

A convergence analysis of the method of codifferential descent

Dolgopolik M.V.*‡

March 4, 2019

Abstract

This paper is devoted to a detailed convergence analysis of the method of codifferential descent (MCD) developed by professor V.F. Demyanov for solving a large class of nonsmooth nonconvex optimization problems. We propose a generalization of the MCD that is more suitable for applications than the original method, and that utilizes only a part of a codifferential on every iteration, which allows one to reduce the overall complexity of the method. With the use of some general results on uniformly codifferentiable functions obtained in this paper, we prove the global convergence of the generalized MCD in the infinite dimensional case. Also, we propose and analyse a quadratic regularization of the MCD, which is the first general method for minimizing a codifferentiable function over a convex set. Apart from convergence analysis, we also discuss the robustness of the MCD with respect to computational errors, possible step size rules, and a choice of parameters of the algorithm. In the end of the paper we estimate the rate of convergence of the MCD for a class of nonsmooth nonconvex functions that arise, in particular, in cluster analysis. We prove that under some general assumptions the method converges with linear rate, and it convergence quadratically, provided a certain first order sufficient optimality condition holds true.

1 Introduction

The class of codifferentiable functions was introduced by professor V.F. Demyanov [14, 15, 16] in the late 1980s. The main feature of this class of nonsmooth functions is the fact that the approximation provided by the codifferential is *continuous* with respect to the reference point and nonhomogeneous w.r.t. the direction, while other standard tools of nonsmooth analysis, such as various generalized derivatives and subdifferentials, provide approximations that are homogeneous w.r.t. the direction, but are continuous w.r.t. the reference point only in the smooth case. Furthermore, the class of codifferentiable functions includes all d.c. functions, forms a vector lattice closed under all standard algebraic operations and composition, and admits a simple and *exact* calculus [19] that can

*Saint Petersburg State University, Saint Petersburg, Russia

†Institute of Problems of Mechanical Engineering, Saint Petersburg, Russia

‡The author was supported by the Russian Foundation for Basic Research under grant no. 16-31-00056.

be easily implemented on a computer (cf. *fuzzy* or *approximate* calculus rules in [46, 50]). The codifferential calculus was extended to the infinite dimensional case in [57, 25, 28].

Codifferentiable functions were introduced as a natural modification of the class of quasidifferentiable functions. Recall that a function f is called quasidifferentiable at a point x , if f is directionally differentiable at x , and there exists a pair of convex compact sets $\mathcal{D}f(x) = [\underline{\partial}f(x), \overline{\partial}f(x)]$, called a *quasidifferential* of f at x , such that $f'(x, h) = \max_{v \in \underline{\partial}f(x)} \langle v, h \rangle + \min_{w \in \overline{\partial}f(x)} \langle w, h \rangle$. Despite their many useful properties and various applications [19, 21, 20, 47, 56], it appeared to be very difficult to design effective numerical methods for minimizing general quasidifferentiable functions, and only methods for minimizing some specific classes of these functions were proposed (see, e.g., [8, 2, 44]). Since quasidifferentiable functions are locally Lipschitz continuous, one can use general methods of nonsmooth optimization, such as bundle methods [45, 32, 34, 31], gradient sampling methods [9, 39, 33, 12], quasi-Newton methods [41, 38], discrete gradient methods [5, 6] and many others [11, 37, 3], in order to minimize a quasidifferentiable function. However, these methods do not utilize any additional information about function's behaviour that a quasidifferential can provide.

Unlike the case of quasidifferentiable functions, it turned out to be possible to design an effective numerical method for minimizing codifferentiable functions called *the method of codifferential descent* (MCD) [19]. A modification of the MCD aiming at reducing the complexity of the original method was proposed in [17]. Methods for minimizing nonsmooth convex and d.c. functions combining the ideas of the MCD and bundle methods were developed in [4, 55, 7, 54]. A trust region method for minimizing codifferentiable functions was studied in [1]. Finally, an infinite-dimensional version of the MCD was applied to some problems of the calculus of variations [22, 23] and optimal control problems [30]. However, theoretical results on a convergence of the MCD are very scarce, and only some results on the global convergence of this method and its modifications are known in the finite dimensional case. Furthermore, in [19, 17] the global convergence of the MCD was proved under the assumption that the objective function is codifferentiable uniformly in directions in a neighbourhood of a limit point, and it is unclear how to verify this assumption in any particular case, while the convergence analysis in [4, 55, 7, 54] heavily relies on the convexity or the d.c. structure of the objective function. Let us finally note that the MCD, as well as the method of truncated codifferential [17], are unrealisable in the general case, since they cannot be directly implemented in practice (see Remark 10 below).

The main goal of this paper is to provide a comprehensive convergence analysis of the MCD in the general case. To this end, we introduce and study a new class of nonsmooth functions called *uniformly codifferentiable* functions. We prove that this class of functions contains almost all codifferentiable functions appearing in applications. We also introduce a new generalized version of the MCD that, in essence, coincides with the practical implementation of the original MCD, and is reduced to the MCD or the method of truncated codifferential [17] under a certain choice of parameters. Furthermore, our version of the MCD allows one to use only a part of a codifferential on every iteration, which might significantly reduce the overall complexity of the method for some specific problems. Under some natural and easily verifiable assumptions we prove the global convergence of the generalized MCD in the infinite dimensional case,

and for the first time prove the convergence of a non-stationarity measure to zero for a sequence generated by the MCD. Finally, we introduce and study a *quadratic regularization of the MCD*, which is the first general method for minimizing codifferentiable functions over convex sets. This method can be viewed as a generalization of the Pschenichnyi-Pironneau-Polak (PPP) algorithm for min-max problems [53, 51, 13, 52] to the case of codifferentiable functions.

In the end of the paper we for the first time study the rate of convergence of the MCD. Since an analysis of the rate of convergence of the MCD in the general case requires the use of very complicated and cumbersome second order approximations of codifferentiable functions such as the so-called second order codifferentials [19], for the sake of simplicity we study only the rate of convergence of the quadratic regularization of the MCD for a particular class of nonsmooth nonconvex functions that arise, in particular, in cluster analysis [17]. Under some general assumptions we prove that the quadratic regularization of the MCD converges with linear rate, and it converges quadratically, if the limit point of a sequence generated by the method satisfies certain first order sufficient optimality conditions. These results can be viewed as an interesting example that sheds some light on the way the MCD performs for various problems, and how the rate of convergence of this method can be estimated in the general case.

The paper is organized as follows. In Section 3 we present a general scheme of the MCD and prove the global convergence of the method. In this section we also analyse the robustness of the MCD with respect to computational errors, and discuss possible step size rules and a choice of parameters of the method. Section 4 is devoted to the quadratic regularization of the MCD, while the rate of convergence of this method for a class of nonsmooth nonconvex functions is studied in Section 5. For the reader's convenience, some general results on codifferentiable function are given in Section 2. This section can be viewed as a brief, but thorough and almost self-contained introduction into the codifferential calculus. However, it should be noted that many results presented in Section 2, such as the Lipschitz continuity of codifferentiable functions and general results on uniformly codifferentiable functions, are completely new.

2 Codifferentiable functions

Let \mathcal{H} be a real Hilbert space, and let a function $f: U \rightarrow \mathbb{R}$ be defined in a neighbourhood U of a point $x \in \mathcal{H}$. Recall that f is said to be *codifferentiable* at x , if there exist weakly compact convex sets $\underline{d}f(x), \bar{d}f(x) \subset \mathbb{R} \times \mathcal{H}$ such that for any $\Delta x \in \mathcal{H}$ one has

$$\lim_{\alpha \rightarrow +0} \frac{1}{\alpha} \left| f(x + \alpha \Delta x) - f(x) - \max_{(a,v) \in \underline{d}f(x)} (a + \alpha \langle v, \Delta x \rangle) - \min_{(b,w) \in \bar{d}f(x)} (b + \alpha \langle w, \Delta x \rangle) \right| = 0,$$

and

$$\max_{(a,v) \in \underline{d}f(x)} a + \min_{(b,w) \in \bar{d}f(x)} b = 0. \quad (1)$$

Here $\langle \cdot, \cdot \rangle$ is the inner product in \mathcal{H} . The pair $Df(x) = [\underline{d}f(x), \bar{d}f(x)]$ is called a *codifferential* of f at x , the set $\underline{d}f(x)$ is called a *hypodifferential* of f at x , while

the set $\bar{d}f(x)$ is referred to as a *hyperdifferential* of f at x .

Remark 1. One can verify that the function f is codifferentiable at x iff there exist continuous convex functions $\Phi, \Psi: \mathcal{H} \rightarrow \mathbb{R}$ such that $\Phi(0) - \Psi(0) = 0$, and for any $\Delta x \in \mathcal{H}$ one has

$$\lim_{\alpha \rightarrow +0} \frac{1}{\alpha} \left| f(x + \alpha \Delta x) - f(x) - (\Phi(\alpha \Delta x) - \Psi(\alpha \Delta x)) \right| = 0$$

(see [28], Example 3.10). Thus, the codifferentiable functions are exactly those nonsmooth functions that can be locally approximated by a d.c. function, which, in particular, implies that any d.c. function is codifferentiable.

Observe that a codifferential of f at x is not unique. Indeed, if $Df(x)$ is a codifferential of f at x , and $C \subset \mathbb{R} \times \mathcal{H}$ is a weakly compact convex set, then the pair $[\underline{d}f(x) + C, \bar{d}f(x) - C]$ is a codifferential of f at x , as well. Note also that not all codifferentials of f at x have the form $[\underline{d}f(x) + C, \bar{d}f(x) + C]$ for some weakly compact convex set C . For instance, applying the inequality $f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle$ one can check that for any $c \geq 0$ the pair $[\text{co}\{(-y^2, 2y) : y \in [-c, c]\}, \{0\}]$ is a codifferential of the function $f(x) = x^2$ at the point $x = 0$.

If there exists a codifferential of f at x of the form $[\underline{d}f(x), \{0\}]$, then the function f is called *hypodifferentiable* at x . Similarly, if there exists a codifferential of f at x of the form $[\{0\}, \bar{d}f(x)]$, then f is said to be *hyperdifferentiable* at x . The function f is called *continuously codifferentiable* at the point x , if f is codifferentiable in a neighbourhood of x (i.e. at every point of this neighbourhood), and there exists a *codifferential mapping* $Df(\cdot)$ defined in neighbourhood of x and such that the multifunctions $\underline{d}f(\cdot)$ and $\bar{d}f(\cdot)$ are Hausdorff continuous at x . In this case the codifferential mapping $Df(\cdot)$ is called *continuous* at x . Finally, a function f defined in a neighbourhood U of a set $A \subset \mathcal{H}$ is called *continuously codifferentiable on the set A* , if f is codifferentiable at every point of the set A , and there exists a codifferential mapping $Df(\cdot)$ such that the corresponding multifunctions $\underline{d}f(\cdot)$ and $\bar{d}f(\cdot)$ are Hausdorff continuous on the set A . Continuously hypodifferentiable and continuously hyperdifferentiable functions are defined in the same way.

Remark 2. From this point onwards, when we consider a continuously codifferentiable function, we suppose that the corresponding codifferential mapping $Df(\cdot)$ is continuous (i.e. we do not consider discontinuous codifferential mappings for continuously codifferentiable functions).

Let us give some simple examples of continuously codifferentiable functions.

Example 1. Let f be Gâteaux differentiable at a point $x \in \mathcal{H}$. Then f is codifferentiable at x , and both pairs $[\{(0, \nabla f(x))\}, \{0\}]$ and $[\{0\}, \{(0, \nabla f(x))\}]$ are codifferentials of f at x , i.e. f is both hypodifferentiable and hyperdifferentiable at x . If, in addition, f is continuously differentiable at x , then f is continuously codifferentiable at this point.

Example 2. Let $f(x) = \|Ax + b\|$, where $A: \mathcal{H} \rightarrow \mathcal{H}$ is a bounded linear operator, and $b \in \mathcal{H}$ is fixed. Then

$$\begin{aligned} f(x + \Delta x) - f(x) &= \max_{v \in B(0,1)} \langle v, A(x + \Delta x) + b \rangle - \|Ax + b\| \\ &= \max_{(a,v) \in \underline{d}f(x)} (a + \langle v, \Delta x \rangle), \end{aligned}$$

where

$$\underline{d}f(x) = \left\{ \left(\langle v, Ax + b \rangle - \|Ax + b\|, A^*v \right) \mid v \in B(0, 1) \right\},$$

the operator A^* is the adjoint of A , and $B(y, r) = \{z \in \mathcal{H} \mid \|z - y\| \leq r\}$. The set $\underline{d}f(x)$ is convex and weakly compact (the weak compactness follows from the fact that the operator A^* is an endomorphism of the space \mathcal{H} endowed with the weak topology). Furthermore, it is easy to see that the mapping $\underline{d}f(\cdot)$ is Hausdorff continuous. Thus, the function $f(x) = \|Ax + b\|$ is continuously hypodifferentiable on the entire space \mathcal{H} .

Example 3. Let f be a piecewise affine function of the form $f(x) = f_1(x) + f_2(x)$ with

$$f_1(x) = \max_{i \in I} (a_i + \langle v_i, x \rangle), \quad f_2(x) = \min_{j \in J} (b_j + \langle w_j, x \rangle),$$

where I and J are some finite index sets, $a_i, b_j \in \mathbb{R}$ and $v_i, w_j \in \mathcal{H}$. Then the function f is continuously codifferentiable on \mathcal{H} , and one can define $\underline{d}f(x) = \{(a_i + \langle v_i, x \rangle - f_1(x), v_i) \mid i \in I\}$ and $\bar{d}f(x) = \{(b_j + \langle w_j, x \rangle - f_2(x), w_j) \mid j \in J\}$.

Example 4. Let f be a convex function, and $U \subset \mathcal{H}$ be a bounded open set such that f is Lipschitz continuous on U . By the definition of subgradient for any $y \in U$, $v \in \partial f(y)$ and $x \in \mathcal{H}$ one has $f(x) \geq f(y) + \langle v, x - y \rangle$, and this inequality turns into an equality when $x = y$. Therefore, $f(x) = \max_{(a,v) \in C} (a + \langle v, x \rangle)$ for any $x \in U$, where

$$C = \{(f(y) - \langle v, y \rangle, v) \in \mathbb{R} \times \mathcal{H} \mid v \in \partial f(y), y \in U\}.$$

Hence taking into account the fact that U is an open set one obtains that for any $x \in U$ there exists $r > 0$ such that

$$f(x + \Delta x) - f(x) = \max_{(a,v) \in \underline{d}f(x)} (a + \langle v, \Delta x \rangle) \quad \forall \Delta x \in B(x, r),$$

where

$$\underline{d}f(x) = \text{clco} \left\{ (f(y) - f(x) + \langle v, x - y \rangle, v) \in \mathbb{R} \times \mathcal{H} \mid v \in \partial f(y), y \in U \right\}.$$

Note that the set $\underline{d}f(x)$ is bounded (and thus weakly compact) due to the fact that f is Lipschitz continuous on U . Furthermore, one can verify that the mapping $\underline{d}f(\cdot)$ is Hausdorff continuous. Hence, in particular, one gets that a continuous convex function $f: \mathcal{H} \rightarrow \mathbb{R}$ is continuously hypodifferentiable at every point $x \in \mathcal{H}$. On the other hand, if f is continuously codifferentiable at a given point, then it is Lipschitz continuous in a neighbourhood of this point (see Corollary 2 below). Thus, a proper convex function $f: \mathcal{H} \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is continuously hypodifferentiable at a point x iff $x \in \text{int dom } f$, and f is continuous at this point. If \mathcal{H} is finite dimensional, then f is continuously hypodifferentiable at every interior point of its effective domain.

