New global optimality conditions for nonsmooth DC optimization problems

Dolgopolik M.V.*

October 3, 2019

Abstract

In this article we propose a new approach to an analysis of DC optimization problems. This approach was largely inspired by codifferential calculus and the method of codifferential descent, and is based on the use of a so-called affine support set of a convex function instead of the Frenchel conjugate function. With the use of affine support sets we define a global codifferential mapping of a DC function and derive new necessary and sufficient global optimality conditions for DC optimization problems. We also provide new simple necessary and sufficient conditions for the global exactness of the ℓ_1 penalty function for DC optimization problems with equality and inequality constraints and present a series of simple examples demonstrating a constructive nature of the new global optimality conditions. These examples show that when the optimality conditions are not satisfied, they can be easily utilised in order to find "global descent" directions of both constrained and unconstrained problems. As an interesting theoretical example, we apply our approach to the analysis of a nonsmooth problem of Bolza.

1 Introduction

For about thirty years DC optimization has been one of the most active areas of research in nonconvex optimization due the abundance of applications and a possibility of the use of the well-developed apparatus of convex analysis and convex optimization [33, 34, 42, 61–63]. Various local search [3, 25, 38, 39, 42, 45, 50, 53, 57] and global search [4–6, 24, 52, 56, 60, 64] methods for solving smooth and nonsmooth DC optimization problems were proposed over the years. It should be noted that global search methods are often based on global optimality conditions, which have attracted a lot of attention of researchers [14, 22, 28–30, 37, 46, 49, 51, 54, 55, 58, 59, 64, 65].

The main goal of this article is to present new necessary and sufficient global optimality conditions for nonsmooth DC optimization problems, including problems with DC equality and DC inequality constraints. These optimality conditions were largely inspired by the codifferential calculus developed by professor V.F. Demyanov [8–10, 13] and are intimately connected to the method of

---

*Institute for Problems in Mechanical Engineering of the Russian Academy of Sciences, Saint Petersburg, Russia

1

codifferential descent [3, 12, 13, 19, 21, 57]. To obtain new global optimality conditions, we introduce and study a so-called *affine support set* of a proper closed convex function. It should be noted that this set has been somewhat implicitly used in multiple monographs and papers on convex analysis and optimization (see, e.g., [23, Sect. I.3], [32, Theorem 1.3.8], [48, Sect. 7.3.3], etc.). However, to the best of author's knowledge, its properties have not been thoroughly investigated earlier.

Affine support sets of convex functions play the same role in the non-positively homogeneous case, as subdifferentials play in Minkowski duality. Furthermore, they are closely related to the abstract convexity theory [48] and Fenchel conjugate functions. In particular, almost all results on affine support sets have natural counterparts in terms of Fenchel conjugate functions. However, the use of affine support sets provides one with a new perspective on convex and DC functions, which allowed us to obtain a new result on convex functions (part 4 of Proposition 2). This result is a key ingredient in our derivation of new global optimality conditions for DC optimization problems.

With the use of affine support sets we define a global codifferential mapping of a DC function, which can be viewed as a "globalization" of Demyanov's definition of codifferential [13]. We provide some simple calculus rules for global codifferentials that are particularly useful in the piecewise affine case. Furthermore, we utilise global codifferentials and some results on affine support sets in order to obtain new necessary and sufficient global optimality conditions for nonsmooth DC optimization problems in terms of global codifferentials (different global optimality conditions in terms of codifferentials in the piecewise affine case were obtained by Polyakova [46]). It turns out that these condition are implicitly incorporated into the method of codifferential descent (see Remark 6 below and [19, 21]) and have a somewhat constructive nature in the piecewise affine case. Namely, we present a series of simple examples demonstrating that the verification of the global optimality conditions at a non-optimal point allows one to find "global descent" directions, which sometimes lead directly towards a global minimizer. In order to apply new global optimality conditions to problems with DC equality and DC inequality constraints we obtain new simple necessary and sufficient conditions for the global exactness of the ℓ_1 penalty function. Finally, as an interesting theoretical example, in the end of the paper we apply some results on global codifferentials of DC functions to an analysis of a nonsmooth problem of Bolza.

It should be noted that in many cases it is difficult to verify the global optimality conditions obtained in this paper, since it is often difficult to compute a global codifferential of a DC function explicitly. However, a similar statement is true for many other global optimality conditions for general DC optimization problems. Nevertheless, it seems possible to design new numerical methods for DC optimization problem utilising a certain approximation of global codifferential (cf. codifferential method in [3], aggregate codifferential method in [57], and Example 9 below).

The paper is organised as follows. In Section 2 we introduce an affine support set of a convex function, study its properties, and point out its connection with the Fenchel conjugate function. Section 3 is devoted to necessary and sufficient global optimality conditions for nonsmooth DC optimization problems in terms of global codifferentials. In this section, we also present a series of simple examples demonstrating a somewhat constructive nature of the global optimal-

ity conditions and obtain simple conditions for the global exactness of the ℓ_1 penalty function for DC optimization problems with equality and inequality constrains. Some connections of the global optimality conditions obtained in this paper with KKT optimality conditions and global optimality conditions in terms of ε-subdifferentials are discussed in Section 4. Finally, different global optimality conditions in terms of global codifferentials and their application to an analysis of a nonsmooth problem of Bolza are given in Section 5.

For the sake of simplicity, in this paper we study DC functions defined on a real Hilbert space. However, it should be noted that most of the results of Sections 2 and 5 (except for part 4 of Proposition 2 and Proposition 3) can be easily extended to the case of locally convex spaces, while the rest of the results of this paper (apart from Theorem 3) remain valid in strictly convex reflexive Banach spaces.

2 Affine support sets of convex functions

In this section we introduce and study a so-called affine support set of a closed convex function. The main ideas and results presented below, in a sense, can be viewed as a natural extension of the Minkowski duality to the case of general, i.e. non-positively homogeneous convex functions (cf. the abstract convexity theory in [48]).

Let $\mathcal{H}$ be a real Hilbert space, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, and $f\colon \mathcal{H} \to \overline{\mathbb{R}}$ be a proper closed convex function. As is well known (see, e.g. [23, Prp. I.3.1]), the function f can be represented as the supremum of a family of affine functions. Taking, if necessary, the closed convex hull of this set, and identifying an affine function $l(x) = a + \langle v, x \rangle$ with the point $(a, v) \in \mathbb{R} \times \mathcal{H}$, one gets that there exists a closed convex set $S_f \subset \mathbb{R} \times \mathcal{H}$ such that

$$f(x) = \sup_{(a,v) \in S_f} (a + \langle v, x \rangle) \quad \forall x \in \mathcal{H},$$

where $\langle \cdot, \cdot \rangle$ is the inner product in $\mathcal{H}$. Any such set S_f is called *an affine support set* of the function f. At first, let us demonstrate how affine support sets are connected with the ε-subdifferential of the function f.

Proposition 1. *For any affine support set S_f of f and for all $\varepsilon \geq 0$ and $x \in \operatorname{dom} f$ one has*

$$\partial_\varepsilon f(x) = \big\{ v \in \mathcal{H} \mid \exists a \in \mathbb{R} \colon (a, v) \in S_f,\ a + \langle v, x \rangle \geq f(x) - \varepsilon \big\}. \qquad (1)$$

Proof. Fix arbitrary $\varepsilon \geq 0$ and $x \in \operatorname{dom} f$, and denote by $D_\varepsilon(x)$ the set on the right-hand side of (1). Observe that for any $(a, v) \in S_f$ such that $a + \langle v, x \rangle \geq f(x) - \varepsilon$ one has

$$f(y) - f(x) \geq a + \langle v, y \rangle - \big(a + \langle v, x \rangle\big) - \varepsilon = \langle v, y - x \rangle - \varepsilon \quad \forall y \in \mathcal{H},$$

which implies that $v \in \partial_\varepsilon f(x)$. Thus, $D_\varepsilon(x) \subseteq \partial_\varepsilon f(x)$.

Arguing by reductio ad absurdum, suppose that $\partial_\varepsilon f(x) \neq D_\varepsilon(x)$. Then there exists $v_0 \in \partial_\varepsilon f(x)$ such that $v_0 \notin D_\varepsilon(x)$. Hence $(a, v_0) \notin S_f$ for any $a \geq f(x) - \langle v_0, x \rangle - \varepsilon$, since otherwise $v_0 \in D_\varepsilon(x)$.

Denote $C_f = \{(b, v) \in \mathbb{R} \times \mathcal{H} \mid \exists a \geq b \colon (a, v) \in S_f\}$. It is clear that the set C_f is convex, and $(f(x) - \langle v_0, x \rangle - \varepsilon, v_0) \notin C_f$. To apply the separation

theorem, let us check that the set C_f is closed. To this end, introduce a function $g\colon \mathcal{H} \to \mathbb{R}$ as follows: $g(v) = \sup\{a \mid (a, v) \in S_f\}$. Observe that $(g(v), v) \in S_f$ for any $v \in \operatorname{dom} g$ due to the fact that the set S_f is closed. Moreover, it is easy to see that C_f is the hypograph of the function g. Therefore, it is sufficient to check that the function g is upper semicontinuous (u.s.c.).

At first, note that g is a proper concave function, since its hypograph is a convex set, and if $g(v) = +\infty$ for some v (i.e. $(a, v) \in S_f$ for any sufficiently large a), then $f(\cdot) \equiv +\infty$, which contradicts the assumption that the function f is proper. Note also that $g(\cdot) \not\equiv -\infty$, since otherwise $S_f = \emptyset$ and $f(\cdot) \equiv -\infty$, which contradicts our assumption. Furthermore, g is bounded above on any bounded set. Indeed, for any bounded set $Q \subset \mathcal{H}$ and $v \in Q$ either $(\mathbb{R} \times \{v\}) \cap S_f = \emptyset$ and $g(v) = -\infty$ or $(a, v) \in S_f$ for some $a \in \mathbb{R}$, and

$$g(v) = \sup\{a \mid (a, v) \in S_f\} = \sup_{a\colon (a,v)\in S_f} \big(a + \langle v, x \rangle - \langle v, x \rangle\big)$$

$$\leq \sup_{(a,v)\in S_f} \big(a + \langle v, x \rangle\big) - \langle v, x \rangle \leq f(x) + q\|x\|,$$

where $q = \sup_{v \in Q} \|v\|$ (recall that $x \in \operatorname{dom} f$, i.e. $f(x) < +\infty$).

Arguing by reductio ad absurdum suppose that g is not u.s.c. at a point $v \in \mathcal{H}$. Let $v \in \operatorname{dom} g$. Then there exists $\theta > 0$ such that for any $n \in \mathbb{N}$ one can find $v_n \in \operatorname{dom} g$ for which $g(v_n) > g(v) + \theta$ and $\|v_n - v\| < 1/n$. Taking into account the fact that g is bounded above on bounded sets one gets that the sequence $\{g(v_n)\}$ is bounded. Therefore, there exists a subsequence $\{v_{n_k}\}$ such that the corresponding subsequence $\{g(v_{n_k})\}$ converges to some $g_* \geq g(v) + \theta$. As was poited out above, $(g(v_{n_k}), v_{n_k}) \in S_f$ for all $k \in \mathbb{N}$. Hence passing to the limit as $k \to \infty$ and applying the closedness of the set S_f one obtains that $(g_*, v) \in S_f$. Consequently, $g(v) \geq g_* \geq g(v) + \theta$, which is impossible.

Let now $v \notin \operatorname{dom} g$. Then there exist $M \in \mathbb{R}$ and a sequence $\{v_n\} \subset \operatorname{dom} g$ converging to v such that $g(v_n) \geq M$ for all $n \in \mathbb{N}$. Applying, as above, the fact that the sequence $\{g(v_n)\}$ is bounded one can extract a subsequence $\{v_{n_k}\}$ such that the sequence $\{g(v_{n_k})\}$ converges to some $g_* \geq M > -\infty$. Therefore $(g_*, v) \in S_f$, and $g(v) \geq g_* > -\infty$, which is impossible. Thus, g is u.s.c., and the set C_f is closed.

Recall that $(f(x) - \langle v_0, x \rangle - \varepsilon, v_0) \notin C_f$, and C_f is a closed convex set. Applying the separation theorem one obtains that there exist $(b, y) \in \mathbb{R} \times \mathcal{H}$ and $\delta > 0$ such that

$$b(f(x) - \langle v_0, x \rangle - \varepsilon) + \langle v_0, y \rangle \geq ba + \langle v, y \rangle + \delta \quad \forall (a, v) \in C_f. \tag{2}$$

By definition for any $(a, v) \in S_f$ one has $(-\infty, a] \times \{v\} \subset C_f$, which implies that $b \geq 0$.

If $b > 0$, then dividing (2) by b and taking the supremum over all $(a, v) \in S_f$ one obtains

$$f(x) + \left\langle v_0, \frac{1}{b}y - x \right\rangle - \varepsilon \geq f\left(\frac{1}{b}y\right) + \frac{\delta}{b}.$$

Recall that $v_0 \in \partial_\varepsilon f(x)$. Therefore

$$f\left(\frac{1}{b}y\right) \geq f(x) + \left\langle v_0, \frac{1}{b}y - x \right\rangle - \varepsilon \geq f\left(\frac{1}{b}y\right) + \frac{\delta}{b},$$

which is impossible. Thus, $\partial_\varepsilon f(x) = D_\varepsilon(x)$.

Suppose now that $b = 0$. Then (2) implies that

$$\frac{f(x + \alpha y) - f(x)}{\alpha} = \frac{1}{\alpha}\Big(\sup_{(a,v) \in S_f} \big(a + \langle v, x + \alpha y \rangle\big) - f(x)\Big)$$

$$\leq \frac{1}{\alpha}\Big(\sup_{(a,v) \in S_f} (a + \langle v, x \rangle)) + \alpha \langle v_0, y \rangle - \alpha \delta - f(x)\Big) = \langle v_0, y \rangle - \delta \quad (3)$$

for any $\alpha > 0$. On the other hand, by the definition of ε-subgradient for any $\alpha > \varepsilon/\delta$ one has

$$\frac{f(x + \alpha y) - f(x)}{\alpha} \geq \langle v_0, y \rangle - \frac{\varepsilon}{\alpha} > \langle v_0, y \rangle - \delta,$$

which contradicts (3). Thus, $\partial_\varepsilon f(x) = D_\varepsilon(x)$, and the proof is complete. $\qquad\square$

Remark 1. By the proposition above the supremum in the definition of affine support set is attained for some $x \in \operatorname{dom} f$ iff f is subdifferentiable at x. In particular, if f is finite-valued, then the supremum in the definition of affine support set is attained for any $x \in \mathcal{H}$ by [23, Proposition I.5.2 and Corollary I.2.5].

Let S_f be any affine support set of f. Our aim now is to show that several important properties of the function f, such as boundedness below and the attainment of minimum, can be described in terms of simple geometric properties of the set S_f.

Observe that if f attains a global minimum at a point x_*, then $0 \in \partial f(x_*)$, and $(f(x_*), 0) \in S_f$ by Proposition 1. Thus, the sets $\mathbb{R} \times \{0\}$ and S_f intersect. In the general case, define $a_f = \sup_{(a,0) \in S_f} a$. By definition $a_f = -\infty$, if the sets $\mathbb{R} \times \{0\}$ and S_f do not intersect. Note also that if they do intersect, then $(a_f, 0) \in S_f$ due to the facts that (i) this intersection is obviously closed, and (ii) if $a_f = +\infty$, then $f(\cdot) \equiv +\infty$, which contradicts the assumption that the function f is proper.

Denote by $N_f = \{(b, w) \in \mathbb{R} \times \mathcal{H} \mid b(a - a_f) + \langle w, v \rangle \leq 0 \ \forall (a, v) \in S_f\}$ the normal cone to the set S_f at the point $(a_f, 0)$, if the sets $\mathbb{R} \times \{0\}$ and S_f intersect, and define $N_f = \emptyset$ otherwise. From this point onwards we suppose that the space $\mathbb{R} \times \mathcal{H}$ is endowed with the norm $\|(a, v)\| = \sqrt{a^2 + \|v\|^2}$.

