Journal of Convex Analysis
Volume 32 (2025), No. 4, 1255–1298
https://doi.org/10.68381/jca32066

Hypodifferentials of Nonsmooth Convex Functions and their Applications to Nonsmooth Convex Optimization

Maksim V. Dolgopolik
Institute for Problems in Mechanical Engineering of the Russian Academy of Sciences, Saint Petersburg, Russia
dmv@ipme.ru

Received: March 23, 2023
Accepted: July 20, 2024

A hypodifferential is a compact family of affine mappings that defines a local max-type approximation of a nonsmooth convex function. We present a general theory of hypodifferentials of nonsmooth convex functions defined on a Banach space. In particular, we provide complete characterizations of hypodifferentiability and hypodifferentials of nonsmooth convex functions, derive calculus rules for hypodifferentials, and study the Lipschitz continuity/Lipschitz approximation property of hypodifferentials that can be viewed as a natural extension of the Lipschitz continuity of the gradient to the general nonsmooth setting. As an application of our theoretical results we study the rate of convergence of several versions of the method of hypodifferential descent for nonsmooth convex optimization and present an accelerated version of this method having the faster rate of convergence $\mathcal{O}(1/k^2)$.

Keywords: Hypodifferential, subdifferential, convex analysis, convex optimization, nonsmooth optimization, first-order method.

2020 Mathematics Subject Classification: 90C25, 90C56, 68Q25.

1. Introduction

Although the idea of approximating objective function with the use of a family of affine mappings has been successfully used in the context of minimax optimization problems (see, e.g. [43, Section 2.4] and [37, Sections 2.3.1–2.3.3]), in the general nonsmooth case only linear approximations are typically used. For example, a subdifferential is always defined as a collection of some sort of local or global *linear* approximations/minorants of a nonsmooth function under consideration [42, 50].

In the late 1980s, Demyanov [7, 8, 9, 10] extended the idea of approximating a nonsmooth function with the use of families of affine mappings to the case of general nonsmooth functions by introducing the concept of *codifferential*. Since then the theory of codifferentials has found a number of applications in optimization, calculus of variations, nonsmooth mechanics, etc. (see [11, 12, 16, 20] and the references therein).

This work was performed in IPME RAS and supported by the Russian Science Foundation (Grant No. 20-71-10032).

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

In the convex case (and some other cases, e.g. in the cases of minimax and composite problems), codifferentials are reduced to the so-called *hypodifferentials*, which can be defined as a compact family of affine mappings that provides a local max-type approximation of a nonsmooth function. A numerical method based on the use of hypodifferentials, called *the method of hypodifferential descent* [10], has been applied to various problems of the calculus of variations [13, 14] and optimal control problems [26, 27]. In [18], it was shown that the method of hypodifferential descent with Armijo's step size rule has the rate of convergence $\mathcal{O}(1/k)$ for general nonsmooth convex functions, which exceeds the standard rate of convergence $\mathcal{O}(1/\sqrt{k})$ of subgradient methods [37], although at the cost of significantly increased complexity of each iteration. The fact that the method of hypodifferential descent has the same rate of convergence in the nonsmooth case as gradient descent methods in the smooth case hints at the possibility of devising accelerated versions of this method that are similar to accelerated gradient methods.

An accelerated version of the gradient descent method having the $\mathcal{O}(1/k^2)$ rate of convergence for smooth convex functions with globally Lipschitz continuous gradient has been first developed by Nesterov [36] (see also [37]). Various improved versions of accelerated gradient descent methods have been studied, e.g. in [33, 34], while a constructive technique for devising accelerated methods for smooth and nonsmooth convex minimization was proposed in [22]. A general approach towards design and analysis of accelerated gradient methods based on analysis of auxiliary ordinary differential equations and differential equations solvers was recently developed in [35, 46].