Let a function f be defined and codifferentiable on an open set $U \subset \mathcal{H}$. Observe that without loss of generality one can suppose that

$$\max_{(a,v) \in \underline{d}f(x)} a = \min_{(b,w) \in \bar{d}f(x)} b = 0 \quad \forall x \in U, \quad (2)$$

since otherwise one can use the codifferential mapping $\widehat{Df}(\cdot) = [\widehat{\underline{d}f}(\cdot), \widehat{\overline{d}f}(\cdot)]$ of the function f of the form

$$\begin{aligned}\widehat{\underline{d}f}(x) &= \{(a - a(x), v) \mid (a, v) \in \underline{d}f(x)\}, \\ \widehat{\overline{d}f}(x) &= \{(b - b(x), w) \mid (b, w) \in \overline{d}f(x)\}\end{aligned}$$

where $a(x) = \max_{(a,v) \in \underline{d}f(x)} a$, and $b(x) = \min_{(b,w) \in \overline{d}f(x)} b$ (recall that $a(x) = -b(x)$ by (1)). Observe also that if the codifferential mapping $Df(\cdot)$ is continuous, then the codifferential mapping $\widehat{Df}(\cdot)$ is continuous as well. Note also that the codifferential mappings from the examples above satisfy equalities (2). Therefore, hereinafter we suppose that (2) always holds true.

Let a function f be defined in a neighbourhood of a point $x \in \mathcal{H}$ and codifferentiable at this point. Then, as it is easy to see, the function f is directionally differentiable at x , and $f'(x, h) = \Phi'(0, h) + \Psi'(0, h)$ for all $h \in \mathcal{H}$, where

$$\Phi(y) = \max_{(a,v) \in \underline{d}f(x)} (a + \langle v, y \rangle), \quad \Psi(y) = \min_{(b,w) \in \overline{d}f(x)} (b + \langle w, y \rangle).$$

Applying the theorem about the subdifferential of the supremum of a family of convex functions (see, e.g., [36, Thm. 4.2.3]) and equalities (2) one obtains that f is quasidifferentiable at x , and the pair $\mathcal{D}f(x) = [\underline{\partial}f(x), \overline{\partial}f(x)]$ with

$$\underline{\partial}f(x) = \{v \in \mathcal{H} \mid (0, v) \in \underline{d}f(x)\}, \quad \overline{\partial}f(x) = \{w \in \mathcal{H} \mid (0, w) \in \overline{d}f(x)\} \quad (3)$$

is a quasidifferential of f at x . Note that both sets $\underline{\partial}f(x)$ and $\overline{\partial}f(x)$ are nonempty, weakly compact and convex. Hereinafter, we consider only the quasidifferential mapping $\mathcal{D}f(\cdot)$ of the form (3) corresponding to a chosen codifferential mapping of the function f .

With the use of the standard first order optimality condition in terms of directional derivative we arrive at the following result, which is well-known in the finite dimensional case.

Proposition 1. *Let a function f be defined in a neighbourhood of a point $x^* \in \mathcal{H}$ and codifferentiable at this point. Suppose also that x^* is a point of local minimum of the function f . Then*

$$0 \in \underline{d}f(x^*) + (0, w) \quad \forall (0, w) \in \overline{d}f(x^*). \quad (4)$$

Moreover, optimality condition (4) holds true iff $f'(x^*, \cdot) \geq 0$; thus, it is independent of the choice of a codifferential, i.e. if (4) holds true for one codifferential of f at x^* , then it holds true for all codifferentials of f at x^* .

Proof. From the fact that x^* is a point of local minimum of f it follows that $f'(x^*, \cdot) \geq 0$. Hence by the definition of quasidifferentiable function one has

$$f'(x^*, h) = \max_{v \in \underline{\partial}f(x^*)} \langle v, h \rangle + \min_{w \in \overline{\partial}f(x^*)} \langle w, h \rangle \geq 0 \quad \forall h \in \mathcal{H}.$$

Clearly, this inequality holds true iff

$$\max_{v \in \underline{\partial}f(x^*)} \langle v + w, h \rangle \geq 0 \quad \forall h \in \mathcal{H} \quad \forall w \in \overline{\partial}f(x^*).$$

In turn, this condition is valid iff $0 \in \underline{\partial}f(x^*) + w$ for all $w \in \overline{\partial}f(x^*)$ or, equivalently, iff (4) holds true (see (3)). Finally, note that since the condition $f'(x^*, \cdot) \geq 0$ is independent of the choice of a codifferential, the optimality condition (4) is independent of the choice of a codifferential as well. \square

Remark 3. (i) The independence of optimality conditions in quasidifferential programming of the choice of quasidifferentials was studied in [42, 43].

(ii) Note that from equalities (2) it follows that if f is codifferentiable at a point x , then there exists at least one pair $(0, w) \in \overline{df}(x)$, i.e. the optimality condition (4) cannot be satisfied vacuously.

Corollary 1. *Let f and x^* be as in the proposition above. Then the optimality condition (4) holds true iff $0 \in \underline{df}(x^*) + (0, w)$ for all $(0, w) \in \text{ext } \overline{df}(x^*)$, where “ext” stands for the set of extreme points of a convex set.*

Proof. The validity of “only if” part of the statement is obvious; therefore, let us prove the “if” part. It is easy to verify that $(0, w) \in \text{ext } \overline{df}(x^*)$ iff $w \in \text{ext } \overline{\partial}f(x^*)$. Hence and from the Krein-Milman theorem (recall that the set $\overline{\partial}f(x^*)$ is nonempty, weakly compact and convex) it follows that there exists at least one point $(0, w) \in \text{ext } \overline{df}(x^*)$.

Arguing by reductio ad absurdum, suppose that (4) does not hold true. Then there exists $w \in \overline{\partial}f(x^*)$ such that $0 \notin \underline{\partial}f(x^*) + w$. By the separation theorem there exist $h \in \mathcal{H}$ and $c > 0$ such that $\langle v + w, h \rangle \geq c$ for all $v \in \underline{\partial}f(x^*)$. Applying the Krein-Milman theorem one obtains that there exists $w_0 \in \text{co ext } \overline{\partial}f(x^*)$ such that $|\langle w - w_0, h \rangle| < c/2$ (recall that $\overline{\partial}f(x^*) = \text{cl co ext } \overline{\partial}f(x^*)$, where the closure is taken in the weak topology due to the fact that the space \mathcal{H} is infinite dimensional in the general case). Therefore $\langle v + w_0, h \rangle \geq c/2$ for all $v \in \underline{\partial}f(x^*)$, which implies that $0 \notin \underline{\partial}f(x^*) + w_0$. On the other hand, from the fact that $w_0 \in \text{co ext } \overline{\partial}f(x^*)$ it obviously follows that $0 \in \underline{\partial}f(x^*) + w_0$, which is impossible. Thus, the proof is complete. \square

With the use of the optimality condition (4) we can prove the mean value theorem for codifferentiable functions. The proof of this result almost literally repeats the proof of the classical mean value theorem for differentiable functions. To the best of author’s knowledge, this proof has never been published before.

Proposition 2. *Let $U \subset \mathcal{H}$ be an open set, and let $f: U \rightarrow \mathbb{R}$ be codifferentiable on a set $C \subseteq U$. Then for any points $x_1, x_2 \in C$ with $\text{co}\{x_1, x_2\} \subseteq C$ there exist $y \in \text{co}\{x_1, x_2\}$, $(0, v) \in \underline{df}(y)$, and $(0, w) \in \overline{df}(y)$ such that $f(x_2) - f(x_1) = \langle v + w, x_2 - x_1 \rangle$.*

Proof. Let a function $g: (-\varepsilon, 1 + \varepsilon) \rightarrow \mathbb{R}$, where $\varepsilon > 0$, be codifferentiable on $[0, 1]$. Applying the definition of codifferentiable function one can easily verify that g is continuous on $[0, 1]$.

Suppose, at first, that $g(0) = g(1) = 0$. Then g attains either a local minimum or a local maximum at a point $\theta \in (0, 1)$. If g attains a local minimum at θ , then applying Proposition 1 one obtains that for all $(0, w) \in \overline{dg}(\theta)$ there exists $(0, v) \in \underline{dg}(\theta)$ such that $v + w = 0$. If g attains a local maximum at θ , then applying Proposition 1 to the function $-g$ one gets that for all $(0, v) \in \underline{dg}(\theta)$ there exists $(0, w) \in \overline{dg}(\theta)$ such that $v + w = 0$ (note that the function $\eta(t) = -g(t)$ is obviously codifferentiable, and $D\eta(\theta) = [-\overline{dg}(\theta), -\underline{dg}(\theta)]$). In either case there exist $(0, v) \in \underline{dg}(\theta)$ and $(0, w) \in \overline{dg}(\theta)$ such that $v + w = 0$.

Suppose, now, that $g(0)$ and $g(1)$ are arbitrary. For any $t \in (-\varepsilon, 1 + \varepsilon)$ define $r(t) = g(t) - g(0) - t(g(1) - g(0))$. Then $r(0) = r(1) = 0$ and, as it is easily seen, the function r is codifferentiable on $[0, 1]$, and one can define $Dr(t) = [\underline{dg}(t) + (0, g(0) - g(1)), \overline{dg}(t)]$. Hence by the first part of the proof there exist $\theta \in (0, 1)$, $(0, v) \in \underline{dg}(\theta)$, and $(0, w) \in \overline{dg}(\theta)$ such that $g(1) - g(0) = v + w$.

Let us, now, return to the function f . Let points $x_1, x_2 \in C$ be such that $\text{co}\{x_1, x_2\} \subset C$. Since U is open and $C \subseteq U$, there exists $\varepsilon > 0$ such that for any $t \in (-\varepsilon, 1 + \varepsilon)$ one has $x(t) = x_1 + t(x_2 - x_1) \in U$. Define $g(t) = f(x(t))$ for all $t \in (-\varepsilon, 1 + \varepsilon)$. One can easily verify that the function g is codifferentiable on $[0, 1]$ (see, e.g., [28, Thm. 4.5]), and one can define

$$\begin{aligned} \underline{d}g(t) &= \{(a, \langle v, x_2 - x_1 \rangle) \mid (a, v) \in \underline{d}f(x(t))\}, \\ \bar{d}g(t) &= \{(b, \langle w, x_2 - x_1 \rangle) \mid (b, w) \in \bar{d}f(x(t))\}. \end{aligned}$$

Hence by the second part of the proof there exist $\theta \in (0, 1)$, $(0, v) \in \underline{d}f(x(\theta))$, and $(0, w) \in \bar{d}f(x(\theta))$ such that $\langle v + w, x_2 - x_1 \rangle = g(1) - g(0) = f(x_2) - f(x_1)$, which completes the proofs. \square

Remark 4. Strictly speaking, in the statement of the proposition above one must indicate that this proposition holds true for any codifferential mapping Df on the set C . However, in order not to overcomplicate the statement of the proposition, as well as the statements of all results below, we implicitly mean that each result below holds true for all codifferential mappings satisfying the assumptions of this result.

Corollary 2. *Let $U \subset \mathcal{H}$ be an open set, and let a function $f: U \rightarrow \mathbb{R}$ be codifferentiable on a convex set $C \subseteq U$. Suppose also that*

$$R = \sup_{x \in C} \sup \{\|v\| \mid (0, v) \in \underline{d}f(x) + \bar{d}f(x)\} < +\infty.$$

Then f is Lipschitz continuous on C with a Lipschitz constant $L = R$. In particular, if f is continuously codifferentiable on U , then f is locally Lipschitz continuous on this set.

Remark 5. (i) Let us note that a different proof of the fact that any continuously codifferentiable function is locally Lipschitz continuous was given in [40] in the finite dimensional case.

(ii) Since a continuously codifferentiable function f is locally Lipschitz continuous, one can consider its Clarke subdifferential. Let us note that the Clarke subdifferential of the function f is connected with a quasidifferential of this function via the Demyanov difference (see [19, Chapter III.4]). Furthermore, it is easy to see that if \mathcal{H} is finite dimensional, and f is hypodifferentiable and regular (in the sense of Clarke; see [10]) at x , then its Clarke subdifferential at x has the form $\partial_{Cl}f(x) = \{v \in \mathcal{H} \mid (0, v) \in \underline{d}f(x)\}$.

Let a function f be defined in a neighbourhood U of a point $x \in \mathcal{H}$ and codifferentiable at this point. For any $r > 0$ such that $B(x, r) \subset U$, and for all $\Delta x \in B(0, r)$ and $\alpha \in [0, 1]$ define

$$\begin{aligned} \varepsilon_f(\alpha, \Delta x, x, r) &= \frac{1}{\alpha} \left(f(x + \alpha \Delta x) - f(x) - \max_{(a, v) \in \underline{d}f(x)} (a + \alpha \langle v, \Delta x \rangle) \right. \\ &\quad \left. - \min_{(b, w) \in \bar{d}f(x)} (b + \alpha \langle w, \Delta x \rangle) \right) \end{aligned} \quad (5)$$

Obviously, the function ε_f depends on the choice of a codifferential $Df(x)$. However, in order not to overcomplicate the notation, we do not indicate this dependence explicitly. Note that from the definition of codifferentiable function

it follows that $\varepsilon_f(\alpha, \Delta x, x, r) \rightarrow 0$ as $\alpha \rightarrow +0$. In essence, the function ε_f measures how well the codifferential $Df(x)$ approximates the function f at the point x in a direction $\Delta x \in B(0, r)$.

Let us derive main calculus rules for continuously codifferentiable functions along with some simple estimates of the function ε_f . These results will be important for understanding the theorems on convergence of the method of codifferential descent presented in the following sections. Let us note that although the codifferential calculus is well-developed (see, e.g., [19, 25, 28]), the estimates of the function ε_f obtained below have never been published before.

For any pairs $[A, B]$ and $[C, D]$ of convex subsets A, B, C and D of a linear space X define $[A, B] + [C, D] = [A + C, B + D]$, and for all $\lambda \in \mathbb{R}$ define

$$\lambda[A, B] = \begin{cases} [\lambda A, \lambda B], & \text{if } \lambda \geq 0, \\ [\lambda B, \lambda A], & \text{if } \lambda < 0. \end{cases}$$

Note that these rules of addition and multiplication by scalar are the same as in the Minkowski-Rådström-Hörmander space (see, e.g., [48]). Let $U \subset \mathcal{H}$ be an open set.