Proposition 2. *The following statements hold true:*

1. *f is bounded below iff $S_f \cap (\mathbb{R} \times \{0\}) \neq \emptyset$;*

2. *if f is bounded below, then $a_f = \inf_{x \in \mathcal{H}} f(x)$;*

3. *f attains a global minimum iff there exists $(b, w) \in N_f$ such that $b > 0$; furthermore, $\arg\min_{x \in \mathcal{H}} f(x) = \{b^{-1}w \in \mathcal{H} \mid (b, w) \in N_f \colon b > 0\}$;*

4. *if $f(x) \geq 0$ for all $x \in \mathcal{H}$, then either $0 \in S_f$ or $a_* > 0$, where (a_*, v_*) is a globally optimal solution of the problem*

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a, v)\|^2 \quad \text{subject to} \quad (a, v) \in S_f;$$

   conversely, if f is bounded below and either $0 \in S_f$ or $a_* > 0$, then $f(x) \geq 0$ for all $x \in \mathcal{H}$. Moreover, in the case $a_* > 0$ one has $a_f > 0$, i.e. $\inf_{x \in \mathcal{H}} f(x) > 0$.

5

Proof. 1. If $S_f \cap (\mathbb{R} \times \{0\}) \neq \emptyset$, then there exists $a_0 \in \mathbb{R}$ such that $(a_0, 0) \in S_f$. By the definition of S_f for all $x \in \mathcal{H}$ one has $f(x) \geq a_0$, i.e. f is bounded below.

Suppose, now, that f is bounded below. Denote $f_* = \inf_{x \in \mathcal{H}} f(x)$. Then for any $\varepsilon > 0$ there exists $x_\varepsilon \in \mathcal{H}$ such that $f(x_\varepsilon) \leq f_* + \varepsilon$. Hence $0 \in \partial_\varepsilon f(x_\varepsilon)$, which with the use of Proposition 1 implies that there exists $a \geq f(x_\varepsilon) - \varepsilon$ such that $(a, 0) \in S_f$, i.e. $S_f \cap (\mathbb{R} \times \{0\}) \neq \emptyset$.

2. As was just proved, for any $\varepsilon > 0$ there exists $a \geq f(x_\varepsilon) - \varepsilon \geq f_* - \varepsilon$ such that $(a, 0) \in S_f$. Therefore $a_f \geq f_*$. On the other hand, for any $(a, 0) \in S_f$ and $x \in \mathcal{H}$ one obviously has $f(x) \geq a$, which implies that $a_f \leq f_*$. Thus, $a_f = f_*$.

3. Let f attain a global minimum at a point $x_* \in \mathcal{H}$. By definition $f(x_*) = \sup_{(a,v) \in S_f} (a + \langle v, x_* \rangle) = f_*$ or, equivalently,

$$(a - f_*) + \langle v, x_* \rangle \leq 0 \quad \forall (a, v) \in S_f,$$

which implies that $(1, x_*) \in N_f$ (note that $(f_*, 0) \in S_f$ and $a_f = f_*$ by the second part of the theorem).

Suppose, now, that $N_f \neq \emptyset$, and there exists $(b, w) \in N_f$ with $b > 0$. By the definition of N_f and the second part of the theorem one has

$$b(a - f_*) + \langle w, v \rangle \leq 0 \quad \forall (a, v) \in S_f.$$

Dividing by b and taking the supremum over all $(a, v) \in S_f$ one obtains

$$f\left(\frac{1}{b} w\right) = \sup_{(a,v) \in S_f} \left(a + \left\langle v, \frac{1}{b} w \right\rangle\right) \leq f_*,$$

which implies that $b^{-1} w$ is a global minimizer of f. Thus, $\arg\min_{x \in \mathcal{H}} f(x) = \{b^{-1} w \in \mathcal{H} \mid (b, w) \in N_f : b > 0\}$.

4. Let $f(x) \geq 0$ for all $x \in \mathcal{H}$. Arguing by reductio ad absurdum, suppose that $0 \notin S_f$ and $a_* \leq 0$. From the definition of (a_*, v_*) and the necessary condition for a minimum of a differentiable function on a convex set it follows that

$$a_*(a - a_*) + \langle v_*, v - v_* \rangle \geq 0 \quad \forall (a, v) \in S_f. \tag{4}$$

If $a_* = 0$, then one gets that $\langle v, -v_* \rangle \leq -\|v_*\|^2 < 0$ for all $(a, v) \in S_f$ (note that $v_* \neq 0$, since otherwise $0 \in S_f$). Therefore for any $\alpha > 0$ and $x \in \operatorname{dom} f$ one has

$$f(x - \alpha v_*) = \sup_{(a,v) \in S_f} (a + \langle v, x \rangle + \alpha \langle v, -v_* \rangle) \leq f(x) - \alpha \|v_*\|^2.$$

Consequently, $f(x - \alpha v_*) \to -\infty$ as $\alpha \to +\infty$, which is impossible.

If $a_* < 0$, then dividing (4) by a_* and taking the supremum over all $(a, v) \in S_f$ one obtains that

$$f\left(\frac{1}{a_*} v_*\right) = \sup_{(a,v) \in S_f} \left(a + \left\langle v, \frac{1}{a_*} v_* \right\rangle\right) \leq a_* + \frac{1}{a_*} \|v_*\|^2 < 0,$$

which contradicts the assumption that f is nonnegative.

Let us prove the converse statement. If $0 \in S_f$, then, obviously, one has $f(x) \geq 0$ for all $x \in \mathcal{H}$. Therefore, let $0 \notin S_f$ and $a_* > 0$. Arguing by reductio ad absurdum, suppose that $f_* = \inf_{x \in \mathcal{H}} f(x) < 0$ (note that $f_* > -\infty$

due to the assumption that f is bounded below). By the second part of the theorem one has $(f_*, 0) \in S_f$. Consequently, for any $\alpha \in [0, 1]$ one has $\alpha(a_*, v_*) + (1 - \alpha)(f_*, 0) \in S_f$. Setting $\alpha = |f_*|/(|f_*| + a_*) \in (0, 1)$ one obtains that $(0, \alpha v_*) \in S_f$, which is impossible due to the definition of (a_*, v_*) and the obvious inequality $\|(0, \alpha v_*)\|^2 < \|(a_*, v_*)\|^2$. Thus, the function f is nonnegative. It remains to note that $a_f > 0$ in the case when $a_* > 0$ by virtue of the facts that $a_f \geq 0$ due to the nonnegativity of the function f, and $a_f \neq 0$, since otherwise $0 \in S_f$ and $a_* = 0$. □

Remark 2. (i) Let us note that the assumption on the boundedness below of the function f cannot be dropped from the last part of the proposition above. Indeed, if $f(x) \equiv a + \langle v, x \rangle$ with $a > 0$ and $v \neq 0$, then defining $S_f = (a, v)$ one obtains that $a_* > 0$, but the function f is not nonnegative.
(ii) From the proof of the last part of the proposition above it follows that if $0 \notin S_f$, but $a_* = 0$, then f is not bounded below. Consequently, if f is bounded below, then f is nonnegative iff $a_* \geq 0$. Furthermore, note that if $a_* < 0$, then $f(\frac{1}{a_*} v_*) < 0$.

Let us give a simple example illustrating the proposition above.

Example 1. Let $\mathcal{H} = \mathbb{R}$, and $S_f = \{(a, v) \in \mathbb{R}^2 \mid (a+1)^2 + (v-1)^2 \leq 1\}$. Then according to Proposition 2 one has $f_* = \inf_{x \in \mathbb{R}} f(x) = -1$. Furthermore, it is easy to check that $N_f = \{(a, v) \in \mathbb{R}^2 \mid a = 0, \ v \leq 0\}$, which by Proposition 2 implies that the function f does not attain a global minimum. Let us verify this directly. Indeed, for any $x \in \mathbb{R}$ one has

$$f(x) = \max_{(a,v) \in S_f} (a + vx) = \max\{(a-1) + (v+1)x \mid a^2 + v^2 \leq 1\} = \sqrt{1 + x^2} + x - 1.$$

Thus, $f_* = -1$, and f does not attain a global minimum.

Let us also obtain an extension of part 4 of Proposition 2 to the case when the nonnegativity of the function f is checked on a set defined by an inequality constraint.

Proposition 3. *Let $g \colon \mathcal{H} \to \overline{\mathbb{R}}$ be a proper closed convex function, and let S_g be any affine support set of g. Suppose also that $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. If $f(x) \geq 0$ for all x satisfying the inequality $g(x) \leq 0$, then either $0 \in \operatorname{cl} \operatorname{co}\{S_f, S_g\}$ or $a_* > 0$, where (a_*, v_*) is a globally optimal solution of the problem*

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a, v)\|^2 \quad subject \ to \quad (a, v) \in \operatorname{cl} \operatorname{co}\{S_f, S_g\}.$$

Conversely, if f is bounded below and continuous on the set $\{x \mid g(x) \leq 0\}$, $0 \notin S_g$, and either $0 \in \operatorname{cl} \operatorname{co}\{S_f, S_g\}$ or $a_* > 0$, then $f(x) \geq 0$ for all x satisfying the inequality $g(x) \leq 0$. Moreover, in the case $a_* > 0$ there exists $\gamma > 0$ such that $f(x) \geq \gamma$ for all x satisfying the inequality $g(x) < \gamma$.

Proof. Let f be nonnegative on the set $\{x \mid g(x) \leq 0\}$. Arguing by reductio ad absurdum, suppose that $0 \notin \operatorname{cl} \operatorname{co}\{S_f, S_g\}$ and $a_* \leq 0$. By the necessary condition for a minimum of a convex function on a convex set one obtains that

$$a_*(a - a_*) + \langle v_*, v - v_* \rangle \geq 0 \quad \forall (a, v) \in \operatorname{cl} \operatorname{co}\{S_f, S_g\}. \tag{5}$$

If $a_* < 0$, then dividing this inequality by a_*, and at first taking the supremum over all $(a, v) \in S_f$, and at second taking the supremum over all $(a, v) \in S_g$ one obtains that

$$f\left(\frac{1}{a_*}v_*\right) \leq a_* + \frac{1}{a_*}\|v_*\| < 0, \quad g\left(\frac{1}{a_*}v_*\right) \leq a_* + \frac{1}{a_*}\|v_*\| < 0,$$

which is impossible. On the other hand, if $a_* = 0$, then from (5) it follows that $\langle v, -v_* \rangle \leq -\|v_*\|^2 < 0$ for all $(a, v) \in S_f \cup S_g$ (note that $v_* \neq 0$, since otherwise $0 \in \mathrm{cl}\,\mathrm{co}\{S_f, S_g\}$). Hence for any $x \in \mathrm{dom}\, f \cap \mathrm{dom}\, g$ and for all $\alpha > 0$ one has

$$f(x - \alpha v_*) \leq f(x) - \alpha\|v_*\|^2, \quad g(x - \alpha v_*) \leq g(x) - \alpha\|v_*\|^2.$$

Consequently, $f(x - \alpha v_*) < 0$ and $g(x - \alpha v_*) < 0$ for any sufficiently large $\alpha > 0$, which is impossible.

Let us prove the converse statement. Define $h(\cdot) = \sup\{f(\cdot), g(\cdot)\}$. It is easily seen that $\mathrm{cl}\,\mathrm{co}\{S_f, S_g\}$ is an affine support set of the function h. Our aim is to verify that $f(x) \geq 0$ on the set $\{x \mid g(x) \leq 0\}$ iff $h(x) \geq 0$ for all $x \in \mathcal{H}$, provided $0 \notin S_g$. Then applying the last part of Proposition 2 to the function h one obtains the desired result.

Clearly, if $f(x) \geq 0$ for all x satisfying the inequality $g(x) \leq 0$, then $h(\cdot) \geq 0$. Let us prove the converse statement. If the set $\{x \mid g(x) \leq 0\}$ is empty, then the statement holds vacuously. Therefore, suppose that this set is not empty. Note that if $\inf_{x \in \mathcal{H}} g(x) = 0$, then $0 \in S_g$ by Proposition 2, which contradicts our assumption. Thus, there exists x_0 such that $g(x_0) < 0$, i.e. Slater's condition holds true.

Suppose that the function h is nonnegative. Then $f(x) \geq 0$ for any x satisfying the strict inequality $g(x) < 0$. From the convexity of the function g it follows that $\{x \mid g(x) \leq 0\} = \mathrm{cl}\{x \mid g(x) < 0\}$, since for any point x such that $g(x) = 0$ one has $g(\alpha x + (1 - \alpha)x_0) < 0$ for all $\alpha \in [0, 1)$. Consequently, applying the fact that f is continuous on $\{x \mid g(x) \leq 0\}$ one obtains that $f(x) \geq 0$ for all x satisfying the inequality $g(x) \leq 0$, and the proof is complete. $\square$

Remark 3. The assumption that f is bounded below on $\{x \mid g(x) \leq 0\}$ is necessary for the validity of the converse statement of the previous proposition. Indeed, if $f(x) = g(x) = a + \langle v, x \rangle$ for some $a > 0$ and $v \neq 0$, then $a_* = a > 0$, but $f(x) < 0$ for any x such that $g(x) < 0$. The assumption $0 \notin S_g$ is also necessary for the validity of the converse statement of the proposition, since if $0 \in S_g$, then $0 \in \mathrm{cl}\,\mathrm{co}\{S_f, S_g\}$ regardless of the behaviour of the function f. Furthermore, note that the assumption $0 \notin S_g$ is, in fact, equivalent to Slater's condition, provided the set $\{x \mid g(x) \leq 0\}$ is not empty.

With the use of Proposition 2 we can point out a direct connection between affine support sets of f and the Frenchel conjugate function f^*.

Proposition 4. *Let S_f be any affine support set of f. Then*

$$\sup\{a \mid (a, v) \in S_f\} = -f^*(v) \quad \forall v \in \mathcal{H}. \tag{6}$$

In particular, any affine support set of the function f is contained in the set $\{(a, v) \in \mathbb{R} \times \mathcal{H} \mid a \leq -f^*(v)\}$. Furthermore, the set

$$\begin{aligned} S_f &= \mathrm{cl}\,\mathrm{co}\{(-f^*(v), v) \in \mathbb{R} \times \mathcal{H} \mid v \in \mathrm{dom}\, f^*\} \\ &= \mathrm{cl}\,\mathrm{co}\{(f(y) - \langle v, y \rangle, v) \in \mathbb{R} \times \mathcal{H} \mid y \in \mathrm{dom}\,\partial f, \, v \in \partial f(y)\} \end{aligned} \tag{7}$$

is the smallest (by inclusion) affine support set of the function f.

Proof. Fix $v \in \mathcal{H}$, and consider the function $g(x) = f(x) - \langle v, x \rangle$. Note that this function is bounded below iff $v \in \operatorname{dom} f^*$. On the other hand, from the fact that the set $S_f - (0, v)$ is an affine support set of this function and the first part of Proposition 2 it follows that g is bounded below iff there exists $a \in \mathbb{R}$ such that $(a, v) \in S_f$. Furthermore, if $v \in \operatorname{dom} f^*$, then applying the second part of Proposition 2 one obtains that

$$-f^*(v) = \inf_{x \in \mathcal{H}} (f(x) - \langle v, x \rangle) = \sup\{a \mid (a, 0) \in S_f - (0, v)\}$$
$$= \sup\{a \mid (a, v) \in S_f\},$$

i.e. (6) holds true, and $S_f \subseteq \{(a, v) \in \mathbb{R} \times \mathcal{H} \mid a \leq -f^*(v)\}$. Hence and from the fact that

$$f(x) = f^{**}(x) = \sup_{v \in \operatorname{dom} f^*} \left(\langle v, x \rangle - f^*(v) \right) \quad \forall x \in \mathcal{H} \tag{8}$$

it follows that set (7) is the smallest affine support set of the function f. It remains to note that the second equality in (7) follows directly from the facts that (i) one can take the supremum in (8) over all $v \in \operatorname{dom} \partial f^*$ (since if $v \in \operatorname{dom} f^* \setminus \operatorname{dom} \partial f^*$, then for any $x \in \mathcal{H}$ by definition there exists $w \in \operatorname{dom} f^*$ such that $\langle w, x \rangle - f^*(w) > \langle v, x \rangle - f^*(v)$), and (ii) $v \in \operatorname{dom} \partial f^*$ iff $v \in \partial f(y)$ for some $y \in \operatorname{dom} \partial f$ iff $f^*(v) = \langle v, y \rangle - f(y)$ by [32, Corollary X.1.4.4]. $\square$

Remark 4. The proposition above demonstrates that there is a direct connection between affine support sets and conjugate functions. Note, in particular, that the function $g(v)$ defined in the proof of Proposition 1 is, in fact, the negative of the conjugate function f^*. Furthermore, Proposition 1 itself is a reformulation of the standard characterization of ε-subgradients via the conjugate function (see, e.g. [32, Proposition XI.1.2.1]) in terms of affine support sets. In the light of Proposition 4 we can also give a simple interpretation of Proposition 2. The first two statements of this proposition is nothing but the obvious equality $\inf_{x \in \mathcal{H}} f(x) = -f^*(0)$. The third one is a combination of the equality $\arg\min_{x \in \mathcal{H}} f(x) = \partial f^*(0)$ and the well-known geometric interpretation of the subdifferential in terms of the normal cone to the epigraph of a convex function (see, e.g. [31, Proposition VI.1.3.1]). However, to the best of author's knowledge, the last statement of Proposition 2 is completely new. Furthermore, the last statement of this propositon is a basis of new global optimality conditions for DC optimization problems derived in the next section.