As is well-known (see, e.g. [37]), in the general nonsmooth case the rate of convergence of convex minimization methods is significantly slower than in the smooth case and is equal to $\mathcal{O}(1/\sqrt{k})$. For a fixed number of iterations, this rate of convergence is achieved by the celebrated Shor's subgradient method [47], while when the number of iterations is not fixed it is achieved by many existing methods, such as Nesterov's primal-dual subgradient methods [40], OSGA [41], a Kelley-like method from [23] (see also restarted subgradient method [54]), etc. A significantly faster rate of convergence can be achieved in the nonsmooth case with the use of some additional assumptions on the objective function, such as Hölderian growth [32], the validity of error bounds and the Kurdyka-Łojasiewicz inequality [3], function growth condition measures [28], etc. A faster rate of convergence can also be achieved by exploiting a particular structure of a problem class under consideration, as is done, e.g. in the context of Nesterov's smoothing [39] and excessive gap [38] techniques and in the setting of composite convex optimization [30, 48, 49, 53].

The main goal of this paper is twofold. Firstly, our aim is to develop a general theory of hypodifferentials of nonsmooth convex functions defined on a Banach space. We provide complete characterizations of hypodifferentiability and hypodifferentials of such functions in terms of subdifferentiability and subdifferentials in the sense of convex analysis. We also present a detailed hypodifferential calculus and study the Lipschitz continuity/Lipschitz approximation property of hypodifferential. In contrast to subdifferentials, hypodifferentials of many nonsmooth convex functions arising in applications are continuous and often even Lipschitz continuous. Thus, the Lipschitz continuity of a hypodifferential can be viewed as a natural extension

of the property of the Lipschitz continuity of the gradient, which is widely used in smooth convex optimization, to the nonsmooth setting. We show that standard operations on convex functions (including the pointwise maximum) preserve the Lipschitz continuity of hypodifferential and, moreover, provide a general existence theorem for Lipschitz continuous hypodifferentials.

Secondly, as an application of our theoretical results, we aim at analysing the rate of convergence of various versions of the method of hypodifferential descent in the nonsmooth convex case and presenting an accelerated version of this method having a faster rate of convergence. Namely, we prove that the method of hypodifferential descent with constant step size has the rate of convergence $\mathcal{O}(1/k)$ and show that under some additional assumptions a specific choice of step size rule can ensure the linear convergence of the method. Following the standard approach to accelerated methods developed by Nesterov [37], we also present an accelerated version of the method of hypodifferential descent for a class of nonsmooth convex functions that can be viewed as a natural extension of the class of smooth convex functions having a globally Lipschitz continuous gradient to the nonsmooth case. We prove that the accelerated method has the rate of convergence $\mathcal{O}(1/k^2)$ that coincides with the rate of convergence of accelerated gradient methods. It must be pointed out that such rate of convergence is achieved by significantly increasing the complexity of each iteration and each oracle call in comparison with classical subgradient/bundle methods, which make the accelerated hypodifferential descent method inapplicable to large scale problems. Nonetheless, this method can be viewed as a method that bridges the gap between nonsmooth and smooth convex optimization methods, at least from the purely theoretical point of view.

Let us finally note that this paper is a continuation of the research started in [18]. Some of the results presented below are completely new (e.g. the characterizations of hypodifferentiability and hypodifferentials given in Theorem 3.4, some of the calculus rules and results on Lipschitz property, the accelerated version of the method of hypodifferential descent, etc.), while others are significantly revised and improved versions of the author's earlier results from [17, 18].

The paper is organized as follows. Some necessary auxiliary results on the Pompeiu-Hausdorff distance are collected in Section 2. The hypodifferential calculus for nonsmooth convex functions is developed in Section 3. Subsection 3.1 contains main definitions connected to hypodifferentials, as well as the theorem providing convenient characterizations of hypodifferentiability and hypodifferentials. Calculus rules for hypodifferentials are presented in Subsection 3.2, while several particular examples are discussed in Subsection 3.3. Subsection 3.4 is devoted to the analysis of Lipschitz continuity/Lipschitz approximation property of hypodifferentials. Some applications of our theoretical results to nonsmooth convex optimization are presented in Section 4. The rate of convergence of the method of hypodifferential descent with two different step size rules is estimated in Subsection 4.1, while an accelerated version of this method having a faster rate of convergence is studied in Subsection 4.2.