Theorem 1. *Let functions $f_i: U \rightarrow \mathbb{R}$, $i \in I = \{1, \dots, m\}$ be continuously codifferentiable on U . Then the functions $f = \max_{i \in I} f_i$ and $g = \min_{i \in I} f_i$ are continuously codifferentiable on U as well, and for any $x \in U$ one can define*

$$Df(x) = \left[\text{co} \left\{ \{(f_i(x) - f(x), 0)\} + \underline{d}f_i(x) - \sum_{j \neq i} \bar{d}f_j(x) \mid i \in I\}, \sum_{i \in I} \bar{d}f_i(x) \right\}, \sum_{i \in I} \bar{d}f_i(x) \right], \quad (6)$$

$$Dg(x) = \left[\sum_{i \in I} \underline{d}f_i(x), \text{co} \left\{ \{(f_i(x) - g(x), 0)\} + \bar{d}f_i(x) - \sum_{j \neq i} \underline{d}f_j(x) \mid i \in I\} \right\} \right].$$

Furthermore, one has

$$\max \left\{ |\varepsilon_f(\alpha, \Delta x, x, r)|, |\varepsilon_g(\alpha, \Delta x, x, r)| \right\} \leq \max_{i \in I} |\varepsilon_{f_i}(\alpha, \Delta x, x, r)|. \quad (7)$$

Proof. We only consider the function f , since the proof for the function g is almost the same. Fix an arbitrary $x \in U$, and let $r > 0$ be such that $B(x, r) \subset U$. Denote

$$\Phi_i(y) = \max_{(a, v) \in \underline{d}f_i(x)} (a + \langle v, y \rangle), \quad \Psi_i(y) = \min_{(b, w) \in \bar{d}f_i(x)} (b + \langle w, y \rangle),$$

and define

$$\Phi(y) = \max_{(a, v) \in \underline{d}f(x)} (a + \langle v, y \rangle), \quad \Psi(y) = \min_{(b, w) \in \bar{d}f(x)} (b + \langle w, y \rangle),$$

where the pair $Df(x) = [\underline{d}f(x), \bar{d}f(x)]$ is defined in (6). Taking into account the fact that the functions f_i are locally Lipschitz continuous on U by Corollary 2 one can easily verify that the mappings $\underline{d}f(\cdot)$ and $\bar{d}f(\cdot)$ are Hausdorff continuous on U . Moreover, the sets $\underline{d}f(x)$ and $\bar{d}f(x)$ are obviously weakly compact and convex.

Fix arbitrary $\Delta x \in B(0, r)$ and $\alpha \in [0, 1]$. By definition one has

$$\begin{aligned} f(x + \alpha\Delta x) - f(x) &= \max_{i \in I} (f_i(x + \alpha\Delta x) - f(x)) = \max_{i \in I} (f_i(x) - f(x)) \\ &+ \Phi_i(\alpha\Delta x) + \Psi_i(\alpha\Delta x) + \alpha\varepsilon_{f_i}(\alpha, \Delta x, x, r) = \max_{i \in I} (f_i(x) - f(x)) \\ &+ \Phi_i(\alpha\Delta x) - \sum_{j \neq i} \Psi_j(\alpha\Delta x) + \alpha\varepsilon_{f_i}(\alpha, \Delta x, x, r) + \sum_{i=1}^m \Psi_i(\alpha\Delta x). \end{aligned} \quad (8)$$

From the obvious equality

$$\Phi(y) = \max_{i \in I} (f_i(x) - f(x) + \Phi_i(y) - \sum_{j \neq i} \Psi_j(y))$$

it follows that

$$\begin{aligned} \left| \max_{i \in I} (f_i(x) - f(x) + \Phi_i(\alpha\Delta x) - \sum_{j \neq i} \Psi_j(\alpha\Delta x) + \alpha\varepsilon_{f_i}(\alpha, \Delta x, x, r)) \right. \\ \left. - \Phi(\alpha\Delta x) \right| \leq \alpha \max_{i \in I} |\varepsilon_{f_i}(\alpha, \Delta x, x, r)|. \end{aligned}$$

Hence applying (8) and the fact that $\Psi(y) = \sum_{i \in I} \Psi_i(y)$ one obtains that $|\varepsilon_f(\alpha, \Delta x, x, r)| \leq \max_{i \in I} |\varepsilon_{f_i}(\alpha, \Delta x, x, r)|$, where $\varepsilon_f(\alpha, \Delta x, x, r)$ is defined as in (5). Therefore the function f is codifferentiable at the point x , $Df(x)$ is a codifferential of f at this point, and inequality (7) holds true. \square

Theorem 2. *Let functions $f_i: U \rightarrow \mathbb{R}$, $i \in I = \{1, \dots, m\}$ be continuously codifferentiable on U , and a real-valued function G be defined and continuously differentiable on an open set V containing the set $\{f(x) \in \mathbb{R}^m \mid x \in U\}$, where $f(x) = (f_1(x), \dots, f_m(x))$. Then the function $g(\cdot) = G(f_1(\cdot), \dots, f_m(\cdot))$ is continuously codifferentiable on U , and for any $x \in U$ one can define*

$$Dg(x) = \sum_{i=1}^m \frac{\partial G}{\partial y_i}(f(x)) Df_i(x). \quad (9)$$

Moreover, for any $r > 0$ such that $B(x, r) \subset U$ and $\text{co } f(B(x, r)) \subset V$ one has

$$\begin{aligned} |\varepsilon_g(\alpha, \Delta x, x, r)| &\leq \sum_{i=1}^m \left| \frac{\partial G}{\partial y_i}(f(x)) \right| |\varepsilon_{f_i}(\alpha, \Delta x, x, r)| \\ &+ \frac{1}{\alpha} \sup_{t \in [0, 1]} \left| \langle \nabla G(y(t)) - \nabla G(f(x)), f(x + \alpha\Delta x) - f(x) \rangle \right|, \end{aligned} \quad (10)$$

where $\langle \cdot, \cdot \rangle$ is the inner product in \mathbb{R}^m , and $y(t) = tf(x) + (1-t)f(x + \alpha\Delta x)$.

Proof. Fix arbitrary $x \in U$ and $r > 0$ such that $B(x, r) \subset U$ and $\text{co } f(B(x, r)) \subset V$. Note that such $r > 0$ exists due to the fact that f_i are locally Lipschitz continuous on U by Corollary 2. Choose $\Delta x \in B(0, r)$ and $\alpha \in [0, 1]$, and denote $y(t) = tf(x) + (1-t)f(x + \alpha\Delta x)$. Applying the mean value theorem one

obtains that there exists $t \in (0, 1)$ such that

$$\begin{aligned} g(x + \alpha\Delta x) - g(x) &= \langle \nabla G(f(x)), f(x + \alpha\Delta x) - f(x) \rangle \\ &\quad + \langle \nabla G(y(t)) - \nabla G(f(x)), f(x + \alpha\Delta x) - f(x) \rangle \\ &= \sum_{i=1}^m \frac{\partial G}{\partial y_i}(f(x)) \left(\Phi_i(\alpha\Delta x) + \Psi_i(\alpha\Delta x) + \alpha\varepsilon_{f_i}(\alpha, \Delta x, x, r) \right) \\ &\quad + \langle \nabla G(y(t)) - \nabla G(f(x)), f(x + \alpha\Delta x) - f(x) \rangle, \end{aligned}$$

where Φ_i and Ψ_i are defined as in the proof of Theorem 1. Hence taking into account the fact that the functions f_i are locally Lipschitz continuous one can easily verify that the function g is codifferentiable at x , the pair $Dg(x)$ defined in (9) is a codifferential of g at x , and (10) holds true. It remains to note that codifferential mapping (9) is obviously continuous. \square

Corollary 3. *Let functions $f_i: U \rightarrow \mathbb{R}$, $i \in I = \{1, \dots, m\}$ be continuously codifferentiable on U . Then for any $\lambda_i \in \mathbb{R}$, $i \in I$ the function $f = \sum_{i \in I} \lambda_i f_i$ is continuously codifferentiable on U , and*

$$Df(x) = \sum_{i \in I} \lambda_i Df_i(x), \quad |\varepsilon_f(\alpha, \Delta x, x, r)| \leq \sum_{i \in I} |\lambda_i| \cdot |\varepsilon_{f_i}(\alpha, \Delta x, x, r)|.$$

Corollary 4. *Let functions $f_1, f_2: U \rightarrow \mathbb{R}$ be continuously codifferentiable on U . Then the function $f = f_1 \cdot f_2$ is continuously codifferentiable on U , for any $x \in U$ one can define $Df(x) = f_2(x)Df_1(x) + f_1(x)Df_2(x)$, and for any sufficiently small $r > 0$ one has*

$$\begin{aligned} |\varepsilon_f(\alpha, \Delta x, x, r)| &\leq |f_2(x)| \cdot |\varepsilon_{f_1}(\alpha, \Delta x, x, r)| + |f_1(x)| \cdot |\varepsilon_{f_2}(\alpha, \Delta x, x, r)| \\ &\quad + 2\alpha L_1 L_2 \|\Delta x\|^2, \end{aligned}$$

where $L_i > 0$ is a Lipschitz constant of f_i on $B(x, r)$.

Corollary 5. *Let a function $f: U \rightarrow \mathbb{R}$ be continuously codifferentiable on U , and $f(x) \neq 0$ for all $x \in U$. Then the function $g = 1/f$ is continuously codifferentiable on U , for any $x \in U$ one can define $Dg(x) = -f(x)^{-2}Df(x)$, and for any sufficiently small $r > 0$ one has*

$$|\varepsilon_g(\alpha, \Delta x, x, r)| \leq \frac{1}{f(x)^2} |\varepsilon_f(\alpha, \Delta x, x, r)| + \frac{2\alpha L^2 \|\Delta x\|^2}{\min\{|f(x) - rL|^3, |f(x) + rL|^3\}},$$

where $L > 0$ is a Lipschitz constant of f on $B(x, r)$.

Remark 6. Note that equalities (2) hold true automatically, if one computes a codifferential of a function f with the use of the calculus rules presented above and Examples 1–4.

In the end of this section, let us introduce a class of *uniformly codifferentiable* function that will play a very important role in the convergence analysis of the method of codifferential descent.

Definition 1. Let a function $f: U \rightarrow \mathbb{R}$ be (continuously) codifferentiable on U , and let a set $C \subset U$ be such that there exists $r > 0$ for which $B(x, r) \subseteq U$ for

all $x \in C$. One says that f is (*continuously*) *uniformly codifferentiable* on a set $C \subseteq U$ if there exists a (continuous) codifferential mapping $Df(\cdot) = [\underline{d}f(\cdot), \bar{d}f(\cdot)]$ with $\underline{d}f(\cdot), \bar{d}f(\cdot): U \rightrightarrows \mathbb{R} \times \mathcal{H}$ such that $\varepsilon_f(\alpha, \Delta x, x, r) \rightarrow 0$ as $\alpha \rightarrow +0$ uniformly for all $\Delta x \in B(0, r)$ and $x \in C$. In this case one says that the (continuous) codifferential mapping Df uniformly approximates the function f on the set C .

One can easily check that if a function $f: U \rightarrow \mathbb{R}$ is Gâteaux differentiable on U , and its Gâteaux derivative is uniformly continuous on the set $\cup_{x \in C} B(0, \tau)$ for some $\tau > 0$, then f is continuously uniformly codifferentiable on the set C . In particular, if \mathcal{H} is finite dimensional, then a continuously differentiable function f is continuously uniformly codifferentiable on any bounded set.

Observe that Theorems 1 and 2 are very useful for verifying whether a given codifferentiable function is uniformly codifferentiable on a given set. In particular, Theorem 1 and Corollary 3 imply that the set of all functions that are continuously uniformly codifferentiable on a given set C is closed under addition, multiplication by scalar, as well as the pointwise maximum and minimum of finite families of functions (i.e. this set is a vector lattice). In turn, Corollaries 4 and 5 imply that the set of all locally (i.e. in a neighbourhood of every point) continuously uniformly codifferentiable functions is closed under all standard algebraic operations. Furthermore, from Theorem 2 it follows that the function $g(x) = G(f_1(x), \dots, f_m(x))$ is continuously uniformly codifferentiable on a set $C \subset U$, provided the functions f_i are continuously uniformly codifferentiable and Lipschitz continuous on U , the function G is continuously differentiable on an open set V containing the set $\cup_{x \in C} \text{co } f(B(x, r))$ for some $r > 0$, and the gradient $\nabla G(y)$ is uniformly continuous on V . For example, the function

$$f(x_1, x_2) = \sin(\min\{x_1, x_2\}) \cdot e^{\max\{x_1, x_2\}} + |\sinh(x_1 + |x_2|)|$$

is continuously uniformly codifferentiable on any bounded subset of \mathbb{R}^2 . Moreover, with the use of Theorems 1 and 2 one can easily compute a continuous codifferential mapping $Df(\cdot)$ of the function f on \mathbb{R}^2 uniformly approximating this function on any bounded set.

3 The method of codifferential descent

In this section we present a general scheme of the method of codifferential descent, and prove the global convergence of this method.

3.1 A description of the method

Let $U \subseteq \mathcal{H}$ be an open set, and let a function $f: U \rightarrow \mathbb{R}$ be codifferentiable on U . Hereinafter, we suppose that a codifferential mapping $Df(\cdot)$ of the function f on the set U is fixed. Recall that if $x^* \in U$ is a point of local minimum of the function f , then

$$0 \in \underline{d}f(x^*) + (0, w) \quad \forall (0, w) \in \text{ext } \bar{d}f(x^*) \quad (11)$$

by Corollary 1. Any point $x^* \in U$ satisfying (11) is called *an inf-stationary point* of the function f . Note that by Proposition 1 the set of inf-stationary points of f is independent of the choice of a codifferential. Let us describe a method for finding inf-stationary points of the function f on the set U called *the method of codifferential descent* (MCD).

Remark 7. Note that our aim is to minimize the function f on an open set U , but not necessarily on the entire space \mathcal{H} . To this end, below we suppose that a sequence generated by the MCD does not leave the set U . This assumption might seem unnatural at first glance; however, it allows one to apply the results on the convergence of the MCD obtained in the article to “barrier-like” functions f , i.e. to those nonsmooth functions f which are equal to $+\infty$ outset an open set U . In particular, the convergence results below can be applied to exact barrier functions for nonsmooth optimization problems (see [24]).

For any $\nu, \mu \in [0, +\infty]$ let set-valued mappings $\underline{d}_\nu f(\cdot), \bar{d}_\mu f(\cdot): U \rightrightarrows \mathbb{R} \times \mathcal{H}$ be such that for any $x \in U$ one has

$$\begin{aligned} \{(a, v) \in \text{ext } \underline{d}f(x) \mid a \geq -\nu\} &\subseteq \underline{d}_\nu f(x) \subseteq \underline{d}f(x), \\ \{(b, w) \in \text{ext } \bar{d}f(x) \mid b \leq \mu\} &\subseteq \bar{d}_\mu f(x) \subseteq \bar{d}f(x). \end{aligned} \quad (12)$$

The pair $[\underline{d}_\nu f(x), \bar{d}_\mu f(x)]$ is called a *truncated codifferential* of f at x , and it can be viewed as a kind of approximation of both quasidifferential $\mathcal{D}f(x)$ and codifferential $Df(x)$ of f at x , depending on the values of the parameters ν and μ . Namely, one has

$$\{0\} \times \text{ext } \partial f(x) \subseteq \underline{d}_0 f(x), \quad \text{ext } \underline{d}f(x) \subseteq \underline{d}_\infty f(x) \subseteq \underline{d}f(x),$$

and similar relations hold true for $\bar{d}_\mu f(x)$. Let us note that in applications the set $\underline{d}f(x)$ is typically a convex hull of a finite set of points (a_i, v_i) . In this case one usually defines $\underline{d}_\nu f(x)$ as the set of all those (a_i, v_i) for which $a_i \geq -\nu$. The set $\bar{d}_\mu f(x)$ is defined in a similar way.

Remark 8. In theory, the best possible choice of the sets $\underline{d}_\nu f(x)$ and $\bar{d}_\mu f(x)$ is

$$\begin{aligned} \underline{d}_\nu f(x) &= \{(a, v) \in \text{ext } \underline{d}f(x) \mid a \geq -\nu\}, \\ \bar{d}_\mu f(x) &= \{(b, w) \in \text{ext } \bar{d}f(x) \mid b \leq \mu\}. \end{aligned}$$

However, in order to define the truncated codifferential this way in practice, one has to find all extreme points of the sets $\underline{d}f(x)$ and $\bar{d}f(x)$, which is a very computationally expensive procedure. That is why in some application it might be more efficient to use larger sets $\underline{d}_\nu f(x)$ and $\bar{d}_\mu f(x)$, but avoid the search of extreme points. The main goal of this article is to analyse the convergence of the MCD in the general case. That is why we do not specify the way the truncated codifferential is defined, and only impose assumptions (12) that are somewhat necessary to ensure convergence.

