that $L < M_P$. Then there exists $\delta > 0$ such that $L < M_P - \delta$. From the definition of $\liminf$ this implies that there exists a subsequence $\{x_{k_l}\}$ such that

$$f(x_{k_l}, h(x_{k_l})) < M_P - \delta.$$

Without loss of generality we can assume that $\{x_{k_l}\}$ converges weakly to $\hat{x}$, which by (d) belongs to P_*. Altogether,

$$M_P = f(\hat{x}, 0) \leq \liminf_l f(x_{k_l}, h(x_{k_l})) \leq M_P - \delta,$$

a contradiction. Hence we must have $\liminf_k f(x_k, h(x_k)) \geq M_P$, as claimed. Now we can write

$$\begin{aligned} 0 &\leq \limsup_k c_k\sigma(h(x_k)) = \limsup_k q(y_k, c_k) - f(x_k, h(x_k)) + \langle y_k, h(x_k)\rangle \\ &\leq \limsup_k q(y_k, c_k) + \limsup_k [-f(x_k, h(x_k))] + \limsup_k \langle y_k, h(x_k)\rangle \\ &\leq M_P - M_P = 0, \end{aligned}$$

where we used the fact that $\{h(x_k)\}$ converges weakly to 0_H in the last equality. The above expression yields $\lim_k c_k\sigma(h(x_k)) = 0$. Because c_k converges to $E(\bar{y}) = \bar{c} > 0$, the sequence $\{\sigma(h(x_k))\}$ must converge to zero. Using now Corollary 5.3 (part (a) implies (b)), we obtain (a). □

7. Final Remarks

In this paper we introduce a penalty map and relate its properties with exact penalty representation of a general augmented Lagrangian scheme. We show that primal convergence properties of subgradient type methods applied to the duality scheme via general augmented Lagrangian, under some mild assumptions, are directed related to strong exactness of the exact penalty map and differentiability of the dual function at the dual limit point.

Acknowledgements

Jefferson G. Melo was partially supported by Procad-UFG/UNB/IMPA.

References

[1] R. S. Burachik, *On primal convergence for augmented Lagrangian duality*, Optimization. 60 (2011), pp. 979–990.
[2] R. S. Burachik, R. N. Gasimov, N. A. Ismayilova and C. Y. Kaya, *On a modified subgradient algorithm for dual problems via sharp augmented Lagrangian*, J. Global Optim. 34 (2006), pp. 55–78.
[3] R. S. Burachik, A. N. Iusem and J. G. Melo, *A Primal Dual Modified Subgradient Algorithm with Sharp Lagrangian*, J. Global Optim. 46 (2010), pp. 347–361.
[4] R. S. Burachik, A. N. Iusem and J. G. Melo, *Duality and Exact Penalization for General Augmented Lagrangians*, J. Optim. Theory Appl. 147 (2010), pp. 125–140.
[5] R. S. Burachik, A. N. Iusem and J. G. Melo, *An Inexact Modified Subgradient Algorithm for Primal-Dual Problems via Augmented Lagrangians*, (*submitted*).
[6] R. S. Burachik, and C. Y. Kaya, *A Deflected Subgradient Method Using a General Augmented Lagrangian Duality with Implications on Penalty Methods*, Variational Analysis and Generalized Differentiation in Optimization and Control, special issue in honour of Boris Mordukhovich. Springer Optimization and Its Applications 47 (2010), pp. 109–132.
[7] R. S. Burachik, and C. Y. Kaya, *An augmented penalty function method with penalty parameter updates for nonconvex optimization*, Nonlinear Anal. 75 (2012), pp. 1158-1167.
[8] R. S. Burachik, and C. Y. Kaya. *An update rule and a convergence result for a penalty function method*. J. Ind. Manag. Optim. 3 (2008), pp. 381–398.
[9] R. S. Burachik, C. Y. Kaya, and M. Mammadov. *An inexact modified subgradient algorithm for nonconvex optimization*, Comput. Optim. Appl. 45 (2010), pp. 1–24.
[10] R. S. Burachik and A. M. Rubinov. *Abstract convexity and augmented Lagrangians*, SIAM J. Optim. 18 (2007), pp. 413–436.
[11] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM , New York, 1990.
[12] R. N. Gasimov. *Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming*. J. Global Optim. 24 (2002), pp. 187–203.
[13] X. X. Huang and X. Q. Yang. *A unified augmented Lagrangian approach to duality and exact penalization*, Math. Oper. Res. 28 (2003), pp. 533–552.
[14] C. S. Lalitha. *A new augmented Lagrangian approach to duality and exact penalization.* J. Global Optim. 46 (2010), pp. 233–245.
[15] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, Springer, Berlin, 1998.
[16] A. M. Rubinov and X. Q. Yang. *Lagrange-type functions in constrained non-convex optimization*, Applied Optimization Vol. 85, Kluwer Academic Publishers, Boston, MA, 2003.
[17] W. Schirotzek. *Nonsmooth Analysis*, Springer, Berlin, 2007.
[18] Y. Y. Zhou and X. Q. Yang. *Duality and Penalization in Optimization via an Augmented Lagrangian Function with Applications*, J. Optim. Theory Appl. 140 (2009), pp. 171–188.