2. Preliminaries

Let us present some auxiliary results on the Pompeiu-Hausdorff metric and the continuity of set-valued maps with respect to this metric that will be utilized throughout the article. It should be noted that all results presented in this section are simple and well-known (see [50, 45]), but scattered in the literature and often presented either without proofs or only in the finite dimensional/compact-valued case. Therefore, for the sake of completeness we provide short proofs of all these results.

Let (M, d) be a metric space. Recall that the *Pompeiu-Hausdorff distance* between nonempty subsets $A, B \subseteq M$ is defined as

$$d_{PH}(A, B) = \sup\Big\{ \sup_{a \in A} \operatorname{dist}(a, B),\ \sup_{b \in B} \operatorname{dist}(b, A) \Big\},$$

where $\operatorname{dist}(a, B) = \inf_{b \in B} d(a, b)$ is the distance between a point $a \in M$ and a set $B \subseteq M$. The following simple result follows directly from the definition of the Pompeiu-Hausdorff distance.

Proposition 2.1. *Let $A, B \subseteq M$ be nonempty sets and $\varepsilon > 0$ be given. Then $d_{PH}(A, B) < \varepsilon$ if and only if there exists $\eta \in (0, \varepsilon)$ for which the following two conditions hold true:*

(1) *for any $a \in A$ there exists $b \in B$ such that $d(a, b) \leq \eta$;*

(2) *for any $b \in B$ there exists $a \in A$ such that $d(a, b) \leq \eta$.*

Let (X, d_X) be a metric space. Recall that a set-valued map $F\colon M \rightrightarrows X$ is called *continuous* in the Pompeiu-Hausdorff metric at a point $x \in M$, if for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d_{PH}(F(x'), F(x)) < \varepsilon \quad \forall x' \in B(x, \delta) := \big\{ y \in M \bigm| d(x, y) \leq \delta \big\}.$$

The map F is called continuous in the Pompeiu-Hausdorff metric, if it is continuous at every point $x \in M$. Note also that if one replaces $B(x, \delta)$ with a neighbourhood of x, then the definition above can be applied to the case when M is merely a topological space.

Remark 2.2. Below we will often call set-valued maps that are continuous in the Pompeiu-Hausdorff metric simply continuous, since we will not consider any other notion of continuity for set-valued mappings in this paper. □

Let us show that many standard operations on set-valued maps preserve continuity in the Pompeiu-Hausdorff metric.

Proposition 2.3. *Let X be a normed space and $F, G\colon M \rightrightarrows X$ be continuous set-valued maps. Then the Minkowski sum $H = F + G$ is continuous as well.*

Proof. Fix any $x \in M$ and $\varepsilon > 0$. Since F and G are continuous, one can find $\delta > 0$ such that

$$d_{PH}(F(x'), F(x)) < \frac{\varepsilon}{2}, \quad d_{PH}(G(x'), G(x)) < \frac{\varepsilon}{2} \quad \forall x' \in B(x, \delta).$$

Choose any $x' \in B(x, \delta)$ and $y \in H(x')$. By definition there exist $y_1 \in F(x')$ and $y_2 \in G(x')$ such that $y = y_1 + y_2$. By Proposition 2.1 there exist $\eta < \varepsilon/2$, $z_1 \in F(x)$,

and $z_2 \in G(x)$ such that $\|y_i - z_i\| \leq \eta$, $i \in \{1,2\}$. Define $z = z_1 + z_2$. Clearly, $z \in H(x)$ and $\|y - z\| \leq 2\eta < \varepsilon$. Thus, for any $y \in H(x')$ we have found $z \in H(x)$ such that $\|y - z\| \leq 2\eta < \varepsilon$. Swapping x and x' we get that for any $z \in H(x)$ one can find $y \in H(x')$ such that $\|y - z\| \leq 2\eta < \varepsilon$. Therefore, by Proposition 2.1 one has $d_{PH}(H(x'), H(x)) < \varepsilon$, which implies that H is continuous. □