Let the space $\mathbb{R} \times \mathcal{H}$ be equipped with the inner product $\langle (a, v), (b, w) \rangle = ab + \langle v, w \rangle$ and the corresponding norm. The scheme of the method of codifferential descent is as follows.

1. Choose sequences $\{\nu_n\}, \{\mu_n\} \subset [0, +\infty]$, the upper bound $\alpha_* \in (0, +\infty)$ on the step size, and an initial point $x_0 \in U$.
2. n th iteration ($n \geq 0$).
 - (a) Compute $\underline{d}_{\nu_n} f(x_n)$ and $\bar{d}_{\mu_n} f(x_n)$.
 - (b) For any $z = (b, w) \in \bar{d}_{\mu_n} f(x_n)$ compute

$$\{(a_n(z), v_n(z))\} = \arg \min \left\{ \|(a, v)\| \mid (a, v) \in \text{cl co } \underline{d}_{\nu_n} f(x_n) + z \right\}.$$

(c) For any $z \in \bar{d}_{\mu_n} f(x_n)$ compute

$$\alpha_n(z) \in \arg \min \left\{ f(x_n - \alpha v_n(z)) \mid \alpha \in [0, \alpha_*]: x_n - \alpha v_n(z) \in U \right\}.$$

(d) Compute

$$z_n \in \arg \min \left\{ f(x_n - \alpha_n(z) v_n(z)) \mid z \in \bar{d}_{\mu_n} f(x_n) \right\},$$

$$\text{and define } x_{n+1} = x_n - \alpha_n(z_n) v_n(z_n).$$

Remark 9. Here the closure is taken in the norm (or, equivalently, weak) topology, and in the case $\alpha_* = +\infty$ we define $[0, \alpha_*] = [0, +\infty)$. Note also that the set $\text{cl co } \underline{d}_{\nu_n} f(x_n) \subseteq \underline{d}f(x_n)$ is weakly compact due to the weak compactness of the hypodifferential $\underline{d}f(x_n)$, which implies that the pairs $(a_n(z), v_n(z))$ are correctly defined.

Observe that at each iteration of the MCD one must perform line search in *several* directions. As we will show below, at least one of those directions is a descent direction (i.e. $f'(x_n, -v_n(z)) < 0$), provided x_n is not an inf-stationary point of the function f . Therefore, for any $n \in \mathbb{N}$ either x_n is an inf-stationary point of f or $f(x_{n+1}) < f(x_n)$. On the other hand, some of the search directions $-v_n(z)$ might not be descent directions of the function f , i.e. f may first increase and then decrease in these directions. This interesting feature of the MCD allows it to “jump over” some points of local minimum, provided the parameters ν_n and μ_n are sufficiently large (for a particular example of this phenomenon see [17], Sect. 4). Note that the fact that some of the search directions $v_n(z)$ are not descent direction forces one to define step sizes via the minimization of the function f along the directions $v_n(z)$ instead of utilizing some more widespread step size rules, such as the Armijo and the Goldstein rules.

The parameters $\nu_n, \mu_n \geq 0$ are introduced into the MCD in order to ensure convergence. If one looks at the form of the necessary optimality condition (11), then it might seem natural to utilize the MCD with $\nu_n \equiv \mu_n \equiv 0$. However, the MCD with $\nu_n \equiv \mu_n \equiv 0$ might converge to a non-stationary point of the function f (cf. [18], Section III.5, and [35], Section VIII.2.2). Let us also note that the choice of parameters ν_n and μ_n is a tradeoff between the complexity of every iteration and the overall performance of the method. In many application, a decrease of the parameters ν_n and μ_n reduces the cost of an iteration, while an increase of these parameters might allow one to find a better local solution.

Finally, from this point onwards we suppose that the step sizes $\alpha_n(z)$ and the vector z_n are well-defined in every iteration of the MCD. The vector z_n is well-defined, provided the sets $\bar{d}_{\mu_n} f(x_n)$ are finite, which is the case in almost all applications. In turn, the step sizes $\alpha_n(z)$ are well-defined if f is l.s.c., $U = \mathcal{H}$ (or $f(x) = +\infty$ outside U) and $\alpha_* < +\infty$. In the case $\alpha_* = +\infty$, one must make an additional assumption on the function f such as the boundedness of the set $\{x \in U \mid f(x) < f(x_0)\}$.

Remark 10. Let us note that the original version of the MCD [19] corresponds to the case when $\alpha_* = +\infty$, $\mu_n \equiv \mu > 0$, $\underline{d}_{\nu_n} f(x_n) \equiv \underline{d}f(x_n)$ and $\bar{d}_{\mu_n} f(x_n) \equiv \{(b, w) \in \bar{d}f(x_n) \mid b \leq \mu\}$ for some $\mu > 0$. However, a direct practical implementation of the original method is impossible, since the set $\{(b, w) \in \bar{d}f(x) \mid b \leq \mu\}$ almost always has the cardinality of the continuum

(unless f is hypodifferentiable). Note also that a direct implementation of the method of truncated codifferential [17] is impossible for a similar reason. The version of the MCD proposed in this paper is, in fact, a mathematical description of the way the original method and the method of truncated codifferential are implemented in practice.

3.2 Auxiliary Results

Before we proceed to the convergence analysis of the MCD, let us first prove a simple result about descent directions in the MCD, and two useful auxiliary lemmas that will be utilized throughout the rest of the article.

Lemma 1. *Let a sequence $\{x_n\} \subset U$ be generated by the MCD. Then for any $n \in \mathbb{N}$ and $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$ such that $0 \notin \text{cl co } \underline{d}_{\nu_n} f(x_n) + z$ one has $f'(x_n, -v_n(z)) \leq -\|(a_n(z), v_n(z))\|^2$, which implies that $v_n(z) \neq 0$, provided $a_n(z) \neq 0$. Moreover, if x_n is not an inf-stationary point of the function f , then $f(x_{n+1}) < f(x_n)$.*

Proof. Fix arbitrary $n \in \mathbb{N}$ and $z = (0, w_0) \in \bar{d}_{\mu_n} f(x_n)$ such that $0 \notin \text{cl co } \underline{d}_{\nu_n} f(x_n) + z$. Applying the necessary condition for a minimum of a convex function on a convex set one obtains that

$$aa_n(z) + \langle v, v_n(z) \rangle \geq \|(a_n(z), v_n(z))\|^2 \quad \forall (a, v) \in \text{cl co } \underline{d}_{\nu_n} f(x_n) + z. \quad (13)$$

Note that by definition one has $\{0\} \times \underline{\partial} f(x_n) \subseteq \text{cl co } \underline{d}_{\nu_n} f(x_n)$ and $w_0 \in \bar{\partial} f(x_n)$. Therefore

$$\begin{aligned} f'(x_n, -v_n(z)) &= \max_{v \in \underline{\partial} f(x_n)} \langle v, -v_n(z) \rangle + \min_{w \in \bar{\partial} f(x_n)} \langle w, -v_n(z) \rangle \\ &\leq \max_{v \in \underline{\partial} f(x_n) + w_0} \langle v, -v_n(z) \rangle \leq -\|(a_n(z), v_n(z))\|^2. \end{aligned}$$

Finally, note that if x_n is not an inf-stationary point of f , then by definition there exists $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$ such that $0 \notin \text{cl co } \underline{d}_{\nu_n} f(x_n) + z$, which implies that $f(x_n - \alpha_n(z)v_n(z)) < f(x_n)$, and hence $f(x_{n+1}) < f(x_n)$. \square

Lemma 2. *Let X be a finite dimensional normed space, and let a sequence $\{A_n\}$ of convex compact subsets of X converge to a convex compact set $A \subset X$ in the Hausdorff metric. Then for any subsequence $\{A_{n_k}\}$ one has $\text{ext } A \subseteq \limsup_{k \rightarrow \infty} \text{ext } A_{n_k}$, where \limsup is the outer limit.*

Proof. Let us verify that $A = \text{co}(\limsup_{k \rightarrow \infty} \text{ext } A_{n_k})$. Then applying a partial converse to the Krein-Milman theorem (if K is a compact convex set, and $K = \text{cl co } B$, then $\text{ext } K \subseteq \text{cl } B$; see, e.g., [29, Proposition 10.1.3]), and the fact that the outer limit set is always closed we arrive at the required result.

From the fact that $A_n \rightarrow A$ in the Hausdorff metric it obviously follows that $\text{co}(\limsup_{k \rightarrow \infty} \text{ext } A_{n_k}) \subseteq A$. Let us prove the converse inclusion. Fix an arbitrary $x \in A$. Clearly, for any $k \in \mathbb{N}$ there exists $x_k \in A_{n_k}$ such that $x_k \rightarrow x$ as $k \rightarrow \infty$. Applying the Krein-Milman and the Carathéodory theorems one obtains that for any $k \in \mathbb{N}$ there exist $y_k^i \in \text{ext } A_{n_k}$, and $\alpha_k^i \geq 0$, $1 \leq i \leq m+1$ (here m is the dimension of X) such that

$$x_k = \sum_{i=1}^{m+1} \alpha_k^i y_k^i, \quad \sum_{i=1}^{m+1} \alpha_k^i = 1. \quad (14)$$

Note that the sequence $\{A_n\}$ lies within a bounded set by virtue of the fact that A is compact, and $A_n \rightarrow A$ in the Hausdorff metric. Hence the sequences $\{y_k^i\}$, $i \in \{1, \dots, m+1\}$ are bounded. Replacing them as well as the sequences $\{\alpha_k^i\}$ with convergent subsequences, and passing to the limit in (14) one obtains that there exists $y^i \in \limsup_{k \rightarrow \infty} \text{ext } A_{n_k}$ and $\alpha^i \geq 0$ such that

$$x = \sum_{i=1}^{m+1} \alpha^i y^i, \quad \sum_{i=1}^{m+1} \alpha^i = 1.$$

Thus, $x \in \text{co}(\limsup_{k \rightarrow \infty} \text{ext } A_{n_k})$, and the proof is complete. \square

Lemma 3. *Let sequences $\{x_n\} \subset U$ and $\{\nu_n\}, \{\mu_n\} \subset [0, +\infty]$ be such that*

1. *there exists $r > 0$ for which $B(x_n, r) \subset U$ for all $n \in \mathbb{N}$;*
2. *the codifferential mapping Df uniformly approximates f on the set $\{x_n\}_{n \in \mathbb{N}}$;*
3. *the sequences $\{\underline{d}f(x_n)\}$ and $\{\bar{d}f(x_n)\}$ lie within a bounded set;*
4. *$\liminf_{n \rightarrow \infty} \nu_n = \nu^* > 0$.*

Suppose also that $\{h_n\} \subset \mathcal{H}$ is a bounded sequence satisfying the inequalities

$$\sup_{(a,v) \in \underline{d}_{\nu_n} f(x_n) + z_n} (a + \langle v, h_n \rangle) \leq -\theta \quad \forall n \in \mathbb{N}, \quad (15)$$

$$f(x_{n+1}) \leq \inf_{\alpha \in [0, \alpha_0]} f(x_n + \alpha h_n) + \varepsilon_n \quad \forall n \in \mathbb{N} \quad (16)$$

for some $z_n = (b_n, w_n) \in \bar{d}_{\mu_n} f(x_n)$, $\theta > 0$, $\alpha_0 > 0$ and ε_n such that $\varepsilon_n \rightarrow 0$ and $b_n \rightarrow 0$ as $n \rightarrow \infty$. Then $f(x_n) \rightarrow -\infty$ as $n \rightarrow \infty$.

Proof. By definition for any $n \in \mathbb{N}$ one has

$$\begin{aligned} f(x_n + \alpha h_n) - f(x_n) &= \max_{(a,v) \in \underline{d}f(x_n)} (a + \alpha \langle v, h_n \rangle) + \min_{(b,w) \in \bar{d}f(x_n)} (b + \alpha \langle w, h_n \rangle) \\ &+ \alpha \varepsilon_n(\alpha) \leq \max_{(a,v) \in \underline{d}f(x_n) + z_n} (a + \alpha \langle v, h_n \rangle) + \alpha \varepsilon_n(\alpha), \end{aligned} \quad (17)$$

where $\varepsilon_n(\alpha) = \varepsilon_f(\|h_n\|\alpha/r, \Delta x_n, x_n, r)$, $\Delta x_n = r h_n / \|h_n\| \in B(x_n, r) \subset U$, and $0 \leq \alpha \leq \hat{\alpha} := \min\{1, r / \sup_n \|h_n\|\}$ is arbitrary (note that $\hat{\alpha} > 0$ due to the fact that the sequence $\{h_n\}$ is bounded). Our aim is to prove that there exist $n_1 \in \mathbb{N}$ and $\alpha_1 \in (0, \hat{\alpha}]$ such that for all $n \geq n_1$ and $\alpha \in [0, \alpha_1]$ the set $\underline{d}f(x_n) + z_n$ in (17) can be replaced by the smaller set $\underline{d}_{\nu_n} f(x_n) + z_n$, i.e.

$$f(x_n + \alpha h_n) - f(x_n) \leq \sup_{(a,v) \in \underline{d}_{\nu_n} f(x_n) + z_n} (a + \alpha \langle v, h_n \rangle) + \alpha \varepsilon_n(\alpha) \quad (18)$$

for all $n \geq n_1$ and $\alpha \in (0, \alpha_1]$.

Before we turn to the proof of inequality (18), let us first demonstrate that the validity of the lemma follows directly from this inequality. Indeed, denote

$$\eta_n(\alpha) = \sup_{(a,v) \in \underline{d}_{\nu_n} f(x_n) + z_n} (a + \alpha \langle v, h_n \rangle).$$

From (18) and the convexity of $\eta_n(\alpha)$ it follows that

$$\begin{aligned} f(x_n + \alpha h_n) - f(x_n) &\leq \alpha \eta_n(1) + (1 - \alpha) \eta_n(0) + \alpha \varepsilon_n(\alpha) \\ &= \alpha \sup_{(a,v) \in \underline{d}_{\nu_n} f(x_n) + z_n} (a + \langle v, h_n \rangle) + (1 - \alpha) \eta_n(0) + \alpha \varepsilon_n(\alpha) \end{aligned}$$

for all $n \geq n_1$ and $\alpha \in [0, \alpha_1]$. Hence applying (15) and taking into account the fact that $\eta_n(0) = b_n$ due to (2) one obtains that

$$f(x_n + \alpha h_n) - f(x_n) \leq -\alpha \theta + (1 - \alpha) b_n + \alpha \varepsilon_n(\alpha)$$

for all $n \geq n_1$ and $\alpha \in [0, \alpha_1]$. Recall that the codifferential mapping Df uniformly approximates the function f on the set $\{x_n\}_{n \in \mathbb{N}}$. Therefore there exists $\alpha_2 > 0$ such that for all $n \in \mathbb{N}$ and $\alpha \in [0, \alpha_2]$ one has $|\varepsilon_n(\alpha)| < \theta/3$. Hence for any $n \geq n_1$ one has

$$f(x_n + \gamma h_n) - f(x_n) \leq -\gamma \frac{2\theta}{3} + (1 - \gamma) b_n, \quad \gamma = \min\{\alpha_0, \alpha_1, \alpha_2\}.$$

By our assumption $b_n \rightarrow 0$ as $n \rightarrow \infty$. Therefore for any sufficiently large $n \in \mathbb{N}$ one has

$$f(x_n + \gamma h_n) - f(x_n) \leq -\gamma \frac{\theta}{3},$$

which with the use of (16) implies that $f(x_{n+1}) \leq f(x_n) - \gamma\theta/3 + \varepsilon_n$ for all n large enough. Consequently, taking into account the fact that $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$ one obtains that $f(x_n) \rightarrow -\infty$ as $n \rightarrow \infty$.