Let us present some simple calculus rules for affine support sets. Their proofs are straightforward and therefore are omitted.

Proposition 5 (linear combination). *Let $f_i \colon \mathcal{H} \to \overline{\mathbb{R}}$, $i \in I = \{1, \ldots, l\}$ be proper closed convex functions, and let S_{f_i} be any affine support set of f_i, $i \in I$. Then for any $\lambda_i \geq 0$, $i \in I$, the set $S_f = \operatorname{cl}(\sum_{i \in I} \lambda_i S_{f_i})$ is an affine support set of the function $f = \sum_{i \in I} \lambda_i f_i$.*

Proposition 6 (affine transformation). *Let $g \colon \mathcal{H} \to \overline{\mathbb{R}}$ be a proper closed convex function, and S_g be any affine support set of g. Suppose also that X is a Hilber*

space, $T\colon X \to \mathcal{H}$ is a bounded linear operator, and $f(x) = g(Tx + b)$ for some $b \in \mathcal{H}$. Then the set

$$S_f = \mathrm{cl}\left\{(a + \langle v, b\rangle, T^*v) \in \mathbb{R} \times X \mid (a, v) \in S_g\right\} \tag{9}$$

is an affine support set of f. Moreover, the closure operator in (9) can be dropped, if S_g is bounded or T is invertible.

Proposition 7 (supremum). *Let Y be a nonempty set, and a function $f\colon \mathcal{H} \times Y \to \overline{\mathbb{R}}$ be such that for any $y \in Y$ the function $f(\cdot, y)$ is proper, closed, and convex. Suppose also that $S(y)$ is an affine support set of the function $f(\cdot, y)$, and the function $g(\cdot) = \sup_{y \in Y} f(\cdot, y)$ is proper. Then $S_g = \mathrm{cl}\,\mathrm{co}\{S(y) \mid y \in Y\}$ is an affine support set of the function g.*

In the end of this section, let us give several simple examples demonstrating how one can compute affine support sets of convex functions with the use of Proposition 4 and some other well-known results.

Example 2. If f is a proper closed positively homogeneous convex function, then the set $S_f = \{0\} \times \partial f(0)$ is an affine support set of f (see, e.g. [31, Theorem V.3.1.1]).

Example 3. If $\mathcal{H} = \mathbb{R}^d$, and f is a finite polyhedral convex function, then $f(x) = \max_{1 \le i \le n}(a_i + \langle v_i, x\rangle)$ for some $n \in \mathbb{N}$ and $(a_i, v_i) \in \mathbb{R}^{d+1}$ (see [47, Sect. 19]). Consequently, the set $S_f = \mathrm{co}\{(a_i, v_i) \mid i \in I\}$ is an affine support set of f. Therefore, a finite convex function f is polyhedral iff there exists an affine support set of this function that is a convex polytope.

Example 4. If f is Gâteaux differentiable on its effective domain, then

$$S_f = \mathrm{cl}\,\mathrm{co}\left\{\big(f(x) - \langle \nabla f(x), x\rangle, \nabla f(x)\big) \in \mathbb{R} \times \mathcal{H} \mid x \in \mathrm{dom}\, f\right\}$$

is an affine support set of f. Here $\nabla f(x)$ is the gradient of f at x In particular, if $f(x) = 0.5\langle x, Ax\rangle + \langle b, x\rangle$, where the linear operator $A\colon \mathcal{H} \to \mathcal{H}$ is positive semidefinite, then $S_f = \mathrm{cl}\,\mathrm{co}\{(-0.5\langle x, Ax\rangle, Ax + b) \mid x \in \mathcal{H}\}$ is an affine support set of f. Note that in this case it is easier to describe the affine support set with the use of the gradient rather than the conjugate function (cf. (7)), since the conjugate function is defined via the pseudoinverse operator of A.

3 Global codifferential calculus and optimality conditions

In this section we apply the main results on affine support sets of convex functions obtained above to DC optimization problems. In particular, with the use of Proposition 2 we obtain new necessary and sufficient conditions for global optimality in DC optimization. Hereinafter we consider only finite-valued DC functions $f\colon \mathcal{H} \to \mathbb{R}$ defined on the entire space $\mathcal{H}$.

Let f be a DC function, i.e. let $f = g - h$, where $g, h\colon \mathcal{H} \to \mathbb{R}$ are closed convex functions. Suppose also that S_g and S_h are any affine support sets of the functions g and h respectively. Introduce the set-valued mappings

$$\begin{aligned}
\underline{d}f(x) &= \left\{(a - g(x) + \langle v, x\rangle, v) \in \mathbb{R} \times \mathcal{H} \mid (a, v) \in S_g\right\}, \\
\overline{d}f(x) &= \left\{(-b + h(x) - \langle w, x\rangle, -w) \in \mathbb{R} \times \mathcal{H} \mid (b, w) \in S_h\right\}.
\end{aligned} \tag{10}$$

Then for any $x, \Delta x \in \mathcal{H}$ the following equality holds true:

$$f(x + \Delta x) - f(x) = \sup_{(a,v)\in \underline{d}f(x)} (a + \langle v, \Delta x \rangle) + \inf_{(b,w)\in \overline{d}f(x)} (b + \langle w, \Delta x \rangle) \quad (11)$$

(in actuality, the supremum and the infimum are attained by Remark 1). Indeed, by definition one has

$$\begin{aligned}
g(x + \Delta x) - g(x) &= \sup_{(a,v)\in S_g} (a + \langle v, x + \Delta x \rangle) - g(x) \\
&= \sup_{(a,v)\in S_g} (a - g(x) + \langle v, x \rangle + \langle v, \Delta x \rangle) \\
&= \sup_{(a,v)\in \underline{d}f(x)} (a + \langle v, \Delta x \rangle). \quad (12)
\end{aligned}$$

Subtracting from this equality the same one for the function $h(x)$ one obtains that (11) is valid. Furthermore, for any $x \in \mathcal{H}$ one has

$$\sup_{(a,v)\in \underline{d}f(x)} a = \sup_{(a,v)\in S_g} (a + \langle v, x \rangle) - g(x) = 0,$$

and, similarly, $\inf_{(b,w)\in \overline{d}f(x)} b = 0$. Finally, observe that the sets $\underline{d}f(x)$ and $\overline{d}f(x)$ are convex and closed due to the fact that the map $(a,v) \mapsto (a - g(x) + \langle v, x \rangle, v)$ is a homeomorphism of $\mathbb{R} \times \mathcal{H}$. Thus, the pair $[\underline{d}f(x), \overline{d}f(x)]$ has similar properties to codifferential of f at x [8–10,13]. Therefore, it is natural to call the pair $Df = [\underline{d}f, \overline{d}f]$ a *global codifferential mapping* (or simply *a global codifferential*) of the function f associated with the DC decomposition $f = g - h$. The multifunction $\underline{d}f$ is called *a global hypodifferential* of f, while the multifunction $\overline{d}f$ is called *a global hyperdifferential* of f. Note that global codifferential mappings are obviously not unique, since there exist infinitely many DC decompositions of a DC function.

Let us point out some simple calculus rules for global codifferentials. Their proofs are straightforward, and we omit them for the sake of shortness (see [21, Proposition 4.4] for some details).

Proposition 8. *Let f_i, $i \in I = \{1, \ldots k\}$, be DC functions, and let Df_i be the global codifferential mapping of f_i associated with a DC decomposition $f_i = g_i - h_i$. The following statements hold true:*

1. *if $f = f_1 + c$ for some $c \in \mathbb{R}$, then $Df = Df_1$;*

2. *if $f = \sum_{i=1}^{k} f_i$, then $Df = [\mathrm{cl}(\sum_{i=1}^{k} \underline{d}f_i), \mathrm{cl}(\sum_{i=1}^{k} \overline{d}f_i)]$ is a global codifferential mapping of the function f associated with the DC decomposition $f = \sum_{i=1}^{k} g_i - \sum_{i=1}^{k} h_i$;*

3. *if $f = \lambda f_1$, then $Df = [\lambda \underline{d}f_1, \lambda \overline{d}f_1]$ is a global codifferential mapping of f associated with the DC decomposition $f = \lambda g_1 - \lambda h_1$ in the case $\lambda \geq 0$, and $Df = [\lambda \overline{d}f_1, \lambda \underline{d}f_1]$ is a global codifferential mapping of f associated with the DC decomposition $f = |\lambda|h_1 - |\lambda|g_1$ in the case $\lambda < 0$;*

4. *if $f = \max_{i\in I} f_i$, then*

$$Df(\cdot) = \left[\mathrm{cl}\,\mathrm{co} \left\{ (f_i(\cdot) - f(\cdot), 0) + \underline{d}f_i(\cdot) - \sum_{j\neq i} \overline{d}f_j(\cdot) \,\Big|\, i \in I \right\}, \mathrm{cl}\left(\sum_{i=1}^{k} \overline{d}f_i(\cdot) \right) \right]$$

is a global codifferential mapping of f associated with the DC decomposition $f = \max_{i \in I}\{g_i + \sum_{j \neq i} h_j\} - \sum_{i=1}^{k} h_i$;

5. *if $f = \min_{i \in I} f_i$, then*

$$Df(\cdot) = \left[\operatorname{cl}\left(\sum_{i=1}^{k} \underline{d}f_i(\cdot) \right), \operatorname{cl}\operatorname{co}\left\{ (f_i(\cdot) - f(\cdot), 0) + \overline{d}f_i(\cdot) - \sum_{j \neq i} \underline{d}f_j(\cdot) \,\Big|\, i \in I \right\} \right]$$

is a global codifferential mapping of f associated with the DC decomposition $f = \sum_{i=1}^{k} g_i - \max_{i \in I}\{h_i + \sum_{j \neq i} g_j\}$.

Remark 5. Let us explain the presence of the terms $(f_i(\cdot) - f(\cdot), 0)$ in the expressions for global codifferentials of the functions $f = \max_{i \in I} f_i$ and $f = \min_{i \in I} f_i$ in the proposition above. The easiest way to see this is by computing the increment of the function f. Namely, let $k = 2$ and $f = \max\{f_1, f_2\}$. Then for any $x, \Delta x \in \mathcal{H}$ one has

$$f(x + \Delta x) - f(x) = \max\{f_1(x + \Delta x) - f(x), f_2(x + \Delta x) - f(x)\}$$
$$= \max\Bigg\{ f_1(x) - f(x) + \sup_{(a,v) \in \underline{d}f_1(x)} (a + \langle v, \Delta x \rangle) + \inf_{(b,w) \in \overline{d}f_1(x)} (b + \langle w, \Delta x \rangle),$$
$$f_2(x) - f(x) + \sup_{(a,v) \in \underline{d}f_2(x)} (a + \langle v, \Delta x \rangle) + \inf_{(b,w) \in \overline{d}f_2(x)} (b + \langle w, \Delta x \rangle) \Bigg\}.$$

Adding and subtracting the terms corresponding to $\overline{d}f_1(x)$ and $\overline{d}f_2(x)$ one obtains

$$f(x + \Delta x) - f(x)$$
$$= \max\Bigg\{ f_1(x) - f(x) + \sup_{(a,v) \in \underline{d}f_1(x)} (a + \langle v, \Delta x \rangle) - \inf_{(b,w) \in \overline{d}f_2(x)} (b + \langle w, \Delta x \rangle),$$
$$f_2(x) - f(x) + \sup_{(a,v) \in \underline{d}f_2(x)} (a + \langle v, \Delta x \rangle) - \inf_{(b,w) \in \overline{d}f_1(x)} (b + \langle w, \Delta x \rangle) \Bigg\}$$
$$+ \inf_{(b,w) \in \overline{d}f_1(x)} (b + \langle w, \Delta x \rangle) + \inf_{(b,w) \in \overline{d}f_2(x)} (b + \langle w, \Delta x \rangle).$$

which implies the required result. The interested reader can also verify this fact in a direct, but slightly more complicated way. Namely, define

$$g = \max\{g_1 + h_2, g_2 + h_1\}, \quad S_g = \operatorname{cl}\operatorname{co}\{S_{g_1} + S_{h_2}, S_{g_2} + S_{h_1}\},$$

and compute $\underline{d}f(x)$ with the use of (10) (cf. [21, Proposition 4.4, part (5)]).

Now we can turn to the study of global optimality conditions for DC optimization problems. At first, we obtain necessary and sufficient global optimality conditions for the unconstrained problem

$$\min_{x \in \mathcal{H}} f(x) \tag{$\mathcal{P}_0$}$$

in terms of a global codifferential of the function f.

Theorem 1. *Let f be a DC function, Df be any global codifferential of f, and $x_* \in \mathcal{H}$ be a given point. Suppose that f is bounded below, and a set $C \subseteq \overline{d}f(x_*)$*

is such that $\overline{d}f(x_*) = \mathrm{cl\,co}\,C$ *(in particular, if* $f = g - h$, *then one can set* $C = \{(h^*(v) + h(x) - \langle v, x \rangle, -v) \mid v \in \mathrm{dom}\, h^*\}$). *Then* x_* *is a globally optimal solution of the problem* $(\mathcal{P}_0)$ *if and only if for any* $z \in C$ *one has* $a(z) \geq 0$, *where* $(a(z), v(z))$ *is a globally optimal solution of the problem*

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a,v)\|^2 \quad \textit{subject to} \quad (a,v) \in \underline{d}f(x_*) + z.$$

Proof. From the definition of global codifferential mapping and the fact that $\overline{d}f(x_*) = \mathrm{cl\,co}\,C$ it follows that

$$f(x) - f(x_*) = \sup_{(a,v) \in \underline{d}f(x_*)} (a + \langle v, x - x_* \rangle) + \inf_{z \in C} (b + \langle w, x - x_* \rangle).$$

Consequently, x_* is a point of global minimum of f iff for any $z \in C$ one has

$$\sup_{(a,v) \in \underline{d}f(x_*)+z} (a + \langle v, x - x_* \rangle) \geq 0 \quad \forall x \in \mathcal{H}.$$

Note that the function on the left hand side of this inequality is bounded below by $\inf_{x \in \mathcal{H}} f(x) - f(x_*) > -\infty$. Hence applying the last part of Proposition 2 one obtains the desired result (see also Remark 2). $\qquad\square$

Corollary 1. *Let* f *be a DC function,* Df *be any global codifferential of* f, *and* $x_* \in \mathcal{H}$ *be a given point. Suppose that* f *is bounded above, and a set* $C \subseteq \underline{d}f(x_*)$ *is such that* $\underline{d}f(x_*) = \mathrm{cl\,co}\,C$. *Then* x_* *is a point of global maximum of the function* f *if and only if for any* $z \in C$ *one has* $b(z) \leq 0$, *where* $(b(z), w(z))$ *is a globally optimal solution of the problem*

$$\min_{(b,w) \in \mathbb{R} \times \mathcal{H}} \|(b,w)\|^2 \quad \textit{subject to} \quad (b,w) \in \overline{d}f(x_*) + z.$$

Remark 6. (i) From the proofs of Proposition 2 and Theorem 1 (see also Remark 2) it follows that if x_* is not a point of global minimum of the function f, then there exists $z \in C$ such that $a(z) < 0$, and for any such $z \in C$ one has $f(x_* + a(z)^{-1}v(z)) < f(x_*)$. Thus, the necessary and sufficient global optimality conditions from the theorem above not only allow one to verify whether a given point is a global minimizer, but also provide a way to compute "better" points, if the optimality conditions are not satisfied. Thus, it is fair to say that the global optimality conditions in terms of global codifferentials are somewhat constructive. Furthermore, it seems possible to propose a numerical method for general DC optimization problems based on the global optimality conditions from Thereom 1 and utilising a certain approximation of global codifferential (cf. [3, 57]).