Thus, it remains to verify that inequality (18) holds true. Denote

$$g_n(\alpha) = \max_{(a,v) \in \underline{d}f(x_n) + z_n} (a + \alpha \langle v, h_n \rangle).$$

Let us prove that there exist $n_1 \in \mathbb{N}$ and $\alpha_1 > 0$ such that $g_n(\alpha) = \eta_n(\alpha)$ for all $n \geq n_1$ and $\alpha \in [0, \alpha_1]$. Then (18) follows directly from (17).

From the inclusion $\underline{d}_{\nu_n} f(x_n) \subseteq \underline{d}f(x_n)$ it follows that $g_n(\alpha) \geq \eta_n(\alpha)$ for all $\alpha \geq 0$ and $n \in \mathbb{N}$. We need to check that the converse inequality holds true for any sufficiently small $\alpha \geq 0$, and for all $n \in \mathbb{N}$ large enough. It is clear that

$$\max_{(a,v) \in \underline{d}f(x_n)} (a + \alpha \langle v, h_n \rangle) = \sup_{(a,v) \in \text{ext } \underline{d}f(x_n)} (a + \alpha \langle v, h_n \rangle). \quad (19)$$

By our assumption the sets $\underline{d}f(x_n)$ and $\overline{d}f(x_n)$ lie within a bounded set K . Define $C_1 = \sup_{z \in K} \|z\|$, $C_2 = \sup_{n \in \mathbb{N}} \|h_n\|$ and $\alpha_1 = \nu^*/4C_1C_2$. Then for any $n \in \mathbb{N}$, $\alpha \in [0, \alpha_1]$ and $(a, v) \in \text{ext } \underline{d}f(x_n)$ one has $\alpha \langle v, h_n \rangle \geq -\nu^*/4$, which implies that

$$a + \alpha \langle v, h_n \rangle \begin{cases} \geq -0.5\nu^*, & \text{if } a \geq -0.25\nu^*, \\ < -0.5\nu^*, & \text{if } a < -0.75\nu^*. \end{cases}$$

Therefore, for any $n \in \mathbb{N}$ and $\alpha \in [0, \alpha_1]$ one has

$$\sup_{(a,v) \in \text{ext } \underline{d}f(x_n)} (a + \alpha \langle v, h_n \rangle) = \sup_{(a,v) \in \text{ext } \underline{d}f(x_n): a \geq -0.75\nu^*} (a + \alpha \langle v, h_n \rangle) \geq -\frac{\nu^*}{2}.$$

By the definition of ν^* there exists $n_1 \in \mathbb{N}$ such that $\nu_n \geq 0.75\nu^*$ for all $n \geq n_1$. Consequently, taking into account the definition of $\underline{d}_{\nu} f(x)$ and equality (19) one obtains that $g_n(\alpha) = \eta_n(\alpha)$ for all $\alpha \in [0, \alpha_1]$ and $n \geq n_1$, and the proof is complete. \square

3.3 Global convergence

Now we can prove the global convergence of the MCD.

Theorem 3. *Suppose that $x^* \in U$ is a cluster point of a sequence $\{x_n\}$ generated by the MCD, the codifferential mapping Df is continuous at x^* , and uniformly approximates the function f in a neighbourhood of this point. Let also f be bounded below on U , $\liminf_{n \rightarrow \infty} \nu_n > 0$ and $\liminf_{n \rightarrow \infty} \mu_n > 0$. Suppose finally that one of the two following assumptions holds true:*

1. \mathcal{H} is finite dimensional;
2. $\{(b, w) \in \bar{d}f(x_n) \mid b \leq \hat{\mu}\} \subseteq \bar{d}_{\mu_n} f(x_n)$ for some $\hat{\mu} > 0$, and for all sufficiently large $n \in \mathbb{N}$.

Then x^ is an inf-stationary point of the function f . If, in addition, f is convex, then x^* is a point of global minimum of the function f .*

Proof. Arguing by reductio ad absurdum, suppose that x^* is not an inf-stationary point of the function f . Then there exists $z^* = (0, w^*) \in \text{ext } \bar{d}f(x^*)$ such that $\theta = \min\{\|(a, v)\|^2 \mid (a, v) \in \underline{d}f(x^*) + z^*\} > 0$. Our aim is to apply Lemma 3.

Since x^* is a cluster point of the sequence $\{x_n\}$, there exists a subsequence $\{x_{n_k}\}$ converging to x^* . Therefore $\bar{d}f(x_{n_k}) \rightarrow \bar{d}f(x^*)$ as $k \rightarrow \infty$ in the Hausdorff metric due to the continuity of the codifferential mapping Df at x^* . Hence applying Lemma 2 and the fact that $\liminf_{n \rightarrow \infty} \mu_n > 0$ in the case when \mathcal{H} is finite dimensional or assumption 2 in the case when \mathcal{H} is infinite dimensional one obtains that there exists a subsequence of the sequence $\{x_{n_k}\}$, which we denote again by $\{x_{n_k}\}$, and there exists $z_{n_k} = (b_{n_k}, w_{n_k}) \in \bar{d}_{\mu_{n_k}} f(x_{n_k})$ such that $z_{n_k} \rightarrow z^*$, i.e. $b_{n_k} \rightarrow 0$ and $w_{n_k} \rightarrow w^*$, as $k \rightarrow \infty$. Consequently, taking into account the fact that the multifunction $\underline{d}f(\cdot)$ is Hausdorff continuous at x^* by our assumption one obtains that $\|(a_{n_k}(z_{n_k}), v_{n_k}(z_{n_k}))\|^2 > \theta/2$ for all k greater than some $k_1 \in \mathbb{N}$.

Denote $a_k = a_{n_k}(z_{n_k})$ and $v_k = v_{n_k}(z_{n_k})$. From the definition of (a_k, v_k) , and the necessary and sufficient conditions for a minimum of a convex function on a convex set it follows that

$$-(a + b_{n_k})a_k - \langle v + w_{n_k}, v_k \rangle \leq -\|(a_k, v_k)\|^2 < -\frac{\theta}{2} \quad (20)$$

for all $(a, v) \in \underline{d}_{\nu_{n_k}} f(x_{n_k})$ and $k \geq k_1$. Therefore

$$\sup_{(a, v) \in \underline{d}_{\nu_{n_k}} f(x_{n_k}) + z_{n_k}} (a + \alpha \langle v, -v_k \rangle) \leq -\alpha \frac{\theta}{2} + \max_{(a, v) \in \underline{d}_{\nu_{n_k}} f(x_{n_k})} ((a + b_{n_k})(1 + \alpha a_k)).$$

for all $\alpha \geq 0$ and $k \geq k_1$. Taking into account the facts that the codifferential mapping Df is continuous at x^* , and $x_{n_k} \rightarrow x^*$ as $k \rightarrow \infty$ one obtains that the sequences $\{\underline{d}f(x_{n_k})\}$ and $\{\bar{d}f(x_{n_k})\}$ lie within a bounded set, which, in particular, implies that the sequences $\{a_k\}$ and $\{v_k\}$ are bounded. Consequently, there exists $\gamma > 0$ such that $-1 < \gamma a_k < 1$ for all $k \geq k_1$, which implies that

$$\sup_{(a, v) \in \underline{d}_{\nu_{n_k}} f(x_{n_k}) + z_{n_k}} (a + \langle v, -\gamma v_k \rangle) \leq -\gamma \frac{\theta}{2} + (1 + \gamma a_k) b_{n_k} \quad \forall k \geq k_1$$

(here we used equalities (2)). By definition $b_{n_k} \rightarrow 0$ as $k \rightarrow \infty$. Therefore there exists $k_2 \geq k_1$ such that $b_{n_k} < \gamma\theta/4$ for all $k \geq k_2$.

Observe that since $x_{n_k} \rightarrow x^*$ as $k \rightarrow \infty$ and $x^* \in U$, there exist $k_3 \in \mathbb{N}$ and $r > 0$ such that $B(x_{n_k}, r) \subset U$ for all $k \geq k_3$. Furthermore, from the fact that the codifferential mapping Df uniformly approximates the function f in a neighbourhood of x^* it follows that Df uniformly approximates f on the set $\{x_{n_k}\}_{k \geq k_4}$ for some $k_4 \in \mathbb{N}$. Therefore applying Lemma 3 with $\alpha_0 = \alpha_*$ (recall that α_* is the upper bound on the step size in the MCD), the sequence $\{x_n\}$ defined as $\{x_{n_k}\}_{k \geq m}$, and the sequence $\{h_n\}$ defined as $\{-\gamma v_k\}_{k \geq m}$, where $m = \max\{k_2, k_3, k_4\}$, one obtains that $f(x_{n_k}) \rightarrow -\infty$ as $k \rightarrow \infty$, which is impossible due to the fact that f is bounded below on U (note that the validity of inequality (16) in this case follows directly from the definitions of $\alpha_n(z)$ and x_{n+1} in the MCD, and the fact that $f(x_{n+1}) \leq f(x_n)$ for all $n \in \mathbb{N}$).

It remains to note that if f is convex, and x^* is an inf-stationary point of f , then $f'(x^*, \cdot) \geq 0$ by Proposition 1, which implies that x^* is a point of global minimum of f due to the convexity assumption. \square

Remark 11. For the validity of the theorem above it is sufficient to suppose that $\{0\} \times \text{ext } \bar{\partial}f(x^*) \subseteq \limsup_{k \rightarrow \infty} \text{ext } \bar{\partial}f(x_{n_k})$, where $x_{n_k} \rightarrow x^*$ as $k \rightarrow \infty$. Lemma 2 guarantees that this inclusion always holds true in the finite dimensional case. In order to ensure the validity of this inclusion in the infinite dimensional case we utilized assumption 2, but it should be noted that in many applications the validity of this inclusion can be verified directly.

Corollary 6. *Let $\liminf_{n \rightarrow \infty} \nu_n > 0$, $\liminf_{n \rightarrow \infty} \mu_n > 0$, and f be bounded below on U . Suppose also that the codifferential mapping Df is continuous on U and locally uniformly approximates the function f on the set U . Let finally one of the two following assumptions be valid:*

1. \mathcal{H} is finite dimensional;
2. $\{(b, w) \in \bar{\partial}f(x_n) \mid b \leq \hat{\mu}\} \subseteq \bar{\partial}_{\mu_n} f(x_n)$ for some $\hat{\mu} > 0$, and for all sufficiently large $n \in \mathbb{N}$.

Then any cluster point $x^ \in U$ of the sequence $\{x_n\}$ generated by the MCD is an inf-stationary point of the function f .*

It is well-known that under some natural assumptions for any gradient method one has $\lim_{n \rightarrow \infty} \|\nabla f(x_n)\| = 0$, where $\{x_n\}$ is a sequence generated by this method (see, e.g., [49], Theorem 2.5). Let us extend this result to the case of the MCD. To this end, for any $\nu \geq 0$ introduce the function

$$\omega(x, \nu) = \sup_{(0, w) \in \bar{\partial}f(x)} \min_{u \in \text{cl co } \underline{d}_\nu f(x) + (0, w)} \|u\|^2$$

that, in a sense, measures how far a point x is from being an inf-stationary point of the function f . In particular, x is an inf-stationary point of f iff $\omega(x, \nu) = 0$ for some $\nu \geq 0$. It should be noted that the function $\omega(x, \nu)$ is not continuous (or even l.s.c.) in x in the general case, even if the function f is continuously codifferentiable. But ω is continuous in x , if f is continuously hypodifferentiable, and $\nu = +\infty$, since in this case $\omega(x, +\infty) = \text{dist}(0, \underline{d}f(x))$.

Theorem 4. Let f be bounded below on U , a sequence $\{x_n\}$ be generated by the MCD, and the sequences $\{\underline{d}f(x_n)\}$ and $\{\bar{d}f(x_n)\}$ be bounded. Suppose also that $\liminf_{n \rightarrow \infty} \nu_n = \nu^* > 0$, there exists $r > 0$ such that $B(x_n, r) \subset U$ for all $n \in \mathbb{N}$, and the codifferential mapping Df uniformly approximates the function f on the set $\{x_n\}_{n \geq m}$ for some $m \in \mathbb{N}$ (in particular, one can suppose that Df uniformly approximates f on the set $\{x \in U \mid f(x) \leq f(x_0)\}$). Then $\omega(x_n, \nu_n) \rightarrow 0$ as $n \rightarrow \infty$. In particular, if f is hypodifferentiable, i.e. $Df(\cdot) = [\underline{d}f(\cdot), \{0\}]$, and $\nu_n \equiv +\infty$, then $\text{dist}(0, \underline{d}f(x_n)) \rightarrow 0$ as $n \rightarrow \infty$.

Proof. Arguing by reductio ad absurdum, suppose that the theorem is false. Then there exist a subsequence $\{x_{n_k}\}$ and $\theta > 0$ such that $\omega(x_{n_k}, \nu_{n_k}) > \theta$ for all $k \in \mathbb{N}$. By the definition of ω for any $k \in \mathbb{N}$ there exists $z_{n_k} = (0, w_{n_k}) \in \bar{d}_{\nu_{n_k}} f(x_{n_k})$ such that $\|(a_{n_k}(z_{n_k}), v_{n_k}(z_{n_k}))\|^2 \geq \theta$. Then taking into account the fact that the sequences $\{\underline{d}f(x_n)\}$ and $\{\bar{d}f(x_n)\}$ are bounded, and arguing in the same way as in the proof of Theorem 3 one can easily check that there exists $\gamma > 0$ such that

$$\sup_{(a,v) \in \underline{d}_{\nu_{n_k}} f(x_{n_k}) + z_{n_k}} (a + \langle v, -\gamma v_{n_k}(z_{n_k}) \rangle) \leq -\gamma\theta \quad \forall k \in \mathbb{N}.$$

Consequently, applying Lemma 3 one obtains that $f(x_{n_k}) \rightarrow -\infty$ as $k \rightarrow \infty$, which contradicts the assumption that f is bounded below on U . \square

Remark 12. It is easy to verify that Theorems 3 and 4 remain to hold true in the case when the search directions $v_n(z_n)$ are replaced by some approximations $\tilde{v}_n(z_n)$ such that

$$\|\tilde{v}_n(z_n) - v_n(z_n)\| \leq \varepsilon_n, \quad (21)$$

where $\varepsilon_n \rightarrow +0$ as $n \rightarrow \infty$. Namely, note that in this case one has

$$-(a + b_{n_k})a_{n_k}(z_{n_k}) - \langle v + w_{n_k}, \tilde{v}_{n_k}(z_{n_k}) \rangle \leq -\frac{\theta}{2} + \varepsilon_{n_k} C$$

for a sufficiently large $C > 0$, and for all $(a, v) \in \underline{d}_{\nu_{n_k}} f(x_{n_k})$ and $k \geq k_1$ (see (20)). With the use of this estimate and the fact that $\varepsilon_{n_k} \rightarrow 0$ one can easily obtain the required results. In particular, one can extend Theorems 3 and 4 to the case when instead of the sets $\underline{d}f(x_n)$ and $\bar{d}f(x_n)$ one uses their approximations, provided these approximations are “good enough”, i.e. provided inequality (21) holds true for the corresponding approximate search directions.

Similarly, Theorems 3 and 4 remain to hold true if the step sizes $\alpha_n(z)$ are computed only approximately. Namely, it is sufficient to suppose that

$$\begin{aligned} & f(x_n - \alpha_n(z)v_n(z)) \\ & \leq \inf \left\{ f(x_n - \alpha_n(z)v_n(z)) \mid \alpha \in [0, \alpha_*]: x_n - \alpha_n(z)v_n(z) \in U \right\} + \varepsilon_n, \end{aligned}$$

where $\varepsilon_n \rightarrow +0$ as $n \rightarrow \infty$. Thus, one can say that the MCD is somewhat robust with respect to computational errors.

Note that in the general case for some $z = (b, w) \in \bar{d}_{\mu_n} f(x_n)$ with $b > 0$ the corresponding search direction $-v_n(z)$ might not be a descent direction of the function f , i.e. $f'(x_n, -v_n(z)) > 0$.

Example 5. Let $\mathcal{H} = \mathbb{R}^2$, $x_0 = (0, 0)$ and

$$f(x_1, x_2) = \max\{x_1 + x_2, x_1^2 + x_2^2 - 1\} + \min\{x_1^3 + x_2^3, -2x_1 - x_2 + 1, -x_1 - 2x_2 + 2\},$$

With the use of Theorem 1, Corollary 3 and Example 1 one gets

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

(note that equalities (2) hold true). Let $\nu = 0.5$ and $\mu = 1$, and define

$$\underline{d}_\nu f(x_0) = \left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}, \quad \bar{d}_\mu f(x_0) = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} \right\}.$$

Observe that f is differentiable at x_0 and $\nabla f(x_0) = (1, 1)$. Hence for $z = (1, -2, -1) \in \bar{d}_\mu f(x_0)$ one has $v_0(z) = (-1, 0)$, and $f'(x_0, -v_0(z)) = 1$.

Thus, the only reasonable step size rule in the case when $z = (b, w) \in \bar{d}_{\mu_n} f(x_n)$ is such that $b > 0$ is the minimization along the direction $-v_n(z)$ over some interval $[0, \alpha_*]$. It should be noted that such directions $v_n(z)$ cannot be excluded, since otherwise the MCD might converge to a non-stationary point (see the proof of Theorem 3). Let us also note that the choice of α_* is usually heuristic, since if α_* is too small, then one might not benefit from the use of a non-decent direction $v_n(z)$ (i.e. the method would be unable to “jump over” a local minimum), while if α_* is too large, then the line search might take unreasonably long time.

Note finally that Lemma 1 implies that for any $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$ one has $f'(x_n, -v_n(z)) \leq -\|(a_n(z), v_n(z))\|^2$. In this case one can compute the step size $\alpha_n(z)$ as follows:

$$\alpha_n(z) = \max_{k \in \mathbb{N} \cup \{0\}} \left\{ \gamma^k \mid f(x_n - \gamma^k v_n(z)) - f(x_n) \leq -\sigma \gamma^k \|(a_n(z), v_n(z))\|^2 \right\}. \quad (22)$$

Here $\sigma, \gamma \in (0, 1)$ are fixed. It is easy to see that if Df uniformly approximates the function f in a neighbourhood of a point x^* , and a sequence $\{x_n\}$ generated by the MCD converges to x^* , then for any x_n in a neighbourhood of x^* step sizes (22) are bounded away from zero. With the use of this result one can verify that Theorems 3 and 4 remain to hold true if in the MCD for any $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$ one uses the step size rule (22).

4 The quadratic regularization of the MCD

In this section we present and analyse a different method for minimizing a codifferentiable function f , which we call *the quadratic regularization of the MCD*. In the case when f is the maximum of a finite family of continuously differentiable functions, this method coincides with the PPP algorithm [53, 51, 13, 52]. Since the quadratic regularization of the MCD can be easily applied not only to the problem of unconstrained minimization of a codifferentiable function, but also to the problem of minimizing a codifferentiable function over a convex set, below we consider the constrained version of the problem.

Let, as above, $U \subseteq \mathcal{H}$ be an open set, a function $f: U \rightarrow \mathbb{R}$ be codifferentiable on U , and Df be its fixed codifferential mapping. Let also $A \subset U$ be a closed convex set. Below, we study the problem of minimizing the function f over the set A . Let us first obtain necessary optimality conditions for this problem.

For any $x \in U$ and $z \in \bar{d}f(x)$ introduce the continuous convex function

$$\varphi(h, z, x, \nu) = \max_{(a,v) \in \text{cl co } \underline{d}_\nu f(x)+z} (a + \langle v, h \rangle) + \frac{1}{2} \|h\|^2 \quad \forall h \in \mathcal{H}. \quad (23)$$

In the case when some sequences $\{\nu_n\} \subset [0, +\infty]$ and $\{x_n\} \subset U$ are given, we denote $\varphi_n(h, z) = \varphi(h, z, x_n, \nu_n)$. Let us note that the quadratic term is introduced into the definition of the function $\varphi(h, z, x, \nu)$ in order to ensure that this function attains a global minimum in h on the set $A - x$.

Proposition 3. *Let $x^* \in A$ be a point of local minimum of the function f on the set A . Then for any $\nu \geq 0$ one has*

$$\{0\} = \arg \min_{h \in A - x^*} \varphi(h, z, x^*, \nu) \quad \forall z = (0, w) \in \text{ext } \bar{d}f(x^*). \quad (24)$$

Furthermore, (24) holds true iff one of the two following statements is valid:

1. 0 is a global minimizer of the function $\varphi(\cdot, z, x^*, \nu) - \|\cdot\|^2/2$ on the set $A - x^*$;
2. $f'(x^*, h) \geq 0$ for all $h \in A - x^*$.

In particular, optimality condition (24) is independent of the choice of a codifferential and parameter $\nu \geq 0$.

Proof. Let $z = (0, w) \in \text{ext } \bar{d}f(x^*)$ be arbitrary. Applying [27], Theorem 2.8 (see also [26], Theorem 5) one obtains that 0 is a point of global minimum of the function $\varphi(\cdot, z, x^*, \nu) - \|\cdot\|^2/2$ on the set $A - x^*$. Hence taking into account the fact that the subdifferential of this convex function at the origin coincides with the subdifferential of the function $\varphi(\cdot, z, x^*, \nu)$ at the origin one obtains that (24) holds true. Furthermore, from the coincidence of the subdifferentials at the origin it follows that (24) holds true iff condition 1 holds true.

Applying the theorem about the subdifferential of the supremum of a family of convex functions (see, e.g., [36, Thm. 4.2.3]), the definition of $\underline{d}_\nu f(x)$ and the first equality in (2) one obtains that $\partial_h \varphi(0, z, x^*, \nu) = \underline{d}f(x^*) + w$, where $\partial_h \varphi(0, z, x^*, \nu)$ is the subdifferential (in the sense of convex analysis) of the function $\varphi(\cdot, z, x^*, \nu)$ at the origin. Hence with the use of the standard necessary and sufficient condition for a minimum of a convex function on a convex set one obtains that

$$\max_{v \in \underline{d}f(x^*)+w} \langle v, h \rangle \geq 0 \quad \forall h \in A - x^* \quad \forall (0, w) \in \text{ext } \bar{d}f(x^*).$$

Taking the infimum over all $w \in \text{ext } \bar{d}f(x^*)$ (clearly, $\text{ext } \bar{d}f(x^*) = \{w \in \mathcal{H} \mid (0, w) \in \text{ext } \bar{d}f(x^*)\}$), and applying the Krein-Milman theorem one finds that

$$f'(x^*, h) = \max_{v \in \underline{d}f(x^*)} \langle v, h \rangle + \min_{w \in \bar{d}f(x^*)} \langle w, h \rangle \geq 0 \quad \forall h \in A - x^*. \quad (25)$$

Arguing backwards one can check that if (25) is valid, then optimality condition (24) holds true as well. \square

A point $x^* \in A$ satisfying optimality condition

$$\{0\} = \arg \min_{h \in A - x^*} \varphi(h, z, x^*, \nu) \quad \forall z = (0, w) \in \text{ext } \bar{d}f(x^*).$$

for some $\nu \geq 0$ is called an inf-stationary point of the function f on the set A . With the use of this optimality condition, which is independent of the choice of a codifferential and $\nu \geq 0$, we can design the quadratic regularization of the MCD (QR-MCD). The scheme of this method is as follows.

1. Choose sequences $\{\nu_n\}, \{\mu_n\} \subset [0, +\infty]$, the upper bound $\alpha_* \in (0, +\infty)$ on the step size, and an initial point $x_0 \in A$.
2. n th iteration ($n \geq 0$).

(a) Compute $\underline{d}_{\nu_n} f(x_n)$ and $\bar{d}_{\mu_n} f(x_n)$.

(b) For any $z = (b, w) \in \bar{d}_{\mu_n} f(x_n)$ compute

$$\{h_n(z)\} = \arg \min \left\{ \varphi_n(h, z) \mid h \in A - x_n \right\}$$

(c) For any $z \in \bar{d}_{\mu_n} f(x_n)$ compute

$$\alpha_n(z) \in \arg \min \left\{ f(x_n + \alpha h_n(z)) \mid \alpha \in [0, \alpha_*]: x_n + \alpha h_n(z) \in A \right\}.$$

(d) Compute

$$z_n \in \arg \min \left\{ f(x_n + \alpha_n(z) h_n(z)) \mid z \in \bar{d}_{\mu_n} f(x_n) \right\},$$

and define $x_{n+1} = x_n + \alpha_n(z_n) h_n(z_n)$.

Hereinafter, we suppose that the step sizes $\alpha_n(z)$, and the vectors z_n are correctly defined.

Remark 13. Observe that the function $\varphi_n(\cdot, z)$ is strictly convex, continuous (since it is obviously bounded on bounded sets), and $\varphi_n(h, z) \rightarrow +\infty$ as $\|h\| \rightarrow +\infty$. Therefore taking into account the facts that \mathcal{H} is a Hilbert space, and the convex set A is closed, one obtains that the search directions $h_n(z)$ are well-defined. Furthermore, note that x_n is an inf-stationary point of the function f on the set A iff $h_n(z) = 0$ for all $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$.

Let us first extend Lemma 1 to the the case of the QR-MCD.

Lemma 4. *Let a sequence $\{x_n\}$ be generated by the QR-MCD. Then for any $n \in \mathbb{N}$ and $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$ such that $h_n(z) \neq 0$ one has $f'(x_n, h_n(z)) \leq -\|h_n(z)\|^2$. In particular, if x_n is not an inf-stationary point of the function f on the set A , then $f(x_{n+1}) < f(x_n)$.*

Proof. Fix an arbitrary $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$ such that $h_n(z) \neq 0$. Equalities (2) imply that $\varphi_n(0, z) = 0$. Therefore $\varphi_n(h_n(z), z) \leq 0$ or, equivalently,

$$\max_{(a, v) \in \text{cl co } \underline{d}_{\nu_n} f(x_n) + z} (a + \langle v, h_n(z) \rangle) \leq -\frac{1}{2} \|h_n(z)\|^2.$$

Recall that by definition $\{0\} \times \text{ext } \underline{d}f(x_n) \subseteq \underline{d}_{\nu_n} f(x_n)$. Consequently, applying the inequality above and the definition of quasidifferential one obtains that

$$f'(x_n, h_n(z)) \leq \max_{v \in \underline{d}f(x_n) + w} \langle v, h_n(z) \rangle \leq -\frac{1}{2} \|h_n(z)\|^2,$$

which completes the proof. \square

Now we can prove the global convergence of the QR-MCD.

Theorem 5. *Suppose that x^* is a cluster point of a sequence $\{x_n\} \subset A$ generated by the QR-MCD, the codifferential mapping Df is continuous at x^* , and uniformly approximates the function f in a neighbourhood of this point. Let also f be bounded below on A , $\liminf_{n \rightarrow \infty} \nu_n > 0$ and $\liminf_{n \rightarrow \infty} \mu_n > 0$. Suppose finally that one of the two following assumptions holds true:*

1. \mathcal{H} is finite dimensional;
2. $\{(b, w) \in \bar{d}f(x_n) \mid b \leq \hat{\mu}\} \subseteq \bar{d}_{\mu_n} f(x_n)$ for some $\hat{\mu} > 0$ and for all sufficiently large $n \in \mathbb{N}$.

Then x^ is an inf-stationary point of the function f on the set A . If, in addition, f is convex, then x^* is a point of global minimum of f on A .*

Proof. Arguing by reductio ad absurdum, suppose that x^* is not an inf-stationary point of the function f on the set A . Then there exists $z^* = (0, w^*) \in \text{ext } \bar{d}f(x^*)$ such that $\min_{h \in A - x^*} \varphi(h, z^*, x^*, +\infty) = -\theta < 0$. Denote by h^* a point of global minimum of the function $\varphi(\cdot, z^*, x^*, +\infty)$ on the set $A - x^*$. Our aim is to apply Lemma 3.

Arguing in the same way as in the proof of Theorem 3 one can check that there exist a subsequence $\{x_{n_k}\}$ and a sequence $z_{n_k} = (b_{n_k}, w_{n_k}) \in \bar{d}_{\mu_{n_k}} f(x_{n_k})$ such that $x_{n_k} \rightarrow x^*$ and $z_{n_k} \rightarrow z^*$ as $k \rightarrow \infty$. From the continuity of the codifferential mapping Df at x^* it follows that $\underline{d}f(x_{n_k}) \rightarrow \underline{d}f(x^*)$ in the Hausdorff metric. Applying this fact it is easy to deduce that there exists $k_0 \in \mathbb{N}$ such that

$$|\varphi(h^*, z_{n_k}, x_{n_k}, +\infty) - \varphi(h^*, z^*, x^*, +\infty)| < \frac{\theta}{2} \quad \forall k \geq k_0,$$

which implies that

$$\varphi_{n_k}(h^*, z_{n_k}) \leq \varphi(h^*, z_{n_k}, x_{n_k}, +\infty) < -\frac{\theta}{2} \quad \forall k \geq k_0,$$

and, moreover, $\varphi_{n_k}(h_{n_k}(z_{n_k}), z_{n_k}) < -\theta/2$ for all $k \geq k_0$. Therefore

$$\sup_{(a, v) \in \underline{d}_{\nu_{n_k}} f(x_{n_k}) + z_{n_k}} (a + \langle v, h_{n_k}(z_{n_k}) \rangle) < \varphi_{n_k}(h_{n_k}(z_{n_k}), z_{n_k}) < -\frac{\theta}{2}. \quad (26)$$

for all $k \geq k_0$.

Clearly, the sequences $\{\underline{d}f(x_{n_k})\}$ and $\{\bar{d}f(x_{n_k})\}$ are bounded due to the continuity of the codifferential mapping Df at x^* . Therefore there exist $c_1, c_2 \in \mathbb{R}$ such that $\varphi_{n_k}(h, z_{n_k}) \geq 0.5\|h\|^2 + c_1\|h\| + c_2$ for all $h \in \mathcal{H}$ and $k \in \mathbb{N}$, which implies that the sequence $\{h_{n_k}(z_{n_k})\}$ is bounded. Hence taking into account inequality (26), and the fact that Df uniformly approximates f in a neighbourhood of x^* , and applying Lemma 3 one obtains that $f(x_{n_k}) \rightarrow -\infty$ as $k \rightarrow \infty$, which is impossible by virtue of the fact that f is bounded below on A . \square

Similarly to the case of the MCD, introduce the function

$$\omega_2(x, \nu) = \sup_{z=(0,w) \in \bar{d}f(x)} \|h(z, x, \nu)\|^2,$$

that measures how far a point x is from being an inf-stationary point of the function f on the set A . Here $h(z, x, \nu)$ is a point of global minimum of the function $\varphi(\cdot, z, x, \nu)$ on the set $A - x$. It is easy to see that x^* is an inf-stationary point of the function f on the set A iff $\omega_2(x^*, \nu) = 0$ for some $\nu \geq 0$.