(ii) It should be noted that the global optimality conditions from Theorem 1 (and part 4 of Proposition 2) were largely inspired by the codifferential calculus and the method of codifferential descent proposed by Demyanov [12, 13, 19, 21]. As was pointed out in [16], a nonsmooth function f is codifferentiable iff its increment can be locally approximated by a DC function. In the light of Theorem 1 one can say that in every iteration of the method of codifferential descent one verifies whether the global optimality conditions from Theorem 1 are satisfied, and then utilises "global descent" directions $v(z)$ of the DC approximation as line search directions for the objective function (see [21] for more details). Note that this observation partly explains the ability of the method of codifferential

13

descent to "jump over" some points of local minimum of the objective function (see [12, 21] for particular examples of this phenomenon).

(iii) It is obvious that in many particular cases the global optimality conditions from Theorem 1 are of theoretical value only, since it is extremely difficult to compute a global codifferential of a DC function. However, the same statement is true for many other general global optimality conditions. In particular, it is true for the well-known global optimality condition in terms of ε-subdifferentials [28–30] due to the fact that ε-subdifferentials can be efficiently computed only in few particular cases (see, e.g. [41]). Let us note that in the case when the function f is piecewise affine, there always exists a global codifferential of the function f such that both sets $\underline{d}f(x)$ and $\overline{d}f(x)$ are convex polytopes [27]. In this case, a global codifferential of the function f can be computed with the aid of Proposition 8. See [21] for applications of the optimality conditions from the theorem above to design and analysis of numerical methods for global optimization of nonconvex piecewise affine functions.

Let us give a simple example illustrating the use of the global optimality conditions from Theorem 1.

Example 5. Let $\mathcal{H} = \mathbb{R}$, $f(x) = \min\{2|x|, |x+2|+1\}$, and $x_0 = -2$. Let us check the optimality conditions at x_0. Note that x_0 is a point of strict local minimum of the function f, while a global minimum is attained at the point $x_* = 0$.

Denote $f_1(x) = 2|x|$ and $f_2(x) = |x+2|+1$. With the use of Proposition 8 one gets that

$$\underline{d}f_1(x_0) = \mathrm{co}\left\{\begin{pmatrix} 0 \\ -2 \end{pmatrix}, \begin{pmatrix} -8 \\ 2 \end{pmatrix}\right\}, \quad \overline{d}f_1(x_0) = \{0\},$$

$$\underline{d}f_2(x_0) = \mathrm{co}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right\}, \quad \overline{d}f_2(x_0) = \{0\}$$

(here the first coordinate is a, and the second one is v). Observe that unlike all subdifferentials, a codifferential is a pair of *two*-dimensional convex sets even in the one-dimensional case. Applying Proposition 8 again one obtains that

$$\underline{d}f(x_0) = \underline{d}f_1(x_0) + \underline{d}f_2(x_0) = \mathrm{co}\left\{\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \end{pmatrix}, \begin{pmatrix} -8 \\ 3 \end{pmatrix}, \begin{pmatrix} -8 \\ 1 \end{pmatrix}\right\},$$

and

$$\overline{d}f(x_0) = \mathrm{co}\left\{\begin{pmatrix} 3 \\ 0 \end{pmatrix} - \underline{d}f_2(x_0), -\underline{d}f_1(x_0)\right\} = \mathrm{co}\left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 8 \\ -2 \end{pmatrix}\right\}.$$

Let C be the set of extreme points of $\overline{d}f(x_0)$, i.e.

$$C = \left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 8 \\ -2 \end{pmatrix}\right\}.$$

Then one can easily verify that

1. $0 \in \underline{d}f(x_0) + z$ for $z = (3, 1) \in C$, $z = (0, 2) \in C$, and $z = (8, -2) \in C$;

2. $(a(z), v(z)) = (-0.2, -0.4)$ for $z = (3, -1) \in C$.

14

Thus, the global optimality conditions from Theorem 1 are not satisfied. Furthermore, note that for $z = (3, -1)$ one has $x(z) = x_0 + a(z)^{-1}v(z) = 0$, i.e. $x(z)$ is a point of global minimum of the function f.

Now we turn to constrained DC optimization problems. We start with the case of inequality constrained problems, since the presence of equality constraints significantly complicates the derivation of optimality conditions. Below, we largely follow Proposition 3, but do not apply it directly, since, as one can verify, a direct application of this theorem leads to more restrictive regularity assumptions on the constraints.

Consider the optimization problem of the form

$$\min_{x \in \mathcal{H}} f_0(x) \quad \text{subject to} \quad f_i(x) \le 0, \quad i \in I, \qquad (\mathcal{P}_I)$$

where f_i, $i \in 0 \cup I$, $I = \{1, \ldots, l\}$, are DC functions. Denote by Ω the feasible region of this problem. To obtain global optimality conditions for this problem we need to impose a regularity assumption on the constraints. Namely, one says that *the interior point constraint qualification* (IPCQ) holds at a point $x_0 \in \Omega$, if $x_0 \in \text{cl}\{x \in \mathcal{H} \mid f_i(x) < 0 \, i \in I\}$ or, equivalently, if for any $\varepsilon > 0$ there exists $y \in \Omega$ such that $\|y - x\| < \varepsilon$, and $f_i(y) < 0$ for all $i \in I$. It is easy to see that in the case when the functions f_i, $i \in I$, are convex, IPCQ is equivalent to Slater's condition. Note also that IPCQ holds at x_0, in particular, if a nonsmooth Mangasarian-Fromovitz constraint qualification (MFCQ) holds true at this point, i.e. if there exists $v \in \mathcal{H}$ such that $f_i'(x_0, v) < 0$ for all $i \in I$ such that $f_i(x_0) = 0$, where $f_i'(x_0, v)$ is the directional derivative of f_i at x_0 in the direction v. Finally, it should be noted that IPCQ is a generalization of the robustness condition from [34].

Theorem 2. *Let there exist a globally optimal solution of the problem $(\mathcal{P}_I)$ such that IPCQ holds true at this solution, and let x_* be a feasible point of $(\mathcal{P}_I)$. Let also the function f_0 be bounded below on Ω, and Df_i be a global codifferential of f_i, $i \in I \cup \{0\}$. Suppose, finally, that $C_i \subseteq \overline{d}f_i(x_*)$ is a nonempty set such that $\overline{d}f_i(x_*) = \text{cl co}\, C_i$, $i \in I \cup \{0\}$. Then x_* is a globally optimal solution of the problem $(\mathcal{P}_I)$ if and only if for any $z_i \in C_i$, $i \in I \cup \{0\}$, one has $a(z) \ge 0$, where $(a(z), v(z))$ with $z = (z_0, z_1, \ldots, z_l)$ is a globally optimal solution of the problem*

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a, v)\|^2 \quad \text{subject to} \quad (a, v) \in L(z)$$

and

$$L(z) = \text{cl co}\{\underline{d}f_0(x_*) + z_0, \, \underline{d}f_i(x_*) + z_i + (f_i(x_*), 0) \mid i \in I\}. \qquad (13)$$

Proof. Let us utilise a global version of the standard trick (see, e.g. the classic paper [35]) to transform the problem $(\mathcal{P}_I)$ into an unconstrained optimization problem. Introduce the function

$$F(x) = \max\{f_0(x) - f_0(x_*), f_1(x), \ldots, f_l(x)\}.$$

Note that $F(x_*) = 0$, since x_* is a feasible point of $(\mathcal{P}_I)$. Let us check that x_* is a globally optimal solution of the problem $(\mathcal{P}_I)$ iff it is a point of global minimum of the function F.

Indeed, suppose that x_* is a globally optimal solution of the problem $(\mathcal{P}_I)$. Observe that if $F(x) < 0$ for some $x \in \mathcal{H}$, then $x \in \Omega$ and $f_0(x) < f_0(x_*)$, which

is impossible. Thus, $F(x) \geq F(x_*) = 0$ for any $x \in \mathcal{H}$, i.e. x_* is a point of global minimum of the function F. Conversely, let x_* be a point of global minimum of F. By definition $F(x) \geq F(x_*) = 0$ for all $x \in \mathcal{H}$. Hence, in particular, for any x such that $f_i(x) < 0$ for all $i \in I$ one has $f_0(x) \geq f_0(x_*)$. Thus, x_* is a globally optimal solution of the problem

$$\min_{x \in \mathcal{H}} f_0(x) \quad \text{subject to} \quad x \in \{x_*\} \cup \{y \in \mathcal{H} \mid f_i(y) < 0 \, \forall i \in I\}.$$

Note that the function f_0 is continuous as the difference of finite closed convex functions that are continuous due to the fact that $\mathcal{H}$ is a Hilbert space (see, e.g. [23, Corollary I.2.5]). Therefore, taking into account the fact that by our assumption IPCQ holds true at some globally optimal solution y_* of the problem $(\mathcal{P}_I)$ one obtains that $f_0(x_*) \leq f_0(y_*)$, which implies that x_* is a globally optimal solution of $(\mathcal{P}_I)$ as well. Thus, x_* is a globally optimal solution of the problem $(\mathcal{P}_I)$ iff x_* is a point of global minimum of the function F.

From the definition of global codifferential it follows that

$$F(x) = \max_{i \in I} \left\{ \sup_{(a,v) \in \underline{d}f_0(x_*)} (a + \langle v, x - x_* \rangle) + \inf_{(b,w) \in \overline{d}f_0(x_*)} (b + \langle w, x - x_* \rangle), \right.$$
$$\left. f_i(x_*) + \sup_{(a,v) \in \underline{d}f_i(x_*)} (a + \langle v, x - x_* \rangle) + \inf_{(b,w) \in \overline{d}f_i(x_*)} (b + \langle w, x - x_* \rangle) \right\}.$$

Therefore, as is easy to see, x_* is a point of global minimum of the function F iff for any $z_i \in C_i$, $i \in I \cup \{0\}$ the function

$$F_z(x) = \max_{i \in I} \left\{ \sup_{(a,v) \in \underline{d}f_0(x_*)+z_0} (a + \langle v, x - x_* \rangle), \right.$$
$$\left. f_i(x_*) + \sup_{(a,v) \in \underline{d}f_i(x_*)+z_i} (a + \langle v, x - x_* \rangle) \right\}$$

is nonnegative. Note that $F_z(x) \geq F(x) \geq f_0(x) - f_0(x_*) \geq \inf_{x \in \Omega} f_0(x) - f_0(x_*) > -\infty$ for any $x \in \Omega$, and $F_z(x) \geq F(x) > 0$ for any $x \notin \Omega$, i.e. the function F_z is bounded below. Consequently, taking into account the fact that the set (13) is an affine support set of F_z, and applying the last part of Proposition 2 one obtains the desired result. $\qquad \square$

Remark 7. (i) As in the case of Theorem 1, the global optimality conditions from the theorem above are somewhat constructive. Namely, one can easily verify that if x_* is not a globally optimal solution of the problem $(\mathcal{P}_I)$, then for any $z_i \in C_i$, $i \in I \cup \{0\}$ such that $a(z) < 0$ (note that such z_i exist by Theorem 2) one has $f_0(x_* + a(z)^{-1}v(z)) < f_0(x_*)$ and $f_i(x_* + a(z)^{-1}v(z)) < 0$ for all $i \in I$. Thus, if x_* is not a globally optimal solution of the problem $(\mathcal{P}_I)$, then with the use of the global optimality conditions from Theorem 2 one can find a "better" point in the interior of the feasible region (see Example 6 below).
(ii) Note that if x_* is a point of global minimum of the function $F(x)$ defined in the proof of the theorem above, but IPCQ does not hold true at any globally optimal solution of the problem $(\mathcal{P}_I)$, then x_* need not be a globally optimal solution of this problem. For example, if $l = 2$, $f_1(x) = \|x\| - 1$, $f_2(x) = 1 - \|x\|$, then IPCQ does not hold true at any feasible point of $(\mathcal{P}_I)$, and any feasible point x_* is a global minimizer of $F(x)$. Thus, the validity of IPCQ is, in essence,

necessary for the validity of the global optimality conditions from the theorem above. Furthermore, this example shows that Theorem 2 cannot be applied to equality constrained problems, since IPCQ fails to hold true, if one rewrites an equality constraint $f_i(x) = 0$ as two inequality constraints $f_i(x) \leq 0$ and $-f_i(x) \leq 0$.

Let us give a simple example illustrating Theorem 2.

Example 6. Let $\mathcal{H} = \mathbb{R}$, and the problem $(\mathcal{P}_I)$ have the form

$$\min_{x \in \mathbb{R}} f_0(x) = |x - 4| \quad \text{subject to} \quad f_1(x) = \min\{|x - 2|, |x + 2|\} - 1 \leq 0. \quad (14)$$

Let also $x_0 = -1$. It is easily seen that $\Omega = [-3, -1] \cup [1, 3]$, IPCQ holds true at the unique globally optimal solution $x_* = 3$ of problem (14), and x_0 is a locally optimal solution of this problem. Let us check the global optimality conditions at the point x_0.

With the use of Proposition 8 one obtains that

$$\underline{d}f_0(x_0) = \mathrm{co}\left\{ \begin{pmatrix} -10 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}, \quad \overline{d}f_0(x_0) = \{0\},$$

$$\underline{d}f_1(x_0) = \mathrm{co}\left\{ \begin{pmatrix} -6 \\ 2 \end{pmatrix}, \begin{pmatrix} -8 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix} \right\},$$

$$\overline{d}f_1(x_0) = \mathrm{co}\left\{ \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Let $C_0 = \{0\}$, and C_1 be the set of extreme points of $\overline{d}f_1(x_0)$. Then applying Theorem 2 one can check that

1. $0 \in L(z)$ for $z = (z_0, z_1)$ with $z_0 = (0,0) \in C_0$, $z_1 = (4,1) \in C_1$, $z_1 = (6,-1) \in C_1$, and $z_1 = (0,1) \in C_1$;

2. $(a(z), v(z)) = (-0.1, -0.3)$ for $z = (z_0, z_1)$ with $z_0 = (0,0) \in C_0$ and $z_1 = (2,-1) \in C_1$.

Thus, the global optimality conditions from Theorem 2 are not satisfied. Furthermore, note that in the case $z_1 = (2,-1)$ one has $x_1 = x_0 + a(z)^{-1}v(z) = 2$, $f_0(x_1) = 2 < 5 = f_0(x_0)$ and $f_1(x_1) = -1 < 0$.

Now we turn to the general constrained optimization problem of the form

$$\min_{x \in \mathcal{H}} f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0, \quad i \in I, \quad f_j(x) = 0, \quad j \in J, \qquad (\mathcal{P}_{IJ})$$

where f_i, $i \in 0 \cup I \cup J$, $I = \{1, \ldots, l\}$, $J = \{l+1, \ldots, m\}$ are DC functions. Denote by Ω the feasible region of the problem $(\mathcal{P}_{IJ})$, and introduce the function

$$\varphi(x) = \sum_{i=1}^{l} \max\{0, f_i(x)\} + \sum_{j=l+1}^{m} |f_j(x)|.$$

Observe that $\Omega = \{x \in \mathcal{H} \mid \varphi(x) = 0\}$.

Our aim is to provide simple sufficient conditions under which the merit function $F_\lambda(x) = f_0(x) + \lambda\varphi(x)$ for the problem $(\mathcal{P}_{IJ})$ is globally exact in the finite dimensional case (note that this function is DC, if the problem $(\mathcal{P}_{IJ})$ is

DC). Apart from its direct applications to the design of numerical methods for solving the problem $(\mathcal{P}_{IJ})$, this result can also be used for the derivation of global optimality conditions for the problem $(\mathcal{P}_{IJ})$.

Recall that the function F_λ is said to be (globally) exact for the problem $(\mathcal{P}_{IJ})$, if there exists $\lambda_* \geq 0$ such that for any $\lambda \geq \lambda_*$ the set of globally optimal solutions of the problem $(\mathcal{P}_{IJ})$ coincides with the set of global minimizers of the function F_λ, i.e. the problem $(\mathcal{P}_{IJ})$ is equivalent (in terms of globally optimal solutions) to the penalized problem

$$\min_{x \in \mathcal{H}} F_\lambda(x) \qquad (\mathcal{P}_\lambda)$$

for any $\lambda \geq \lambda_*$. The greatest lower bound of all such λ_* is called *the least exact penalty parameter* of the function F_λ.

Theorem 3. *Let $\mathcal{H}$ be finite dimensional. Suppose that φ has a local error bound at every globally optimal solution of the problem $(\mathcal{P}_{IJ})$, i.e. for any globally optimal solution x_* of this problem there exist $\tau > 0$ and a neighbourhood U of x_* such that*

$$\varphi(x) \geq \tau \operatorname{dist}(x, \Omega) \quad \forall x \in U. \qquad (15)$$

Then the function F_λ is globally exact if and only if there exists $\lambda \geq 0$ such that the set $\{x \in \mathcal{H} \mid F_\lambda(x) < f_*\}$ is either bounded or empty, where f_* is the optimal value of the problem $(\mathcal{P}_{IJ})$. In particular, F_λ is globally exact, provided this function is bounded below for some $\lambda \geq 0$, and the set

$$C_\alpha = \{x \in \mathcal{H} \mid f_0(x) < f_* + \alpha, \ f_i(x) < \alpha, \ i \in I, \ |f_j(x)| < \alpha, \ j \in J\}$$

is bounded for some $\alpha > 0$.

Proof. Let x_* be a globally optimal solution of $(\mathcal{P}_{IJ})$. Note that the function f_0 is locally Lipschitz continuous, since it is a finite DC function. Consequently, taking into account (15) and applying [17, Thrm 2.4 and Prp. 2.7] one obtains that the function F_λ is locally exact at x_*, i.e. there exist $\lambda_*(x_*) \geq 0$ and a neighbourhood U of x_* such that $F_\lambda(x) \geq F_\lambda(x_*)$ for all $x \in U$ and $\lambda \geq \lambda_*(x_*)$. Then applying the localization principle for linear penalty functions (see [17, Thrm. 3.17] and [20, Thrm 4.1]) one gets that the function F_λ is globally exact if and only if there exists $\lambda \geq 0$ such that the set $\{x \in \mathcal{H} \mid F_\lambda(x) < f_*\}$ is either bounded or empty.

Suppose that F_{λ_0} is bounded below for some $\lambda_0 \geq 0$, and the set C_α is bounded for some $\alpha > 0$. Let us check that in this case $\{x \mid F_\lambda(x) < f_*\} \subset C_\alpha$ for any sufficiently large λ.

Indeed, if $x \notin C_\alpha$, then either $f_0(x) \geq f_* + \alpha$ or $\varphi(x) \geq \alpha$. In the former case one has $F_\lambda(x) \geq f_0(x) > f_*$ for any $\lambda \geq 0$, while in the latter case one has

$$F_\lambda(x) = F_{\lambda_0}(x) + (\lambda - \lambda_0)\varphi(x) \geq c + (\lambda - \lambda_0)\alpha > f_*$$

for all $\lambda > \lambda_0 + (f_* - c)/\alpha$, where $c = \inf_{x \in \mathcal{H}} F_{\lambda_0}(x)$. Thus, $\{x \mid F_\lambda(x) < f_*\} \subset C_\alpha$ for any $\lambda > \lambda_0 + (f_* - c)/\alpha$. $\qquad\square$

Remark 8. (i) Our proof of the global exactness of the ℓ_1 penalty function is based on the assumption that the penalty term $\varphi(x)$ has a local error bound. This assumption can be verified with the use of general results on metric subregularity and local error bounds [2, 26, 40]. In particular, in the case when the

functions f_i are continuously differentiable at a globally optimal solution x_* of $(\mathcal{P}_{IJ})$, the function $\varphi(x)$ has a local error bound at this optimal solution, provided MFCQ holds at x_* (see, e.g. [7, Corollary 2.2]). Let us note that in some cases it is possible to prove the existence of a local error bound with the use of the DC structure of the problem alone (i.e. without any constraint qualifications). See [43] for this kind of results on exact penalty functions and error bounds for DC optimization problems with inequality constraints.

(ii) Note that Theorem 3 significantly improves [55, Proposition 1], since we do not assume that the objective function f_0 is globally Lipschitz continuous, and utilise a local error bound instead of the global one in [55]. Furthermore, we obtained necessary and sufficient conditions for the global exactness of the function $F_\lambda(x)$, while only sufficient conditions were considered in [55].

Applying the global optimality conditions from Theorem 1 to $F_\lambda(x)$ one can easily obtain new necessary and sufficient global optimality conditions for the problem $(\mathcal{P}_{IJ})$ that are valid under the assumptions of Theorem 3. Namely, the following result holds true.

Theorem 4. *Let $\mathcal{H}$ be finite dimensional. Suppose that φ has a local error bound at every globally optimal solution of the problem $(\mathcal{P}_{IJ})$, and there exists $\lambda \geq 0$ such that the set $\{x \in \mathcal{H} \mid F_\lambda(x) < f_*\}$ is either bounded or empty. Suppose also that x_* is a feasible point of the problem $(\mathcal{P}_{IJ})$, Df_k is a global codifferential of the function f_k, $C_k \subseteq \overline{d}f_k(x_*)$ is a nonempty set such that $\overline{d}f_k(x_*) = \mathrm{cl}\,\mathrm{co}\,C_k$, $k \in I \cup J \cup \{0\}$, and $D_j \subseteq \underline{d}f_j(x_*)$ is a nonempty set such that $\underline{d}f_j(x_*) = \mathrm{cl}\,\mathrm{co}\,D_j$, $j \in J$. Then x_* is a globally optimal solution of the problem $(\mathcal{P}_{IJ})$ if and only if there exists $\lambda \geq 0$ such that for any $z_k \in C_k$, $k \in I \cup J \cup \{0\}$ and $u_j \in D_j$, $j \in J$, one has $a(z) \geq 0$, where $(a(z), v(z))$ with $z = (z_0, z_1, \ldots, z_m, u_{l+1}, \ldots, u_m)$ is a globally optimal solution of the problem*

$$\min_{(a,v)\in\mathbb{R}\times\mathcal{H}} \|(a,v)\|^2 \quad subject\ to \quad (a,v) \in Q_\lambda(z),$$

where

$$Q_\lambda(z) = \mathrm{cl}\left\{\underline{d}f_0(x_*) + \lambda\sum_{i\in I}\mathrm{co}\left\{(f_i(x_*),0) + \underline{d}f_i(x_*), -\overline{d}f_i(x_*) \mid i \in I\right\}\right.$$

$$\left. + \lambda\sum_{j\in J}\mathrm{co}\left\{\underline{d}f_j(x_*) + \underline{d}f_j(x_*), -\overline{d}f_j(x_*) - \overline{d}f_j(x_*) \mid j \in J\right\}\right\}$$

$$+ z_0 + \lambda\sum_{i\in I}z_i + \lambda\sum_{j\in J}(z_j - u_j).$$

Proof. As was noted above, F_λ is a DC function. With the use of Proposition 8 one can verify that the pair $DF_\lambda = [\underline{d}F_\lambda, \overline{d}F_\lambda]$ with

$$\underline{d}F_\lambda(x) = \mathrm{cl}\left\{\underline{d}f_0(x) + \lambda\sum_{i\in I}\mathrm{co}\left\{(f_i(x),0) + \underline{d}f_i(x), -\overline{d}f_i(x) \mid i \in I\right\}\right.$$

$$\left. + \lambda\sum_{j\in J}\mathrm{co}\left\{\underline{d}f_j(x) + \underline{d}f_j(x), -\overline{d}f_j(x) - \overline{d}f_j(x) \mid j \in J\right\}\right\},$$

$$\overline{d}F_\lambda(x) = \mathrm{cl}\left\{\overline{d}f_0(x) + \lambda\sum_{i\in I}\overline{d}f_i(x) + \lambda\sum_{j\in J}\left(\overline{d}f_j(x) - \underline{d}f_j(x)\right)\right\}$$

for any feasible point x is a global codifferential of F_λ. Observe also that

$$\overline{d}F_\lambda(x_*) = \text{cl co}\left\{C_0 + \lambda\sum_{i \in I} C_i + \lambda\sum_{j \in J}(C_j - D_j)\right\}$$

by the definitions of the sets C_k and D_j.

By Theorem 3 the function F_λ is globally exact. Therefore, if x_* is a globally optimal solution of the problem $(\mathcal{P}_{IJ})$, then it is a point of global minimum of F_λ for any sufficiently large λ. Now, applying the global optimality conditions from Theorem 1 to the function F_λ at x_* with λ large enough one obtains that the "only if" part of the theorem is valid.

Coversely, if there exists $\lambda \geq 0$ such that $a(z) \geq 0$ for any z from the formulation of the theorem, then by Theorem 1 the point x_* is a global minimizer of F_λ. Hence taking into account the facts that x_* is a feasible point of $(\mathcal{P}_{IJ})$ and $F_\lambda(x) = f_0(x)$ for any feasible point x of this problem one gets that x_* is a globally optimal solution of the problem $(\mathcal{P}_{IJ})$. Thus, the "if" part of the theorem is valid as well. $\square$

Let us consider two simple examples illustrating Theorems 3 and 4. The first example allows one to compare exact penalty approach with "interior point" approach from Theorem 2, while in the second example we analyse an equality constrained problem.

Example 7. Let us consider the same problem as in Example 6, i.e. the problem

$$\min_{x \in \mathbb{R}} f_0(x) = |x - 4| \quad \text{subject to} \quad f_1(x) = \min\{|x - 2|, |x + 2|\} - 1 \leq 0. \quad (16)$$

In this case the merit function F_λ has the form

$$F_\lambda(x) = |x - 4| + \lambda\max\left\{0, \min\{|x - 2|, |x + 2|\} - 1\right\}.$$

It is easily seen that this function is globally exact, and its least exact penalty parameter is equal to 1. We set $\lambda = 2$. Furthermore, one can check that for any $\lambda \geq 1$ the point $x_0 = -1$ is a local minimizer of F_λ, i.e. F_λ is locally exact at x_0. Let us apply the global optimality conditions from Theorem 1 to the function F_λ at the point x_0, as it is done in Theorem 4.

Let, as above, $\varphi(x) = \max\{0, f_1(x_0)\}$. Applying Proposition 8 one gets

$$D\varphi(x_0) = \left[\text{co}\left\{-\overline{d}f_1(x_0), \underline{d}f_1(x_0)\right\}, \overline{d}f_1(x_0)\right],$$
$$DF_2(x_0) = \left[\underline{d}f_0(x_0) + 2\underline{d}\varphi(x_0), \overline{d}f_0(x_0) + 2\overline{d}\varphi(x_0)\right]$$

Recall that global codifferentials of the functions f_0 and f_1 at x_0 were computed in Example 6. Therefore, with the use of Example 6 one gets that

$$\underline{d}F_2(x_0) = \text{co}\left\{\binom{-22}{5}, \binom{-26}{1}, \binom{-10}{1}, \binom{-14}{-3}, \binom{-14}{3}, \binom{-18}{-1}, \binom{-22}{3},\right.$$

$$\left.\binom{-10}{-1}, \binom{-12}{3}, \binom{-16}{-1}, \binom{0}{-1}, \binom{-4}{-5}, \binom{-4}{1}, \binom{-8}{-3}, \binom{-12}{1}, \binom{0}{-3}\right\},$$

and

$$\overline{d}F_2(x_0) = \text{co}\left\{\begin{pmatrix} 4 \\ -2 \end{pmatrix}, \begin{pmatrix} 8 \\ 2 \end{pmatrix}, \begin{pmatrix} 12 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}\right\}.$$

Let C be the set of extreme points of $\overline{d}F_2(0)$. Then one can check that

1. $0 \in \underline{d}F_0(x_0) + z$ for $z = (8, 2) \in C$, $z = (12, -2) \in C$, and $z = (0, 2) \in C$;

2. $(a(z), v(z)) = (-4/17, -16/17)$ for $z = (4, -2) \in C$.

Thus, by Theorem 1 the point x_0 is not a point of global minimum of the function $F_2(x)$ and, therefore, is not a globally optimal solution of problem (16). However, note that for $z = (4, -2)$ one has $x_1 = x_0 + a(z)^{-1}v(z) = -3$, i.e. x_1 is a globally optimal solution of problem (16) (cf. Example 6).

Example 8. Let $\mathcal{H} = \mathbb{R}^2$. Consider the following optimization problem:

$$\min_{x \in \mathbb{R}^2} f_0(x) = |x_1 - 2| + 2|x_2| \quad \text{subject to} \quad f_1(x) = |x_1| - |x_2| = 0. \quad (17)$$

The merit function F_λ for this problem has the form

$$F_\lambda(x) = |x_1 - 2| + 2|x_2| + \lambda\big||x_1| - |x_2|\big|.$$

It is easily seen that that the penalty term $\varphi(x) = \big||x_1| - |x_2|\big|$ has a local error bound at the unique globally optimal solution $x_* = (0, 0)$ of problem (17). Consequently, taking into account the fact that $f_0(x) \to +\infty$ as $\|x\| \to +\infty$ one obtains that the function F_λ is globally exact for problem (17). Let us estimate the least exact penalty parameter of F_λ.

One can easily verify that the function f_0 is globally Lipschitz continuous with Lipschitz constant $L = \sqrt{5}$, and

$$\varphi^\downarrow(x) = \liminf_{y \to x} \frac{\varphi(y) - \varphi(x)}{\|y - x\|} \le -1 \quad \forall x \notin \Omega,$$

where $\|\cdot\|$ is the Euclidean norm. The quantity $\varphi^\downarrow(x)$ is called *the rate of steepest descent* of φ at x (see, e.g. [11, 18]). For any $\lambda > \sqrt{5}$ and $x \notin \Omega$ one has $F_\lambda^\downarrow(x) \le L + \lambda\varphi^\downarrow(x) < 0$. Therefore, local/global minimizers of the function F_λ do not belong to the set $\mathbb{R}^2 \setminus \Omega$ for any $\lambda > \sqrt{5}$, since $F_\lambda^\downarrow(x) \ge 0$ is a necessary optimality condition. Thus, one can conclude that the least exact penalty parameter of F_λ does not exceed $\sqrt{5}$. That is why we set $\lambda = 3$.

Let us apply the global optimality conditions from Theorem 1 to the function F_λ at the point $x_0 = (2, 0)$, which is infeasible for problem (17) and is a point of unconstrained global minimum of the objective function f_0. With the use of Proposition 8 one obtains that

$$\underline{d}f_0(x_0) = \text{co}\left\{\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix}\right\}, \quad \overline{d}f_0(x_0) = \{0\},$$

$$\underline{d}f_1(x_0) = \text{co}\left\{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ -1 \\ 0 \end{pmatrix}\right\}, \quad \overline{d}f_1(x_0) = \text{co}\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}\right\}.$$

Furthermore, one has $DF_3(x_0) = [\underline{d}f_0(x_0) + 3\underline{d}\varphi(x_0), \overline{d}f_0(x_0) + 3\overline{d}\varphi(x_0)]$, where

$$\underline{d}\varphi(x_0) = \mathrm{co}\left\{\underline{d}f_1(x_0) + \underline{d}f_1(x_0), \begin{pmatrix} -4 \\ 0 \\ 0 \end{pmatrix} - \overline{d}f_1(x_0) - \overline{d}f_1(x_0)\right\},$$

and $\overline{d}\varphi(x_0) = \overline{d}f_1(x_0) - \underline{d}f_1(x_0)$. Utilising these expressions for global codifferentials one can easily compute $\underline{d}F_3(x_0)$, which is the convex hull of 20 points and we do not present it here for the sake of shortness, and check that

$$\overline{d}F_3(x_0) = \mathrm{co}\left\{\begin{pmatrix} 0 \\ -3 \\ 3 \end{pmatrix}, \begin{pmatrix} 12 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \\ -3 \end{pmatrix}, \begin{pmatrix} 12 \\ 3 \\ -3 \end{pmatrix}\right\}.$$

Let C be the set of extreme points of $\overline{d}F_3(x_0)$. Then solving the problem

$$\min_{(a,v)\in\mathbb{R}\times\mathbb{R}^2} \|(a,v)\|^2 \quad \text{subject to} \quad (a,v) \in \underline{d}F_3(x_0) + z$$

one can check that

1. $(a(z), v(z)) = (-1, 1.5, 0.5)$ for $z = (0, -3, 3) \in C$;

2. $(a(z), v(z)) = (-0.8, 1.6, 0)$ for $z = (12, 3, 3) \in C$ and $z = (12, 3, -3) \in C$;

3. $(a(z), v(z)) = (-1, 1.5, -0.5)$ for $z = (0, -3, -3) \in C$.

Thus, the global optimality conditions from Theorem 1 are not satisfeid at x_0. Moreover, observe that for $z = (12, 3, 3) \in C$ and $z = (12, 3, -3) \in C$ one has $x_1 = x_0 + a(z)^{-1}v(z) = (0, 0)$, and x_1 is a globally optimal solution of (17).