Arguing in a similar way to the proof of Theorem 4 one can easily verify that the following result holds true.

Theorem 6. *Let f be bounded below on A , a sequence $\{x_n\} \subset A$ be generated by the QR-MCD, and the sequences $\{\underline{d}f(x_n)\}$ and $\{\bar{d}f(x_n)\}$ be bounded. Suppose also that there exists $r > 0$ such that $B(x_n, r) \subset U$ for all $n \in \mathbb{N}$, the codifferentiable mapping Df uniformly approximates the function f on the set $\{x_n\}_{n \geq m}$ for some $m \in \mathbb{N}$, and $\liminf_{n \rightarrow \infty} \nu_n > 0$. Then $\omega_2(x_n, \nu_n) \rightarrow 0$ as $n \rightarrow \infty$.*

Remark 14. (i) As in the case of the MCD, one can verify that Theorems 5 and 6 remain to hold true if instead of the search directions $h_n(z)$ one uses their approximations $\tilde{h}_n(z)$ such that $\|\tilde{h}_n(z) - h_n(z)\| < \varepsilon_n$, where $\varepsilon_n \rightarrow +0$ as $n \rightarrow \infty$. In order to prove this result one needs to note that if the sequences $\{\underline{d}f(x_n)\}$ and $\{\bar{d}f(x_n)\}$ are bounded, then the corresponding functions $\varphi_n(h, z)$ are uniformly (in n) bounded on any bounded set, which, as it is well-known from convex analysis, implies that these functions are Lipschitz continuous on any bounded set with a Lipschitz constant independent of n . Also, one can verify that Theorems 5 and 6 remain valid, if one uses the step size rule similar to step size rule (22) for any $z = (0, w) \in \bar{d}_{\mu_n} f(x_n)$.

(ii) For any $x \in U$ and $z \in \bar{d}f(x)$ introduce the function

$$\psi(h, z, x, \nu) = \sup_{(a,v) \in \underline{d}_\nu f(x)+z} (a + \langle v, h \rangle).$$

From Proposition 3 it follows that x^* is an inf-stationary point of the function f on the set A iff $\{0\} \in \arg \min_{h \in A - x^*} \psi(h, z, x, \nu)$ for all $z = (0, w) \in \text{ext } \bar{d}f(x^*)$ or, equivalently, iff $\{0\} \in \arg \min_{h \in K \cap (A - x^*)} \psi(h, z, x, \nu)$ for all $z = (0, w) \in \text{ext } \bar{d}f(x^*)$, where K is a set containing a neighbourhood of zero. Bearing this result in mind, one can propose a different method for minimizing a codifferentiable function on a convex set. Namely, instead of minimizing the function $\varphi_n(h, z)$ one can compute a search direction as follows:

$$\{h_n(z)\} \in \arg \min \left\{ \psi(h, z, x_n, \nu_n) \mid h \in K \cap (A - x_n) \right\}.$$

Here $K \subset \mathcal{H}$ is a bounded closed convex set containing a neighbourhood of zero. It is natural to call the modification of the QR-MCD utilizing these search directions *the primal regularization of the MCD*. Let us note that one can easily extend the results of this section to the case of the primal regularization of the MCD.

5 The rate of convergence: an example

In this section we discuss the rate of convergence of the MCD. Note that if the function f is smooth, then the MCD coincides with the gradient descent. Therefore it is natural to expect that the MCD converges linearly, provided a suitable second order sufficient optimality condition holds true at the limit point. A rigorous analysis of the rate of convergence obviously requires the use of very cumbersome second order approximation of codifferentiable functions (such as the so-called *second order codifferentials*, see [19]), and is complicated by the fact that in every iteration of the method one uses multiple search directions some of which might not be descent directions. Furthermore, the MCD (unlike the optimality conditions) is not invariant with respect to the choice of a codifferential, and one can provide examples in which a poor choice of a codifferential mapping significantly slows down the method. That is why we do not present a detailed analysis of the rate of convergence of the MCD here, and leave it as an open problem for future research. Instead, we consider only a certain class of nonconvex codifferentiable functions for which the rate of convergence can be easily estimated with the use of some existing results.

Let the function f have the form

$$f(x) = \sum_{k=1}^m \max_{i \in I_k} g_{ki}(x) + \sum_{l=1}^s \min_{j \in J_l} u_{lj}(x), \quad (27)$$

where $g_{ki}, u_{lj}: \mathcal{H} \rightarrow \mathbb{R}$ are continuously differentiable functions, $I_k = \{1, \dots, p_k\}$, $J_l = \{1, \dots, q_l\}$. Particular examples of functions of the form (27), including some functions arising in cluster analysis, can be found in [17].

Observe that f is a nonconvex nonsmooth function (even if all functions g_{ki} and u_{lj} are convex). Furthermore, f is continuously codifferentiable, and applying Theorem 1 and Corollary 3 one can define

$$Df(x) = \left[\sum_{k=1}^m \text{co} \{ (g_{ki}(x) - g_k(x), \nabla g_{ki}(x)) \mid i \in I_k \}, \right. \\ \left. \sum_{l=1}^s \text{co} \{ (u_{lj}(x) - u_l(x), \nabla u_{lj}(x)) \mid j \in J_l \} \right],$$

where $g_k(x) = \max_{i \in I_k} g_{ki}(x)$, and $u_l(x) = \min_{j \in J_l} u_{lj}(x)$. Additionally, one can easily check that the codifferential mapping $Df(x)$ locally uniformly approximates the function f , provided the gradients of the functions g_{ki} and u_{lj} are locally uniformly continuous (see Theorem 1).

Let us apply the QR-MCD to the function f . Fix arbitrary $\nu \geq 0$ and $\mu \geq 0$. For any $x \in \mathcal{H}$, introduce the index sets

$$I_{k,\nu}(x) = \{i \in I_k \mid g_{ki}(x) \geq g_k(x) - \nu\}, \quad J_{l,\mu}(x) = \{j \in J_l \mid u_{lj}(x) \leq u_l(x) + \mu\}, \\ I_\nu(x) = I_{1,\nu}(x) \times \dots \times I_{m,\nu}(x), \quad J_\mu(x) = J_{1,\mu}(x) \times \dots \times J_{s,\mu}(x),$$

and denote $I(x) = I_0(x)$ and $J(x) = J_0(x)$. Define also

$$\begin{aligned}\underline{d}_\nu f(x) &= \sum_{k=1}^m \{ (g_{ki}(x) - g_k(x), \nabla g_{ki}(x)) \mid i \in I_{k,\nu}(x) \}, \\ \bar{d}_\mu f(x) &= \sum_{l=1}^s \{ (u_{lj}(x) - u_l(x), \nabla u_{lj}(x)) \mid j \in J_{l,\mu}(x) \}.\end{aligned}$$

It is easy to see that inclusions (12) hold true. Thus, we can minimize the function f with the use of the QR-MCD with the sets $\underline{d}_\nu f(x)$ and $\bar{d}_\mu f(x)$ defined as above (note that these sets are finite).

For any multi-index $\lambda = (j_1, \dots, j_s) \in J_1 \times \dots \times J_s$ introduce the function

$$f_\lambda(x) = \sum_{k=1}^m \max_{i \in I_k} g_{ki}(x) + \sum_{l=1}^s u_{lj_l}(x).$$

By definition one has $f(x) = \min_\lambda f_\lambda(x)$. In essence, the QR-MCD for the function f can be viewed as a method that simultaneously minimizes the functions f_λ , $\lambda \in J_\mu(x)$, with the use of the PPP algorithm. This idea will allow us to estimate the rate of convergence of this method with the use of some existing results on the convergence of the PPP algorithm (see [52], Sections 2.4.1–2.4.4).

Under the assumption that the functions g_{ki} and u_{lj} are strongly convex, we can prove that the QR-MCD converges with a linear rate. In order to utilize the existing results on the convergence of the PPP algorithm, below we suppose that $\nu = +\infty$, and the space \mathcal{H} is finite dimensional. It should be noted that the following theorem remains valid in the general case, but for the sake of shortness we do not present the proof of this result here.

Theorem 7. *Let $\mathcal{H} = \mathbb{R}^d$, $\alpha_* \geq 1$, $\nu_n \equiv +\infty$, and $\lim_{n \rightarrow \infty} \mu_n > 0$. Let also the functions g_{ki} and u_{lj} be twice continuously differentiable. Suppose that there exist $M > m > 0$ such that*

$$m\|y\|^2 \leq \langle y, \nabla^2 g_{ki}(x)y \rangle \leq M\|y\|^2, \quad m\|y\|^2 \leq \langle y, \nabla^2 u_{lj}(x)y \rangle \leq M\|y\|^2,$$

for any $x, y \in \mathbb{R}^d$, and for all $i \in I_k$, $k \in \{1, \dots, m\}$ and $j \in J_l$, $l \in \{1, \dots, s\}$. Suppose, finally, that a sequence $\{x_n\}$ generated by the QR-MCD for the function f converges to a point x^* . Then x^* is a point of strict local minimum of the function f , and there exist $c \in (0, 1)$, $Q > 0$, and $n_0 \in \mathbb{N}$ such that

$$[f(x_{n+1}) - f(x^*)] \leq c[f(x_n) - f(x^*)], \quad \|x_{n+1} - x^*\| \leq Qc^{n/2}, \quad (28)$$

for any $n \geq n_0$, i.e. the QR-MCD converges linearly.

Proof. By Theorem 5 the point x^* is an inf-stationary point of the function f . Therefore, as it is easy to check, $0 \in \partial f_\lambda(x^*)$ for any $\lambda \in J(x^*)$, where $\partial f_\lambda(x^*)$ is the subdifferential of the convex function f_λ at x^* in the sense of convex analysis. Consequently, x^* is a point of global minimum of f_λ for all $\lambda \in J(x^*)$. Observe that every function f_λ is strictly convex as the sum of the maximum of convex functions and a strongly convex function. Hence x^* is a point of *strict* global minimum of f_λ for all $\lambda \in J(x^*)$, which, as it is easily seen, implies that x^* is a point of strict local minimum of the function f .

Let us now estimate the rate of convergence. Since the functions u_{lj} are continuous, there exists a neighbourhood U of x^* such that $J(x) \subseteq J(x^*)$ for all $x \in U$. Hence taking into account the fact that $x_n \rightarrow x^*$ as $n \rightarrow \infty$, one obtains that there exists $n_0 \in \mathbb{N}$ such that $x_n \in U$ for all $n \geq n_0$.

Fix arbitrary $\sigma, \gamma \in (0, 1)$, $n \geq n_0$ and $\lambda = (j_1, \dots, j_s) \in J(x_n)$. Define $z_n = (0, \sum_{l=1}^s \nabla u_{lj_1}(x_n))$. Then $z_n \in \bar{d}_\mu f(x_n)$. Define also

$$\{\gamma_n\} = \arg \max_{k \in \mathbb{N} \cup \{0\}} \left\{ \gamma^k \mid f_\lambda(x_n + \gamma^k h_n(z_n)) - f_\lambda(x_n) \leq \gamma^k \sigma \varphi_n(h_n(z_n), z_n) \right\}.$$

Let us check that γ_n is correctly defined. Indeed, from the definitions of f_λ and $\varphi_n(h, z)$ (see (23)) it follows that $f'_\lambda(x_n, h) = (\varphi_n(\cdot, z_n))'(0, h)$ for all $h \in \mathcal{H}$. Taking into account the convexity of the function $\varphi_n(\cdot, z_n)$ and the fact that $\varphi_n(0, z_n) = 0$ one obtains that $(\varphi_n(\cdot, z_n))'(0, h) \leq \varphi_n(h, z_n)$ for all $h \in \mathcal{H}$, which yields that $f(x_n + \alpha h_n(z_n)) - f(x_n) \leq \alpha \sigma \varphi_n(h_n(z_n), z_n)$ for any sufficiently small α (note that if $\varphi_n(h_n(z_n), z_n) = 0$, then $h_n(z_n) = 0$). Therefore $\gamma_n \in (0, 1]$ is correctly defined.

Observe that $h_n(z_n)$ is the search direction, and γ_n is the step size of the PPP algorithm for the function f_λ (see [52], Algorithm 2.4.1). Consequently, applying [52], Theorem 2.4.5 one gets that there exists $c \in (0, 1)$ depending only on γ, σ, m and M such that

$$[f_\lambda(x_n + \gamma_n h_n(z_n)) - f_\lambda(x^*)] \leq c[f_\lambda(x_n) - f_\lambda(x^*)].$$

Recall that $\alpha_* \geq 1$. Therefore

$$f(x_n + \alpha_n(z_n) h_n(z_n)) \leq f(x_n + \gamma_n h_n(z_n)) \leq f_\lambda(x_n + \gamma_n h_n(z_n)).$$

Hence and from the fact that $\lambda \in J(x_n) \subseteq J(x^*)$ it follows that

$$[f(x_n + \alpha_n(z_n) h_n(z_n)) - f(x^*)] \leq c[f(x_n) - f(x^*)].$$

Consequently, taking into account the fact that $f(x_{n+1}) \leq f(x_n + \alpha_n(z_n) h_n(z_n))$ by definition one gets that the first inequality in (28) holds true.

Applying inequality (2.4.9h) from [52] one obtains that $f_\lambda(x) - f_\lambda(x^*) \geq m \|x - x^*\|^2 / 2$ for all $x \in \mathbb{R}^d$ and $\lambda \in J(x^*)$. Consequently, $f(x) - f(x^*) \geq m \|x - x^*\|^2 / 2$ for any $x \in U$. Combining this inequality with the first inequality in (28) one obtains that the second inequality in (28) is valid. \square

Let us show that if the QR-MCD converges to a point satisfying a certain *first order sufficient optimality condition*, then it converges *quadratically*. To this end, let us recall the definition of *Haar point* [52] (or *Chebyshev point* in the terminology of the paper [13]).

Let $\mathcal{H} = \mathbb{R}^d$. Fix arbitrary $x^* \in \mathbb{R}^d$ and $\lambda = (j_1, \dots, j_s) \in J(x^*)$. Suppose that x^* is an inf-stationary point of the function f_λ , i.e. $0 \in \underline{d}f_\lambda(x^*)$ (note that f_λ is obviously hypodifferentiable). Then there exist $\alpha_\zeta \geq 0$, $\zeta = (i_1, \dots, i_m) \in I := I_1 \times \dots \times I_m$, such that $\alpha_\zeta = 0$ for all $\zeta \notin I(x^*)$, and

$$\sum_{\zeta \in I} \alpha_\zeta \left(\sum_{k=1}^m \nabla g_{ki_k}(x^*) \right) + \sum_{l=1}^s \nabla u_{lj_1}(x^*) = 0, \quad \sum_{\zeta \in I} \alpha_\zeta = 1.$$

Denote the set of all such $\alpha = \{\alpha_\zeta\}_{\zeta \in I}$ by $\alpha_\lambda(x^*)$. The elements of the set $\alpha_\lambda(x^*)$ are sometimes called *Danskin-Demyanov multipliers* [52].