As was noted above, in many particular cases the global optimality conditions from Theorems 1, 2, and 4 are of theoretical value only, since it is extremely difficult (if at all possible) to compute a global codifferential of a general DC function and verify the global optimality conditions. Apparently, our optimality conditions can be readily checked only in the piecewise affine case, i.e. when the global codifferential is a pair of convex polytopes (see Example 3). Nevertheless, it seems possible to design new numerical methods for DC optimization problems based on the global optimality conditions obtained in this article and utilising certain polyhedral approximations of global codifferentials (cf. codifferential method in [3], aggregate codifferential method in [57], and a method based on successive piecewise-affine approximations in [25]). A design and analysis of such numerical methods lie outside the scope of this article. Here we only present a simple example of the usage of approximations of global codifferentials without trying to outline the idea behind possible numerical methods.

Let f be a DC function, and Df be its global codifferential associated with a DC decomposition $f = g - h$. From Proposition 4 and the definition of global codifferential (10) it follows that for any $x, y \in \mathcal{H}$, $v \in \partial g(y)$, and $w \in \partial h(y)$ one has $(g(y) - g(x) + \langle v, x - y \rangle, v) \in \underline{d}f(x)$ and $(h(x) - h(y) + \langle w, y - x \rangle, -w) \in \overline{d}f(x)$. Therefore, if a point x is fixed, one can choose sampling points $x_1, \ldots, x_m \in \mathcal{H}$, compute $v_k \in \partial g(x_k)$ and $w_k \in \partial h(x_k)$, $k \in \{1, \ldots, m\}$, and consider the following inner approximations of the global hypodifferential and the global hyperdifferential at x respectively:

$$\begin{aligned} &\mathrm{co}\left\{(g(x_k) - g(x) + \langle v_k, x - x_k \rangle, v_k) \mid k \in \{1, \ldots, m\}\right\} \subset \underline{d}f(x), \\ &\mathrm{co}\left\{(h(x) - h(x_k) + \langle w, x_k - x \rangle, -w_k) \mid k \in \{1, \ldots, m\}\right\} \subset \overline{d}f(x). \end{aligned} \tag{18}$$

The following example demonstrates that even if these inner approximations are very crude, one can still utilise them along with the global optimality conditions from Theorems 1, 2, and 4 to escape from a local minimum (or stationary point).

Example 9. Let $\mathcal{H} = \mathbb{R}^2$. Consider the following DC optimization problem:

$$\min_{x \in \mathbb{R}^2} f_0(x) = x_1^2 - x_2^2 \quad \text{subject to} \quad -1 \leq x_2 \leq 2. \tag{19}$$

We rewrite the constraints of this problem as follows: $f_1(x) = x_2 - 2 \leq 0$ and $f_2(x) = -x_2 - 1 \leq 0$. Let $x_0 = (0, -1)$. Clearly, x_0 is a locally (but not globally) optimal solution of problem (19). We would like to escape from this local minimum.

The merit function $F_\lambda(x) = f_0(x) + \lambda(\max\{0, x_2 - 2\} + \max\{0, -x_2 - 1\})$ for problem (19) is not bounded below for any $\lambda \geq 0$. Therefore, we will use the "interior point" approach of Theorem 2. With the use of Example 4 one obtains that the global codifferential $Df = [\underline{d}f, \overline{d}f]$ of the function f_0 associated with the DC decomposition $f_0 = g_0 - h_0$, where $g_0(x) = x_1^2$ and $h_0(x) = x_2^2$, has the form:

$$\underline{d}f_0(x) = \text{co}\left\{ \begin{pmatrix} -(x_1 - y_1)^2 \\ 2y_1 \\ 0 \end{pmatrix}, y_1 \in \mathbb{R} \right\}, \overline{d}f_0(x) = \text{co}\left\{ \begin{pmatrix} (x_2 - y_2)^2 \\ 0 \\ -2y_2 \end{pmatrix}, y_2 \in \mathbb{R} \right\}.$$

A direct usage of this global codifferential leads to rather cumbersome and complicated computations. That is why we will use inner approximations (18) instead. As sampling points we choose five points: x_0 and $x_0 + \xi_{ij}$, $i, j \in \{1, 2\}$, where $\xi_{ij} = ((-1)^i 2, (-1)^j 2)$ are the extereme points of the ball of radius 2 in the ℓ_∞ norm. Applying (18) one gets the following inner approximations:

$$\text{co}\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ -4 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ 4 \\ 0 \end{pmatrix} \right\} \subset \underline{d}f_0(x_0), \text{co}\left\{ \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \\ 6 \end{pmatrix} \right\} \subset \overline{d}f_0(x_0).$$

Note also that $Df_1(x_0) = [\{(0, 0, 1)\}, \{0\}]$ and $Df_2(x_0) = [\{(0, 0, -1)\}, \{0\}]$.

Let us apply the global optimality conditions from Theorem 2. In our case these conditions take the form: for any $z \in \overline{d}f_0(x_0)$ one has $a(z) \geq 0$, where $(a(z), v(z))$ is a globally optimal solution of the problem

$$\min_{(a,v) \in \mathbb{R} \times \mathbb{R}^2} \|(a, v)\|^2 \quad \text{s.t.} \quad (a, v) \in L(z) = \text{cl co}\left\{ \underline{d}f_0(x_0) + z, \underline{d}f_1(x_0), \underline{d}f_2(x_0) \right\}.$$

Replacing $\underline{d}f_0(x_0)$ with its inner approximation computed above one obtains that $(a(z), v(z)) \approx (-0.1034, 0, -0.2414)$ for $z = (4, 0, -2) \in \overline{d}f_0(x_0)$, i.e. the optimality conditions are not satisfied for the approximation. Following Remark 7 define $\overline{x} = x_0 + (a(z))^{-1}v(z) \approx (0, 1.333)$. Note that $\overline{x}$ belongs to the interior of the feasible region, and $f(\overline{x}) \approx -1.777 < -1 = f(x_0)$. Thus, the use of inner approximations (18) helped us escape from the local minimum.

4 Some connections between optimality conditions

Let us point out some connections between global optimality conditions obtained in the previous section, well-known global optimality conditions in terms of ε-subdifferentials, and KKT optimality conditions.

We start with global optimality conditions in terms of ε-subdifferentials. Let us consider the unconstrained DC optimization problem

$$\min_{x \in \mathcal{H}} f(x) = g(x) - h(x), \qquad (\mathcal{P}_0)$$

where g and h are finite closed convex functions. Suppose that f is bounded below. Recall that x_* is a point of global minimum of the function f if and only if

$$\partial_\varepsilon h(x_*) \subseteq \partial_\varepsilon g(x_*) \quad \forall \varepsilon \geq 0$$

(see [28]). On the other hand, by Theorem 1 the point x_* is a globally optimal solution of the problem $(\mathcal{P}_0)$ if and only if for any $z \in \overline{d}f(x_*)$ one has $a(z) \geq 0$, where $(a(z), v(z))$ is a globally optimal solution of the problem

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a,v)\|^2 \quad \text{subject to} \quad (a,v) \in \underline{d}f(x_*) + z,$$

and Df is a global codifferential of f associated with the DC decomposition $f = g - h$. Let us point out a direct connection between these optimality conditions.

Theorem 5. *Let $x_* \in \mathcal{H}$ and $\varepsilon_0 \geq 0$ be given. Then $\partial_\varepsilon h(x_*) \subseteq \partial_\varepsilon g(x_*)$ for any $\varepsilon \leq \varepsilon_0$ if and only if for any $z = (b, w) \in \overline{d}f(x_*)$ with $b \leq \varepsilon_0$ one has $a(z) \geq 0$.*

Proof. Fix any $z = (b, w) \in \overline{d}f(x_*)$ such that $b \leq \varepsilon$ for some $\varepsilon \geq 0$. By the last part of Proposition 2 and the definition of global codifferential (10) (see also (12)) one obtains that $a(z) \geq 0$ if and only if

$$g(x) - g(x_*) + b + \langle w, x - x_* \rangle \geq 0 \quad \forall x \in \mathcal{H}. \qquad (20)$$

Hence with the use of the inequality $b \leq \varepsilon$ one gets that $-w \in \partial_\varepsilon g(x_*)$. Observe also that

$$\partial_\varepsilon h(x_*) = \{w \in \mathcal{H} \mid \exists b \in [0, \varepsilon] \colon (b, -w) \in \overline{d}f(x_*)\}$$

by Proposition 1 and the definition of global codifferential.

Suppose that $a(z) \geq 0$ for any $z = (b, w) \in \overline{d}f(x_*)$ with $b \leq \varepsilon_0$, and fix any $\varepsilon \leq \varepsilon_0$ and $w \in \partial_\varepsilon h(x_*)$. Then $z = (b, -w) \in \overline{d}f(x_*)$ for some $b \leq \varepsilon$, which implies that $a(z) \geq 0$ and $w \in \partial_\varepsilon g(x_*)$. Thus, $\partial_\varepsilon h(x_*) \subseteq \partial_\varepsilon g(x_*)$ for any $\varepsilon \leq \varepsilon_0$.

Conversely, suppose that $\partial_\varepsilon h(x_*) \subseteq \partial_\varepsilon g(x_*)$ for any $\varepsilon \leq \varepsilon_0$. Choose any $z = (b, w) \in \overline{d}f(x_*)$ with $b \leq \varepsilon_0$. Then $-w \in \partial_b h(x_*)$, which implies that (20) holds true. Hence $a(z) \geq 0$ by Proposition 2, and the proof is complete. $\square$

Let us also consider the reverse convex minimization problem of the form:

$$\min_{x \in \mathcal{H}} f_0(x) \quad \text{subject to} \quad g(x) \geq 0. \qquad (\mathcal{P}_1)$$

Here f_0 and g are finite closed convex functions. We suppose that the feasible region of this problem is nonempty, and there exists an infeasible point x_0 such that

$$f_0(x_0) < \inf\{f_0(x) \mid x \in \mathcal{H} \colon g(x) \geq 0\}, \qquad (21)$$

i.e. the optimal value of the problem $(\mathcal{P}_1)$ is strictly grater than the infimum of f_0 over the entire space $\mathcal{H}$. Note that if $x_* \in \mathcal{H}$ is a globally optimal solution of $(\mathcal{P}_1)$ and $g(x) > 0$, then x_* is obviously a point of unconstrained local minimum of the function f_0, which due to the convexity of f_0 implies that x_* is a global

24

minimizer of f_0 and $0 \in \partial f_0(x_*)$. Therefore, below we suppose that all global minimizers x_* of $(\mathcal{P}_1)$ satisfy the equality $g(x_*) = 0$.

Recall that a feasible point x_* is a globally optimal solution of $(\mathcal{P}_1)$ if and only if $\partial_\varepsilon g(x_*) \subset \bigcup_{\alpha \geq 0} \partial_\varepsilon(\alpha f_0)(x_*)$ for all $\varepsilon \geq 0$ by [30, Theorem 3.5 and Remark 3.7]. On the other hand, rewriting the constraint $g(x) \geq 0$ as $f_1(x) \leq 0$ with $f_1(x) = -g(x)$ one can apply global optimality conditions from Theorem 2 to this problem. Let Df_0 be a global codifferential of f_0 associated with the DC decomposition $f_0 = f_0 - 0$, while Df_1 be a global codifferential of f_1 associated with the DC decomposition $f_1 = 0 - g$. Clearly, $\overline{d}f_0(\cdot) \equiv \{0\}$ and $\underline{d}f_1(\cdot) \equiv \{0\}$. Consequently, under the assumptions of Theorem 2 a feasible point x_* is a globally optimal solution of $(\mathcal{P}_1)$ if and only if for any $z \in \overline{d}f_1(x_*)$ one has $a(z) \geq 0$, where $(a(z), v(z))$ is a globally optimal solution of the problem

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a,v)\|^2 \quad \text{subject to} \quad (a,v) \in \operatorname{cl}\operatorname{co}\{\underline{d}f_0(x_*), z\}.$$

Let us describe how these two optimality conditions are connected.

Theorem 6. *Let a feasible point x_* of $(\mathcal{P}_1)$ and $\varepsilon_0 \geq 0$ be given. Then $\partial_\varepsilon g(x_*) \subset \bigcup_{\alpha \geq 0} \partial_\varepsilon(\alpha f_0)(x_*)$ for any $\varepsilon \leq \varepsilon_0$ if and only if for any $z = (b, w) \in \overline{d}f_1(x_*)$ with $b \leq \varepsilon_0$ one has $a(z) \geq 0$.*

Proof. Fix any $z = (b, w) \in \overline{d}f_1(x_*)$. By the last part of Proposition 2 and the definition of global codifferential (see (10) and (12)) one obtains that $a(z) \geq 0$ if and only if

$$\max\{f_0(x) - f_0(x_*), b + \langle w, x - x_* \rangle\} \geq 0 \quad \forall x \in \mathcal{H}. \tag{22}$$

Define $C_0 = \{x \in \mathcal{H} \mid f_0(x) < f_0(x_*)\}$ and $C = \{x \in \mathcal{H} \mid f_0(x) \leq f_0(x_*)\}$. From (21) it follows that C_0 is nonempty, while by [31, Proposition VI.1.3.3] one has $C = \operatorname{cl} C_0$. Clearly, inequality (22) is satisfied iff $b + \langle w, x - x_* \rangle \geq 0$ for any $x \in C_0$. In turn, this inequality is satisfied iff $\langle w, x - x_* \rangle \geq -b$ for all $x \in C$ due to the fact that $C = \operatorname{cl} C_0$. Thus, $a(z) \geq 0$ for some $z = (b, w) \in \overline{d}f_1(x_*)$ iff $\langle -w, x - x_* \rangle \leq b$ for all $x \in C$ or equivalently $-w \in N_b(C, x_*)$, where $N_b(C, x_*)$ is the set of b-normal directions to the set C at x_* (see, e.g. [32, Definition XI.1.1.3]). Note also that by Proposition 1 and the definition of global codifferential (10) for any $\varepsilon \geq 0$ one has $\partial_\varepsilon g(x_*) = \{w \in \mathcal{H} \mid \exists b \in [0, \varepsilon] \colon (b, -w) \in \overline{d}f_1(x_*)\}$.

Suppose that for any $z = (b, w) \in \overline{d}f_1(x_*)$ with $b \leq \varepsilon_0$ one has $a(z) \geq 0$, and fix any $\varepsilon \leq \varepsilon_0$ and $w \in \partial_\varepsilon g(x_*)$. Then $z = (b, -w) \in \overline{d}f(x_*)$ for some $b \leq \varepsilon$, which implies that $a(z) \geq 0$ and $w \in N_b(C, x_*) \subseteq N_\varepsilon(C, x_*)$. Applying [32, Corollary XI.3.6.2] one gets that $N_\varepsilon(C, x_*) = \bigcup_{\alpha \geq 0} \partial_\varepsilon(\alpha f_0)(x_*)$. Thus, for any $\varepsilon \leq \varepsilon_0$ one has $\partial_\varepsilon g(x_*) \subset \bigcup_{\alpha \geq 0} \partial_\varepsilon f_0(x_*)$.

Conversely, suppose that $\partial_\varepsilon g(x_*) \subset \bigcup_{\alpha \geq 0} \partial_\varepsilon(\alpha f_0)(x_*)$ for any $\varepsilon \leq \varepsilon_0$. Choose any $z = (b, w) \in \overline{d}f_1(x_*)$ with $b \leq \varepsilon_0$. Then $-w \in \partial_b g(x_*) \subset \bigcup_{\alpha \geq 0} \partial_b(\alpha f_0)(x_*) = N_b(C, x_*)$, which, as we proved above, is equivalent to the inequality $a(z) \geq 0$. $\square$

The theorem above can be further extended to the case of convex maximization problems of the form:

$$\max_{x \in \mathcal{H}} f(x) \quad \text{subject to} \quad x \in C. \tag{$\mathcal{P}_{\max}$}$$

Here $C = \{x \in \mathcal{H} \mid f_i(x) \leq 0, \ i \in I = \{1, \ldots, l\}\}$, and f and f_i, $i \in I$, are finite closed convex functions. We suppose that Slater's condition holds true, and the infimum of f over C is strictly smaller than the maximum.

Recall that a feasible point x_* is a globally optimal solution of $(\mathcal{P}_{\max})$ if and only if $\partial_\varepsilon f(x_*) \subset N_\varepsilon(C, x_*)$ for all $\varepsilon \geq 0$ by [28, Proposition 3.9]. Recasting the problem $(\mathcal{P}_{\max})$ as the problem of minimizing the function $f_0(x) = -f(x)$ over C one can apply global optimality conditions from Theorem 2 to this problem. Let Df_0 be a global codifferential of f_0 associated with the DC decomposition $f_0 = 0 - f$, while Df_i be a global codifferential of f_i associated with the DC decomposition $f_i = f_i - 0$, $i \in I$. Then $\underline{d}f_0(\cdot) \equiv \{0\}$ and $\overline{d}f_i(\cdot) \equiv \{0\}$, $i \in I$. Therefore, by Theorem 2 a feasible point x_* is a globally optimal solution of $(\mathcal{P}_{\max})$ if and only if for any $z = (b, w) \in \overline{d}f_0(x_*)$ one has $a(z) \geq 0$, where $(a(z), v(z))$ is a globally optimal solution of the problem

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a,v)\|^2 \quad \text{subject to} \quad (a,v) \in \operatorname{cl}\operatorname{co}\{z, \underline{d}f_i(x_*) + (f_i(x_*), 0) \mid i \in I\}.$$

From Proposition 2 and the definition of global codifferential (10) it follows that $a(z) \geq 0$ iff

$$\max\{b + \langle w, x - x_* \rangle, f_1(x), \ldots, f_l(x)\} \geq 0 \quad \forall x \in \mathcal{H}.$$

In turn, this inequality is satisfied iff $-w \in N_b(C, x_*)$. Utilising this result and arguing in the same way as in the proof of Theorem 6 one can easily check that the following connection between the two global optimality conditions for the problem $(\mathcal{P}_{\max})$ exists.

Theorem 7. *Let a feasible point x_* of $(\mathcal{P}_{\max})$ and $\varepsilon_0 \geq 0$ be given. Then $\partial_\varepsilon f(x_*) \subset N_\varepsilon(C, x_*)$ for any $\varepsilon \leq \varepsilon_0$ if and only if for any $z = (b, w) \in \overline{d}f_0(x_*)$ with $b \leq \varepsilon_0$ one has $a(z) \geq 0$.*

Thus, one can say that there is an intimate relation between global optimality conditions for DC optimization problems in terms of global codifferentials and in terms of ε-subdifferentials.

Now we turn to KKT optimality conditions. For the sake of simplicity, let us consider the inequality constrained problem

$$\min_{x \in \mathcal{H}} f_0(x) \quad \text{subject to} \quad f_i(x) \leq 0, \quad i \in I = \{1, \ldots, l\}, \qquad (\mathcal{P}_I)$$

where $f_i = g_i - h_i$ are DC functions such that the convex functions g_i and h_i are differentiable, $i \in I \cup \{0\}$. Let Df_i be a global codifferential of f_i associated with the DC decomposition $f_i = g_i - h_i$, $i \in I \cup \{0\}$. Recall that under the assumptions of Theorem 2 a feasible point x_* is a globally optimal solution of the problem $(\mathcal{P}_I)$ if and only if for any $z_i \in \overline{d}f_i(x_*)$, $i \in I \cup \{0\}$, one has $a(z) \geq 0$, where $(a(z), v(z))$ with $z = (z_0, z_1, \ldots, z_l)$ is a globally optimal solution of the problem

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a,v)\|^2 \quad \text{subject to} \quad (a,v) \in L(z)$$

with

$$L(z) = \operatorname{cl}\operatorname{co}\{\underline{d}f_0(x_*) + z_0, \ \underline{d}f_i(x_*) + z_i + (f_i(x_*), 0) \mid i \in I\}. \qquad (23)$$

(in the case $a(z) = 0$ one actually has $0 \in L(z)$; see Proposition 2 and Remark 2). Let us show that in the case when $z_i = (b_i, w_i) \in \overline{d}f_i(x_*)$ are such that $b_i = 0$, these optimality conditions are closely connected to the KKT optimality conditions.

Theorem 8. *Let x_* be a feasible point of the problem $(\mathcal{P}_I)$, the function f_0 be bounded below on the feasible region of this problem, and let MFCQ hold at x_*, i.e. there exists $y \in \mathcal{H}$ for which $\langle \nabla f_i(x_*), y \rangle < 0$ for any $i \in I$ such that $f_i(x_*) = 0$. Then KKT optimality conditions hold true at x_* if and only if for any $z_i = (b_i, w_i) \in \overline{d} f_i(x_*)$ with $b_i = 0$, $i \in I \cup \{0\}$, one has $0 \in L(z)$, where $z = (z_0, z_1, \ldots, z_l)$.*

Proof. Observe that $z_i = (0, w_i) \in \overline{d} f_i(x_*)$, $i \in I \cup \{0\}$, if and only if $w_i = -\nabla h_i(x_*)$ by Proposition 1 and the definition of global codifferential (10). Therefore, let $z = (z_0, z_1, \ldots, z_l)$ with $z_i = (0, -\nabla h_i(x_*)) \in \overline{d} f_i(x_*)$ for all $i \in I \cup \{0\}$.

From (23) and (12) it follows that the set $L(z)$ is an affine support set of the function

$$F(x) = \max_{i \in I} \big\{ g_0(x) - g_0(x_*) - \langle \nabla h_0(x_*), x - x_* \rangle,$$
$$g_i(x) - g_i(x_*) + f_i(x_*) - \langle \nabla h_i(x_*), x - x_* \rangle \big\}.$$

Clearly, $F(x_*) = 0$. Moreover, from the convexity of the functions h_i it follows that $h_i(x) - h_i(x_*) \geq \langle \nabla h_i(x_*), x - x_* \rangle$, which implies that for all $x \in \mathcal{H}$ the inequality $F(x) \geq \max_{i \in I}\{f_0(x) - f_0(x_*), f_i(x)\}$ holds true. Consequently, for any feasible point x one has $F(x) \geq f_0(x) - f_0(x^*)$, and $F(x) > 0$ otherwise, i.e. the function F is bounded below, since f_0 is bounded below on the feasible region by our assumption. Therefore, by the last part of Proposition 2 one gets that $0 \in L(z)$ if and only if $F(x) \geq 0$ for all $x \in \mathcal{H}$. Hence taking into account the facts that F is a convex function and $F(x_*) = 0$ one obtains that $0 \in L(z)$ iff $0 \in \partial F(x_*)$.

Bearing in mind the fact that x_* is a feasible point of the problem $(\mathcal{P}_I)$, and applying the well-known formula for the subdifferential of the maximum of a finite family of convex functions (see, e.g. [31, Corollary VI.4.3.2]) one obtains that

$$\partial F(x_*) = \text{co}\big\{\nabla f_0(x_*), \nabla f_i(x_*) \,\big|\, i \in I \colon f_i(x_*) = 0\big\}.$$

Thus, $0 \in L(z)$ if and only if there exist $\alpha_i \geq 0$, $i \in I \cup \{0\}$, such that

$$\sum_{i=0}^{l} \alpha_i \nabla f_i(x_*) = 0, \quad \sum_{i=0}^{l} \alpha_i = 1,$$

and $\alpha_i = 0$ whenever $f_i(x_*) < 0$. Note that $\alpha_0 \neq 0$ due to the fact that MFCQ holds at x_*, since otherwise $0 = \sum_{i=1}^{l} \alpha_i \langle \nabla f_i(x_*), y \rangle < 0$, which is impossible. Hence dividing by α_0 and denoting $\lambda_i = \alpha_i / \alpha_0$ one obtains that $0 \in L(z)$ iff there exist $\lambda_i \geq 0$, $i \in I$, such that

$$\nabla f_0(x_*) + \sum_{i=1}^{l} \lambda_i \nabla f_i(x_*) = 0, \quad \lambda_i f_i(x_*) = 0, \quad \lambda_i \geq 0 \quad \forall i \in I,$$

i.e. if and only if KKT optimality conditions are satisfied at x_*. $\qquad\square$

5   A problem of Bolza

In some applications it might be extremely difficult to solve the problem

$$\min_{(a,v) \in \mathbb{R} \times \mathcal{H}} \|(a, v)\|^2 \quad \text{subject to } (a, v) \in \underline{d} f(x_*) + z \qquad (24)$$

in order to find $(a(z), v(z))$, which renders the global optimality conditions presented above useless. The aim of this section is to demonstrate that in this case one can utilise different global optimality condition in terms of global codifferentials. Below we derive these conditions and apply them to a nonsmooth problem of Bolza.

Theorem 9. *Let f be a DC function, Df be any global codifferential of f, and $x_* \in \mathcal{H}$ be a given point. Suppose that $C \subseteq \overline{d}f(x_*)$ is a nonempty set such that $\overline{d}f(x_*) = \mathrm{cl}\,\mathrm{co}\,C$. Then x_* is a point of global minimum of the function f if and only if for any $z \in C$ there exists $\xi(z) \geq 0$ such that $(\xi(z), 0) \in \underline{d}f(x_*) + z$.*

Proof. Clearly, x_* is a point of global minimum of the function f iff for any $z \in C$ the function

$$G_z(x) = \sup_{(a,v)\in \underline{d}f(x_*)+z} (a + \langle v, x \rangle)$$

is nonnegative (note that $G_z(x) \geq f(x_* + x) - f(x_*)$). Applying the second part of Proposition 2 one obtains that if x_* is a point of global minimum, then for any $z \in C$ one has $(\xi(z), 0) \in \underline{d}f(x_*) + z$, where $\xi(z) = \inf_{x\in\mathcal{H}} G_z(x) \geq 0$. Conversely, if for any $z \in C$ there exists $\xi(z) \geq 0$ such that $(\xi(z), 0) \in \underline{d}f(x_*)+z$, then $\inf_{x\in\mathcal{H}} G_z(x) \geq \xi(z) \geq 0$, and x_* is a point of global minimum. $\square$

With the use of the first part of Proposition 2 and the fact that by the definition of global codifferential $f(x) - f(x_*) = \inf_{z\in C} G_z(x - x_*)$ for all $x \in \mathcal{H}$ one can easily obtain the following result.

Theorem 10. *Let f be a DC function, Df be any global codifferential of f, and $x_* \in \mathcal{H}$ be a given point. Suppose that $C \subseteq \overline{d}f(x_*)$ is a nonempty set such that $\overline{d}f(x_*) = \mathrm{cl}\,\mathrm{co}\,C$. Then f is bounded below if and only if there exists $\xi \in \mathbb{R}$ such that for any $z \in C$ one has $([\xi, +\infty) \times \{0\}) \cap (\underline{d}f(x_*) + z) \neq \emptyset$.*

Let us present an example in which global optimality conditions from Section 3 become too complicated and unverifiable, while Theorems 9 and 10 can be easily applied. This example also demonstrates how one can compute a global codifferential of a variational functional (cf. [15]). Namely, let us analyse the following nonsmooth problem of Bolza:

$$\min \mathcal{I}(u) = u(0) - e^{-1}u(1) + \int_0^1 \max\{|u'(x)| - |u(x)|, 0\}\, dx. \qquad (25)$$

Here u is from the Sobolev space $W^{1,1}(0,1)$. As was demonstrated in [15, 36], the function $u_*(x) = \theta e^x$ with $\theta > 0$ satisfies several necessary optimality conditions for problem (25). Our main goal is to demonstrate that this solution is not globally optimal. Furthermore, we will show that the functional $\mathcal{I}(\cdot)$ is unbounded below and thus does not attain a global minimum.

To convert the problem to the Hilbert space setting, below we suppose that $u \in H^1(0,1) = W^{1,2}(0,1)$. Clearly, if u_* is not a globally optimal solution in $H^1(0,1)$, then it is not a globally optimal solution in $W^{1,1}(0,1)$. Let us compute a global codifferential mapping of the restriction of the functional $\mathcal{I}$ to the Hilbert space $H^1(0,1)$. To this end, for any $x \in [0,1]$ introduce the function

$$f_x(u, \xi) = \max\{|\theta e^x + \xi| - |\theta e^x + u|, 0\} = \max\{|\theta e^x + \xi|, |\theta e^x + u|\} - |\theta e^x + u|.$$

Applying Proposition 8 one obtains that the pair $Df_x(0,0) = [\underline{d}f_x(0,0), \overline{d}f_x(0,0)]$ with

$$\underline{d}f_x(0,0) = \mathrm{co}\left\{\begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}-2\theta e^x\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}-2\theta e^x\\0\\-1\end{pmatrix}\right\}, \qquad (26)$$

$$\overline{d}f_x(0,0) = \mathrm{co}\left\{\begin{pmatrix}0\\-1\\0\end{pmatrix}, \begin{pmatrix}2\theta e^x\\1\\0\end{pmatrix}\right\}. \qquad (27)$$

is a global codifferential of f_x at $(0,0)$. Then by the definition of global codifferential and the fact that $f_x(0,0) = 0$ one gets that

$$\mathcal{I}(u_* + u) - \mathcal{I}(u_*) = u(0) - e^{-1}u(1) + \int_0^1 \Big(\max_{(a,v)\in\underline{d}_x f(0,0)} \big(a + v_1 u(x) + v_2 u'(x)\big)$$

$$+ \min_{(b,w)\in\overline{d}_x f(0,0)} \big(b + w_1 u(x) + w_2 u'(x)\big) \Big) dx$$

for any $u \in H^1(0,1)$. Clearly, the mapping $x \mapsto \underline{d}f_x(0,0)$ is measurable. Therefore, by the Filippov Theorem (see, e.g. [1, Thrm 8.2.10]) for any $u \in H^1(0,1)$ there exists a measurable selection $(a(x), v_1(x), v_2(x))$ of the map $x \mapsto \underline{d}f_x(0,0)$ such that

$$\max_{(a,v)\in\underline{d}f_x(0,0)} (a + v_1 u(x) + v_2 u'(x)\rangle) = a(x) + v_1(x)u(x) + v_2(x)u'(x)$$

for a.e. $x \in (0,1)$. Hence for any $u \in H^1(0,1)$ one has

$$\mathcal{I}(u_* + u) - \mathcal{I}(u_*) = u(0) - e^{-1}u(1) + \max\Big(\int_0^1 \big(a(x) + v_1(x)u(x) + v_2(x)u'(x)\big)\, dx\Big)$$

$$+ \min\Big(\int_0^1 \big(b(x) + w_1(x)u(x) + w_2(x)u'(x)\big)\, dx\Big), \quad (28)$$

where the maximum is taken over all measurable selections of the map $x \mapsto \underline{d}f_x(0,0)$, and the minimum is taken over all measurable selections of the map $x \mapsto \overline{d}f_x(0,0)$.

Recall that $u \in H^1(0,1)$ iff $u(x) = u(0) + \int_0^x \eta(s)ds$ for some $\eta \in L_2(0,1)$ (see, e.g. [44]). Therefore, instead of $\mathcal{I}(u)$ one can consider the functional $\mathcal{J}\colon \mathbb{R} \times L_2(0,1) \to \mathbb{R}$ defined as $\mathcal{J}(u_0, \eta) = \mathcal{I}(u)$, where $u(x) = u_0 + \int_0^x \eta(s)\, ds$. Denote $\eta_*(x) = \theta e^x$. Applying (28) and integrating by parts one obtains that

$$\mathcal{J}(\theta + u_0, \eta_* + \eta) - \mathcal{J}(\theta, \eta_*) = \max_{(A,v_0,v)\in\underline{d}\mathcal{J}(\theta,\eta_*)} \big(A + v_0 u_0 + \langle v, \eta\rangle\big)$$

$$+ \min_{(B,w_0,w)\in\overline{d}\mathcal{J}(\theta,\eta_*)} \big(B + w_0 u_0 + \langle w, \eta\rangle\big),$$

where $\langle v, \eta\rangle = \int_0^1 v(x)\eta(x)\, dx$ is the inner product in $L_2(0,1)$,

$$\underline{d}\mathcal{J}(\theta, \eta_*) = \Big\{(A, v_0, v) \in \mathbb{R} \times \mathbb{R} \times L_2(0,1) \,\Big|\, A = \int_0^1 a(x)\, dx,$$

$$v_0 = 1 - e^{-1} + \int_0^1 v_1(x)\, dx, \ v(x) = \int_x^1 v_1(s)\, ds + v_2(x) - e^{-1},$$

$$(a(\cdot), v_1(\cdot), v_2(\cdot)) \text{ is a measurable selection of the map } x \mapsto \underline{d}f_x(0,0)\Big\},$$

and

$$\overline{d}\mathcal{J}(\theta, \eta_*) = \left\{ (B, w_0, w) \in \mathbb{R} \times \mathbb{R} \times L_2(0,1) \ \middle| \ B = \int_0^1 b(x)\, dx, \right.$$

$$w_0 = \int_0^1 w_1(x)\, dx, \ w(x) = \int_x^1 w_1(s)\, ds + w_2(x),$$

$$\left. (b(\cdot), w_1(\cdot), w_2(\cdot)) \text{ is a measurable selection of the map } x \mapsto \overline{d}f_x(0,0) \right\}.$$

The sets $\underline{d}\mathcal{J}(\theta, \eta_*)$ and $\overline{d}\mathcal{J}(\theta, \eta_*)$ are obviously convex. Let us verify that they are closed. For the sake of shortness, we consider only the set $\underline{d}\mathcal{J}(\theta, \eta_*)$.