Definition 2. The point x^* is called a *Haar point* of the function f_λ , if

1. $0 \in \underline{d}f_\lambda(x^*)$,
2. $|I(x^*)| = |I_{1,0}(x^*)| \times \dots \times |I_{m,0}(x^*)| = d + 1$, where $|M|$ is the cardinality of a set M ,
3. the vectors $\nabla g_{1i_1}(x^*) + \dots + \nabla g_{mi_m}(x^*)$, $(i_1, \dots, i_m) \in I(x^*)$, are affinely independent,
4. for any $\alpha \in \alpha_\lambda(x^*)$ one has $\alpha_\zeta > 0$ for all $\zeta \in I(x^*)$.

Remark 15. Note that if assumptions 2 and 3 from the definition above are satisfied, then assumptions 1 and 4 hold true iff $0 \in \text{int } \underline{\partial}f_\lambda(x^*)$. Furthermore, it is easy to see that if x^* is a Haar point of f_λ , then the set $\alpha_\lambda(x^*)$ consists of only one element.

Let x^* be a Haar point of f_λ . Then by [52], Theorem 2.4.22 (see also [13]) for any starting point x_0 sufficiently close to x^* the PPP algorithm for the function f_λ coincides with the Newton method for solving a certain system of nonlinear equations, and converges quadratically to x^* . Thus, there exists a neighbourhood U of x^* such that for any starting point $x_0 \in U$ the sequence $\{x_n\}$ generated by the PPP algorithm for the function f_λ stays in U and converges quadratically to x^* .

Theorem 8. Let $\mathcal{H} = \mathbb{R}^d$, $\alpha_* \geq 1$, $\nu_n \equiv +\infty$, and $\lim_{n \rightarrow \infty} \mu_n > 0$. Let also the gradients of the functions g_{ki} and u_{ij} be locally Lipschitz continuous. Suppose, finally, that a sequence $\{x_n\}$ generated by the QR-MCD for the function f converges to a point x^* such that for any $\lambda \in J(x^*)$ the point x^* is a Haar point of the function f_λ . Then x^* is a point of strict local minimum of the function f , and $x_n \rightarrow x^*$ as $n \rightarrow \infty$ quadratically.

Proof. Note that by Theorem 5 the point x^* is an inf-stationary point of the function f , which implies that $0 \in \underline{d}f_\lambda(x^*)$ for all $\lambda \in J(x^*)$. Applying [52], Lemma 2.4.19 (see also inequality (2.4.24g)), and the fact that x^* is a Haar point of every function f_λ , $\lambda \in J(x^*)$, one obtains that there exist a neighbourhood U_0 of x^* and $\varkappa > 0$ such that

$$f_\lambda(x) - f_\lambda(x^*) \geq \varkappa \|x - x^*\| \quad \forall x \in U_0 \quad \forall \lambda \in J(x^*). \quad (29)$$

Hence, as it is easily seen, there exists a neighbourhood $U \subseteq U_0$ of x^* such that $f(x) - f(x^*) \geq \varkappa \|x - x^*\|$ for all $x \in U$, i.e. x^* is a point of strict local minimum of the function f .

From the continuity of the functions u_{ij} it follows that there exists a neighbourhood $V \subseteq U$ of x^* such that $J(x) \subseteq J(x^*)$ for all $x \in V$. Taking into account the fact that $x_n \rightarrow x^*$ as $n \rightarrow \infty$, one obtains that there exists $n_0 \in \mathbb{N}$ such that $x_n \in V$ for all $n \geq n_0$.

Fix arbitrary $\sigma, \gamma \in (0, 1)$, $n \geq n_0$ and $\lambda = (j_1, \dots, j_s) \in J(x_n)$. Define $z_n = (0, \sum_{l=1}^s \nabla u_{lj_1}(x_n))$, and let γ_n be defined as in the proof of Theorem 7. Then γ_n is the step size in the PPP algorithm for the function f_λ . Note that one can choose n_0 so large that x_n belongs to the region of quadratic convergence of the PPP algorithm for all function f_λ , $\lambda \in J(x^*)$, and a sequence generated by this algorithms for any function f_λ , $\lambda \in J(x^*)$ with the starting point x_n stays

in V . Therefore there exists $Q > 0$ (independent of $n \geq n_0$ and $\lambda \in J(x^*)$) such that

$$\|x_n + \gamma_n h_n(z_n) - x^*\| \leq Q \|x_n - x^*\|^2, \quad x_n + \gamma_n h_n(z_n) \in V.$$

By Corollary 2 the function f is Lipschitz continuous on any bounded set. Hence and from (29) it follows that there exists $L > 0$ such that

$$\begin{aligned} f(x_n + \gamma_n h_n(z_n)) - f(x^*) &\leq L \|x_n + \gamma_n h_n(z_n) - x^*\| \\ &\leq LQ \|x_n - x^*\|^2 \leq \frac{LQ}{\varkappa^2} [f(x_n) - f(x^*)]^2 \end{aligned}$$

Recall that $\alpha_* \geq 1$. Hence $f(x_{n+1}) \leq f(x_n + \alpha_n(z_n)h_n(z_n)) \leq f(x_n + \gamma_n h_n(z_n))$. Consequently, one has

$$\varkappa \|x_{n+1} - x^*\| \leq f(x_{n+1}) - f(x^*) \leq \frac{LQ}{\varkappa^2} [f(x_n) - f(x^*)]^2 \leq \frac{L^3 Q}{\varkappa^2} \|x_n - x^*\|^2,$$

i.e. $x_n \rightarrow x^*$ as $n \rightarrow \infty$ quadratically. \square

6 Conclusion

In this paper we studied two methods for minimizing a codifferentiable function defined on a Hilbert space. Namely, we proposed and analysed a generalization of the method of codifferential descent and a quadratic regularization of the method of codifferential descent. In order to study a convergence of these methods, we introduced a class of uniformly codifferentiable functions, and derived some calculus rules that allow one to verify whether a given nonsmooth function is uniformly codifferentiable, and compute the corresponding codifferential mapping. Under some natural assumptions we proved the global convergence of the methods, as well as the convergence of the inf-stationarity measure $\omega(x_n, \nu_n)$ to zero. It should be noted that the quadratic regularization of the MCD (as well as the primal regularization of this method) is the first method for minimizing a codifferentiable function over a convex set. In the end of the paper we estimated the rate of convergence of the MCD for a class of nonsmooth nonconvex functions.

7 Acknowledgements

The author is sincerely grateful to his colleagues T. Angelov, A. Fominyh and G.Sh. Tamasyan for numerous useful discussions on the method of codifferential descent and its applications that played a significant role in the preparation of this paper. Also, the author wishes to express his thanks to professor V.N. Malozemov for pointing out the fact that the PPP algorithm converges quadratically to a Chebyshev (Haar) point.

References

- [1] M. Y. Andramonov. The method of confidence neighbourhoods for minimization of codifferentiable functions. *Russ. Math.*, 48:1–7, 2004.

- [2] A. Bagirov. A method for minimization of quasidifferentiable functions. *Optim. Methods Softw.*, 17:31–60, 2002.
- [3] A. Bagirov, N. Kar Mitsa, and M. M. Mäkelä. *Introduction to Nonsmooth Optimization*. Springer International Publishing, Cham, 2014.
- [4] A. M. Bagirov, A. N. Ganjehlou, J. Ugon, and A. H. Tor. Truncated codifferential method for nonsmooth convex optimization. *Pac. J. Optim.*, 6:483–496, 2010.
- [5] A. M. Bagirov, B. Karasözen, and M. Sezer. Discrete gradient method: derivative-free method for nonsmooth optimization. *J. Optim. Theory Appl.*, 137:317–334, 2008.
- [6] A. M. Bagirov, A. M. Rubinov, and J. Zhang. A multidimensional descent method for global optimization. *Optim.*, 58:611–625, 2009.
- [7] A. M. Bagirov and J. Ugon. Codifferential method for minimizing nonsmooth dc functions. *J. Glob. Optim.*, 50:3–22, 2011.
- [8] A. M. Bagirov. Numerical methods for minimizing quasidifferentiable functions: a survey and comparison. In V. Demyanov and A. Rubinov, editors, *Quasidifferentiability and Related Topics*, pages 33–71. Kluwer Academic Publishers, Dordrecht, 2000.
- [9] J. V. Burke, A. S. Lewis, and M. L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J. Optim.*, 15:751–779, 2005.
- [10] F. H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley–Interscience, New York, 1983.
- [11] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia, 2009.
- [12] F. E. Curtis and X. Que. An adaptive gradient sampling algorithm for non-smooth optimization. *Optim. Methods Softw.*, 28:1302–1324, 2013.
- [13] V. A. Daugavet and V. N. Malozemov. Quadratic rate of convergence of a linearization method for solving discrete minimax problems. *USSR Comput. Math. Math. Phys.*, 21:19–28, 1981.
- [14] 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.
- [15] V. F. Demyanov. On codifferentiable functions. *Vestn. Leningr. Univ., Math.*, 2:22–26, 1988.
- [16] V. F. Demyanov. Smoothness of nonsmooth functions. In F. H. Clarke, V. F. Demyanov, and F. Giannessi, editors, *Nonsmooth Optimization and Related Topics*, pages 79–88. Springer, Boston, 1989.

- [17] 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.
- [18] V. F. Demyanov and V. N. Malozemov. *Introduction to Minimax*. Wiley and Sons, New York, 1990.
- [19] V. F. Demyanov and A. M. Rubinov. *Constructive Nonsmooth Analysis*. Peter Lang, Frankfurt am Main, 1995.
- [20] V. F. Demyanov and A. M. Rubinov, editors. *Quasidifferentiability and Related Topics*. Kluwer Academic Publishers, Dordrecht, 2000.
- [21] V. F. Demyanov, G. E. Stavroulakis, L. N. Polyakova, and P. D. Panagiotopoulos. *Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economics*. Kluwer Academic Publishers, Dordrecht, 1996.
- [22] V. F. Demyanov and G. S. Tamasyan. Exact penalty functions in isoperimetric problems. *Optim.*, 60:153–177, 2011.
- [23] V. F. Demyanov and G. S. Tamasyan. Direct methods in the parametric moving boundary variational problem. *Numer. Funct. Anal. Optim.*, 35:932–961, 2014.
- [24] G. Di Pillo and F. Facchinei. Exact barrier function methods for lipschitz programs. *Appl. Math. Optim.*, 32:1–31, 1995.
- [25] M. V. Dolgopolik. Codifferential calculus in normed spaces. *J. Math. Sci.*, 173:441–462, 2011.
- [26] M. V. Dolgopolik. Inhomogeneous convex approximations of nonsmooth functions. *Russ. Math.*, 56:28–42, 2012.
- [27] M. V. Dolgopolik. Nonsmooth problems of calculus of variations via codifferentiation. *ESAIM: Control, Optim. Calc. Var.*, 20:1153–1180, 2014.
- [28] M. V. Dolgopolik. Abstract convex approximations of nonsmooth functions. *Optim.*, 64:1439–1469, 2015.
- [29] R. E. Edwards. *Functional Analysis: Theory and Applications*. Dover Publications, New York, 1995.
- [30] A. V. Fominyh, V. V. Karelin, and L. N. Polyakova. Application of the hypodifferential descent method to the problem of constructing an optimal control. *Optim. Lett.*, 2017. doi: 10.1007/s11590-017-1222-x.
- [31] A. Fuduli, M. Gaudioso, and E. A. Nurminski. A splitting bundle approach for non-smooth non-convex minimization. *Optim.*, 64:1131–1151, 2015.
- [32] N. Haarala, K. Miettinen, and M. Mäkelä. Globally convergent limited memory bundle method for large-scale nonsmooth optimization. *Math. Program.*, 109:181–205, 2007.

- [33] W. Hare and J. Nutini. A derivative-free approximate gradient sampling algorithm for finite minimax problems. *Comput. Optim. Appl.*, 56:1–38, 2013.
- [34] W. Hare and C. Sagastizábal. A redistributed proximal bundle method for nonconvex optimization. *SIAM J. Optim.*, 20:2442–2473, 2010.
- [35] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, Berlin, Heidelberg, 1993.
- [36] A. D. Ioffe and V. M. Tihomirov. *Theory of Extremal Problems*. North-Holland Publishing Company, Amsterdam etc., 1979.
- [37] N. Karmitsa, A. Bagirov, and M. M. Mäkelä. Comparing different non-smooth minimization methods and software. *Optim. Methods Softw.*, 27:131–153, 2012.
- [38] N. Keskar and A. Wächter. A limited-memory quasi-newton algorithm for bound-constrained non-smooth optimization. *Optim. Methods Softw.*, 2017. doi: 10.1080/10556788.2017.1378652.
- [39] K. C. Kiwiel. A nonderivative version of the gradient sampling algorithm for nonsmooth nonconvex optimization. *SIAM J. Optim.*, 20:1983–1994, 2010.
- [40] L. Kuntz. A characterization of continuously codifferentiable function and some consequences. *Optim.*, 22:539–547, 1991.
- [41] A. S. Lewis and M. L. Overton. Nonsmooth optimization via quasi-newton methods. *Math. Program.*, 141:135–163, 2013.
- [42] B. Luderer. Does the special choice of quasidifferentials influence necessary minimum conditions? In W. Oettli and D. Pallaschke, editors, *Advances in Optimization*, pages 256–266. Springer-Verlag, Berlin, Heidelberg, 1992.
- [43] B. Luderer, R. Rösiger, and U. Würker. On necessary minimum conditions in quasidifferential calculus: independence of the specific choice of quasidifferentials. *Optim.*, 22:643–660, 1991.
- [44] B. Luderer and J. Weigelt. A solution method for a special class of nondifferentiable unconstrained optimization problems. *Comput. Optim. Appl.*, 24:83–93, 2003.
- [45] M. Mäkelä. Survey of bundle method for nonsmooth optimization. *Optim. Methods Softw.*, 17:1–29, 2002.
- [46] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation I. Basic Theory*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [47] D. Pallaschke, P. Recht, and R. Urbański. On locally-lipschitz quasidifferentiable functions in banach spaces. *Optim.*, 17:287–295, 1986.
- [48] D. Pallaschke and R. Urbański. *Pairs of Compact Convex Sets: Fractional Arithmetic with Convex Sets*. Kluwer Academic Publishers, Dordrecht, 2002.

- [49] J.-P. Penot. On the convergence of descent algorithms. *Comput. Optim. Appl.*, 23:279–284, 2002.
- [50] J.-P. Penot. *Calculus Without Derivatives*. Springer Science+Business Media, New York, 2013.
- [51] O. Pironneau and E. Polak. On the rate of convergence of certain methods of centers. *Math. Program.*, 2:230–258, 1972.
- [52] E. Polak. *Optimization. Algorithms and Consistent Approximations*. Springer-Verlag, New York, 1997.
- [53] B. N. Pschenichny and Y. M. Danilin. *Numerical Methods in Extremal Problems. English ed.* Mir Publishers, Moscow, 1978.
- [54] A. H. Tor, A. Bagirov, and B. Karasözen. Aggregate codifferential method for nonsmooth dc optimization. *J. Comput. Appl. Math.*, 259:851–867, 2014.
- [55] A. H. Tor, B. Karasözen, and A. Bagirov. Truncated codifferential method for linearly constrained nonsmooth optimization. In *the ISI Proc. of 24th Mini EURO Conference “Continuous Optimization and Information-Based Technologies in the Financial Sector”*, Izmir, Turkey, pages 87–93, 2010.
- [56] A. Uderzo. Fréchet quasidifferential calculus with applications to metric regularity of continuous maps. *Optim.*, 54:469–493, 2005.
- [57] A. Zaffaroni. Continuous approximations, codifferentiable functions and minimization methods. In V. F. Demyanov and A. M. Rubinov, editors, *Quasidifferentiability and related Topics*, pages 361–391. Kluwer Academic Publishers, Dordrecht, 2000.