Note that the set K of measurable selections of the map $x \mapsto \underline{d}f_x(0,0)$ is obviously convex, closed and bounded in $L_2(0,1)$. Therefore it is weakly compact in $L_2(0,1)$. It is easily seen that $\underline{d}\mathcal{J}(\theta, \eta_*)$ is the image of the set K under a continuous map from the space $L_2(0,1)$ endowed with the weak topology to the space $\mathbb{R} \times \mathbb{R} \times L_2(0,1)$ endowed with the weak topology as well. Hence the set $\underline{d}\mathcal{J}(\theta, \eta_*)$ is weakly compact, which implies that it is closed in the norm topology due to the fact that this set is convex.

Thus, the pair $D\mathcal{J}(\theta, \eta_*) = [\underline{d}\mathcal{J}(\theta, \eta_*), \overline{d}\mathcal{J}(\theta, \eta_*)]$ is a global codifferential of $\mathcal{J}$ at the point (θ, η_*). Let us verify that this point is not a global minimizer of $\mathcal{J}$ with the use of Theorem 9.

Remark 9. It should be noted that a direct application of the global optimality conditions from Theorem 1 to problem (25) is very difficult, since it is unclear how to compute points $(a(z), v(z))$ defined in (24) for this problem.

The mapping $(b(x), w_1(x), w_2(x)) = (2\theta e^x, 1, 0)$ is a measurable selection of the map $x \mapsto \overline{d}f_x(0,0)$ (see (27)). Therefore, the point $z_* = (2\theta(e-1), 1, w(\cdot))$ with $w(x) \equiv 1 - x$ belongs to $\overline{d}\mathcal{J}(\theta, \eta_*)$. With the use of (26) and the Filippov Theorem one can easily check that any measurable selection $(a(\cdot), v_1(\cdot), v_2(\cdot))$ of the map $x \mapsto \underline{d}f_x(0,0)$ has the form

$$\begin{pmatrix} a(x) \\ v_1(x) \\ v_2(x) \end{pmatrix} = \begin{pmatrix} -2\theta(\alpha_2(x) + \alpha_4(x))e^x \\ \alpha_1(x) - \alpha_2(x) \\ \alpha_3(x) - \alpha_4(x) \end{pmatrix}$$

for a.e. $x \in [0,1]$ and for some $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) \in S_4$, where the set $S_4 \subset (L_2(0,1))^4$ consists of all those $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ for which all α_i are nonnegative and $\alpha_1(x) + \alpha_2(x) + \alpha_3(x) + \alpha_4(x) = 1$ for a.e. $x \in (0,1)$. Therefore $[A, v_0, v] \in \underline{d}\mathcal{J}(\theta, \eta_*)$ iff there exists $\alpha \in S_4$ such that

$$A = -2\theta \int_0^1 e^x(\alpha_2(x) + \alpha_4(x))\, dx, \quad v_0 = 1 - e^{-1} + \int_0^1 (\alpha_1(x) - \alpha_2(x))\, dx,$$

$$v(x) = -e^{-1} + \int_x^1 (\alpha_1(s) - \alpha_2(s))\, ds + \alpha_3(x) - \alpha_4(x)$$

for a.e. $x \in [0,1]$. Consequently, the point $(\xi, 0, 0)$ belongs to $\underline{d}\mathcal{J}(\theta, \eta_*) + z_*$ for

some $\xi \in \mathbb{R}$ iff there exists $\alpha \in S_4$ such that

$$-2\theta \int_0^1 e^x(\alpha_2(x) + \alpha_4(x)) \, dx + 2\theta(e-1) = \xi,$$

$$\int_0^1 (\alpha_1(x) - \alpha_2(x)) \, dx + 2 - e^{-1} = 0,$$

$$\int_x^1 (\alpha_1(s) - \alpha_2(s)) \, ds + \alpha_3(x) - \alpha_4(x) - e^{-1} + 1 - x = 0$$

for a.e. $x \in [0, 1]$. However, note that

$$\int_0^1 (\alpha_1(x) - \alpha_2(x)) \, dx \geq -\int_0^1 \alpha_2(x) \, dx \geq -1 > e^{-1} - 2$$

due to the fact that $\alpha_1(x) \geq 0$ and $\alpha_2(x) \leq 1$ for a.e. $x \in [0, 1]$. Thus, the sets $\mathbb{R} \times \{0\} \times \{0\}$ and $\underline{d}\mathcal{J}(\theta, \eta_*) + z_*$ do not intersect, which by Theorems 9 and 10 implies that the pair $(\theta, \eta_*(\cdot))$ with $\eta_*(x) = \theta e^x$ is not a point of global minimum of $\mathcal{J}$, and this functional is unbounded below. Consequently, the function $u_*(x) = \theta e^x$ is not a global minimizer of $\mathcal{I}$, and this functional is unbounded below as well.

6   Conclusions

In this article we obtained new necessary and sufficient global optimality conditions for DC optimization problems in terms of global codifferentials. These optimality conditions are closely related to the method of codifferential descent and are somewhat constructive, in the sense that they allow one to find "global descent" directions at non-optimal points. On the other hand, a direct usage of the global optimality conditions requires the knowledge of a global codifferential of a DC function, and global codifferentials can be relatively easily computed (and manipulated with) only in the piecewise affine case. Nevertheless, it seems possible to propose new methods for general DC optimization problems utilising an approximation of global codifferential (cf. codifferential method [3] and aggregate codifferential method [57]). A development and analysis of such methods are interesting topics of future research.

References

[1] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhauser, Boston, 1990.

[2] D. Azé. A unified theory for metric regularity of multifunctions. *J. Convex Anal.*, 13:225–252, 2006.

[3] A. M. Bagirov and J. Ugon. Codifferential method for minimizing nonsmooth DC functions. *J. Glob. Optim.*, 50:3–22, 2011.

[4] G. Bigi, A. Frangioni, and Q. Zhang. Approximate optimality conditions and stopping criteria in canonical DC programming. *Optim. Methods Softw.*, 25:19–27, 2010.

[5] G. Bigi, A. Frangioni, and Q. Zhang. Outer approximation algorithms for canonical DC problems. *J. Glob. Optim.*, 46:163–189, 2010.

[6] R. Blanquero and E. Carrizosa. On covering methods for D.C. optimization. *J. Glob. Optim.*, 18:265–274, 2000.

[7] R. Cominetti. Metric regularity, tangent sets, and second-order optimality conditions. *Appl. Math. Optim.*, 21:265–287, 1990.

[8] V. F. Demyanov. Continuous generalized gradients for nonsmooth functions. In A. Kurzhanski, K. Neumann, and D. Pallaschke, editors, *Optimization, Parallel Processing and Applications*, pages 24–27. Springer, Berlin, Heidelberg, 1988.

[9] V. F. Demyanov. On codifferentiable functions. *Vestn. Leningr. Univ., Math.*, 2:22–26, 1988.

[10] V. F. Demyanov. Smoothness of nonsmooth functions. In F. Clarke, V. Demyanov, and F. Giannesssi, editors, *Nonsmooth Optimization and Related Topics*, pages 79–88. Springer, Boston, 1989.

[11] V. F. Demyanov. Conditions for an extremum in metric spaces. *J. Glob. Optim.*, 17:55–63, 2000.

[12] V. F. Demyanov, A. M. Bagirov, and A. M. Rubinov. A method of truncated codifferential with applications to some problems of cluster analysis. *J. Glob. Optim.*, 23:63–80, 2002.

[13] V. F. Demyanov and A. M. Rubinov. *Constructive Nonsmooth Analysis.* Peter Lang, Frankfurt am Main, 1995.

[14] N. Dinh, T. T. A. Nghia, and G. Vallet. A closedness condition and its applications to DC programs with convex constraints. *Optim.*, 59:541–560, 2010.

[15] M. V. Dolgopolik. Nonsmooth problems of calculus of variations via codifferentiation. *ESAIM: Control Optim. Calc. Var.*, 20:1153–1180, 2014.

[16] M. V. Dolgopolik. Abstract convex approximations of nonsmooth functions. *Optim.*, 64:1439–1469, 2015.

[17] M. V. Dolgopolik. A unifying theory of exactness of linear penalty functions. *Optim.*, 65:1167–1202, 2016.

[18] M. V. Dolgopolik. A unifying theory of exactness of linear penalty functions II: parametric penalty functions. *Optim.*, 66:1577–1622, 2017.

[19] M. V. Dolgopolik. A convergence analysis of the method of codifferential descent. *Comput. Optim. Appl.*, 71:879–913, 2018.

[20] M. V. Dolgopolik. A unified approach to the global exactness of penalty and augmented lagrangian functions I: parametric exactness. *J. Optim. Theory Appl.*, 176:728–744, 2018.

[21] M. V. Dolgopolik. The method of codifferential descent for convex and global piecewise affine optimization. *Optim. Methods Softw.*, 2019. DOI: 10.1080/10556788.2019.1571590.

[22] M. Dür, R. Horst, and M. Locatelli. Necessary and sufficient global optimality conditions for convex maximization revisited. *J. Math. Analysis Appl.*, 217:637–649, 1998.

[23] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* SIAM, Philadelphia, 1999.

[24] A. Ferrer and J. E. Martínez-Legaz. Improving the efficiency of DC global optimization methods by improving the DC representation of the objective function. *J. Glob. Optim.*, 43:513–531, 2009.

[25] M. Gaudioso, G. Giallombardo, G. Miglionico, and A. M. Bagirov. Minimizing nonsmooth DC functions via successive DC piecewise-affine approximations. *J. Glob. Optim.*, 71:37–55, 2018.

[26] H. Gfrerer. First order and second order characterization of metric subregularity and calmness of constraint set mappings. *SIAM J. Optim.*, 21:1439–1474, 2011.

[27] V. V. Gorokhovik and O. I. Zorko. Piecewise affine functions and polyhedral sets. *Optim.*, 31:209–221, 1994.

[28] J.-B. Hiriart-Urruty. From convex minimization to nonconvex minimization: necessary and sufficient conditions for global optimality. In F. N. Clarke, V. F. Demyanov, and F. Giannessi, editors, *Nonsmooth optimization and related topics*, pages 219–240. Plenum, New York, 1989.

[29] J.-B. Hiriart-Urruty. Conditions for global optimality. In R. Horst and P. M. Pardalos, editors, *Handbook of Global Optimization*, pages 1–26. Kluwer Academic Publishers, Dordrecht, 1995.

[30] J.-B. Hiriart-Urruty. Conditions for global optimality 2. *J. Glob. Optim.*, 13:349–367, 1998.

[31] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I. Fundamentals.* Springer-Verlag, Berlin, Heidelberg, 1993.

[32] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II. Advanced Theory and Bundle Methods.* Springer-Verlag, Berlin, Heidelber, 1993.

[33] R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization.* Kluwer Academic Publishers, Dordrecht, 2000.

[34] R. Horst and N. V. Thoai. DC programming: Overview. *J. Optim. Theory Appl.*, 103:1–43, 1999.

[35] A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 1: A reduction theorem and first order conditions. *SIAM J. Control Optim.*, 17:245–250, 1979.

[36] A. D. Ioffe and R. T. Rockafellar. The Euler and Weierstrass conditions for nonsmooth variational problems. *Calc. Var. Partial Differ. Equ.*, 4:59–87, 1996.

[37] V. Jeyakumar and B. M. Glover. Characterizing global optimality for DC optimization problems under convex inequality constraints. *J. Glob. Optim.*, 8:171–187, 1996.

[38] K. Joki, A. M. Bagirov, N. Karmitsa, and M. Mäkelä. A proximal bundle method for nonsmooth DC optimization utilizing nonconvex cutting planes. *J. Glob. Optim.*, 68:501–535, 2017.

[39] K. Joki, A. M. Bagirov, N. Karmitsa, M. Mäkelä, and S. Taheri. Double bundle method for finding Clarke stationary points in nonsmooth DC programming. *SIAM J. Optim.*, 28:1892–1919, 2018.

[40] A. Y. Kruger. Error bounds and metric subregularity. *Optim.*, 64:49–79, 2015.

[41] D. Kumar and Y. Lucet. Computation of the epsilon-subdifferential of convex piecewise linear-quadratic functions in optimal worst-case time. *Set-Valued Var. Anal.*, 27:623–641, 2019.

[42] H. A. Le Thi and P. Dinh. DC programming and DCA: thirty years of development. *Math. Program.*, 169:5–68, 2018.

[43] H. A. Le Thi, T. P. Dinh, and H. Van Ngai. Exact penalty and error bounds in DC programming. *J. Glob. Optim.*, 52:509–535, 2012.

[44] G. Leoni. *A First Course in Sobolev spaces.* American Mathematical Society, Providence, RI, 2009.

[45] L. Polyakova, V. Karelin, S. Myshkov, and E. Stankova. Some methods for minimizing of d.c. functions. In S. Misra et al., editor, *Computational Science and Its Applications — ICCSA 2019. Lecture Notes in Computer Science, vol. 11622*, pages 667–677. Springer, Cham, 2019.

[46] L. N. Polyakova. On global unconstrained minimization of the difference of polyhedral functions. *J. Glob. Optim.*, 50:179–195, 2011.

[47] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

[48] A. M. Rubinov. *Abstract Convexity and Global Optimization.* Kluwer Academic Publishers, Boston–Dordrecht–London, 2000.

[49] I. Singer. *Duality for Nonconvex Approximation and Optimization.* Springer, New York, 2006.

[50] A. Strekalovsky. On a local search for reverse convex problems. In L. Liberti and N. Maculan, editors, *Global Optimization. Nonconvex Optimization and Its Applications, vol. 84*, pages 33–43. Springer, Boston, MA, 2006.

[51] A. S. Strekalovsky. Global optimality conditions for nonconvex optimization. *J. Glob. Optim.*, 12:415–434, 1998.

[52] A. S. Strekalovsky. On the minimization of the difference of convex functions on a feasible set. *Comput. Math. Math. Phys.*, 43:380–390, 2003.

[53] A. S. Strekalovsky. On local search in d.c. optimization problems. *Appl. Math. Comput.*, 255:73–83, 2015.

[54] A. S. Strekalovsky. Global optimality conditions in nonconvex optimization. *J. Optim. Theory Appl.*, 173:770–792, 2017.

[55] A. S. Strekalovsky. Global optimality conditions and exact penalization. *Optim. Lett.*, 13:597–615, 2019.

[56] A. S. Strekalovsky and M. V. Yanulevich. On global search in nonconvex optimal control problems. *J. Glob. Optim.*, 65:119–135, 2016.

[57] A. H. Tor, A. Bagirov, and B. Karasözen. Aggregate codifferential method for nonsmooth DC optimization. *J. Comput. Appl. Math.*, 259:851–867, 2014.

[58] H. Tuy. Convex programs with an additional reverse convex constraint. *J. Optim. Theory Appl.*, 52:463–486, 1987.

[59] H. Tuy. Global minimization of a difference of two convex functions. In B. Cornet, V. H. Nguyen, and J. P. Vial, editors, *Nonlinear Analysis and Optimization. Mathematical Programming Studies, vol. 30*, pages 150–182. Springer, Berling, Heidelberg, 1987.

[60] H. Tuy. Canonical DC programming problem: outer approximation methods revisited. *Oper. Res. Lett.*, 18:99–106, 1995.

[61] H. Tuy. D.C. optimization: Theory, methods and algorithms. In R. Horst and P. Pardalos, editors, *Handbook of Global Optimization*, pages 149–216. Kluwer Academic Publishers, Dordrecht, 1995.

[62] H. Tuy. *Convex Analysis and Global Optimization.* Kluwer Academic Publishers, Dordrecht, 1998.

[63] H. Tuy. On some recent advances and applications of D.C. optimization. In V. H. Nguyen, J. J. Strodiot, and P. Tossings, editors, *Optimization. Lecture Notes in Economics and Mathematical Systems, vol. 481*, pages 473–497. Springer, Berling, Heidelberg, 2000.

[64] H. Tuy. On global optimality conditions and cutting plane algorithms. *J. Optim. Theory Appl.*, 118:201–216, 2003.

[65] Q. Zhang. A new necessary and sufficient global optimality condition for canonical DC problems. *J. Glob. Optim.*, 55:559–577, 2013.