Proposition 2.4. *Let X be a normed space, $F\colon M \rightrightarrows X$ be a continuous set-valued map, and $g\colon M \to \mathbb{R}$ be a continuous function. Suppose also that $F(x)$ is a bounded set for any $x \in X$. Then the set-valued map $H(\cdot) = g(\cdot)F(\cdot)$ is continuous.*

Proof. Fix any $x \in M$ and $\varepsilon > 0$. From the continuity of g it follows that there exist $\theta > 0$ and $\delta_1 > 0$ such that $|g(x')| \leq \theta$ for any $x' \in B(x, \delta_1)$. In turn, from the continuity of F it follows that there exists $\delta_2 > 0$ such that $d_{PH}(F(x'), F(x)) < \varepsilon/2\theta$ for any $x' \in B(x, \delta_2)$. Therefore by Proposition 2.1 there exists $\eta < \varepsilon$ such that for any $x' \in B(x, \delta_2)$ and $y' \in F(x')$ one can find $y \in F(x)$ for which $\|y' - y\| \leq \eta/2\theta$.

Finally, applying the continuity of g again one gets that there exists $\delta_3 > 0$ such that $|g(x') - g(x)| < \eta/2C$ for any $x' \in B(x, \delta_3)$, where $C = \sup\{\|y\| \mid y \in F(x)\}$.

Set $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Choose any $x' \in B(x, \delta)$ and $z' \in H(x')$. Then $z' = g(x')y'$ for some $y' \in F(x')$. As was noted above, one can find $y \in F(x)$ such that $\|y' - y\| \leq \eta/2\theta$. Define $z = g(x)y$. Then $z \in H(x)$ and

$$\|z' - z\| = \|g(x')y' - g(x)y\| \leq |g(x')|\|y' - y\| + |g(x') - g(x)|\|y\| \leq \theta\frac{\eta}{2\theta} + \frac{\eta}{2C}C \leq \eta.$$

Thus, for any $z' \in H(x')$ we have found $z \in H(x)$ such that $\|z' - z\| \leq \eta$. Arguing in a similar way one can check that for any $z \in H(x)$ there exists $z' \in H(x')$ satisfying the same inequality. Hence by Proposition 2.1 one has $d_{PH}(H(x'), H(x)) < \varepsilon$, which implies the required result. □

Proposition 2.5. *Let T be a nonempty set and $F : M \times T \rightrightarrows X$ a given set-valued map. Suppose that the maps $F(\cdot, t)$ are continuous uniformly in t. Then the set-valued map $G(\cdot) = \bigcup_{t \in T} F(\cdot, t)$ is continuous. In particular, if T is a compact topological space, then it is sufficient to assume that the set-valued map F is continuous.*

Proof. Fix any $x \in M$ and $\varepsilon > 0$. From the uniform in t continuity of $F(\cdot, t)$ it follows that one can find $\delta > 0$ such that $d_{PH}(F(x', t), F(x, t)) < \varepsilon/2$ for any $x' \in B(x, \delta)$ and $t \in T$.

Choose any $x' \in B(x, \delta)$ and $z' \in G(x')$. Then by definition there exists $t' \in T$ such that $y' \in F(x', t')$. By Proposition 2.1 one can find $y \in F(x, t') \subseteq G(x)$ such that $\|y' - y\| < \varepsilon/2$. Similarly, if $y \in G(x)$, then $y \in F(x, t)$ for some $t \in T$ and Proposition 2.1 implies that there exists some $y' \in F(x', t) \subseteq G(x')$ such that $\|y' - y\| < \varepsilon/2$. Hence applying Proposition 2.1 one more time one obtains that $d_{PH}(G(x'), G(x)) < \varepsilon$, which implies that G is continuous.

It remains to note that if T is a compact topological space and the set-valued map F is continuous, then, as one can readily check using the standard compactness argument, the maps $F(\cdot, t)$ are continuous uniformly in t. □

Proposition 2.6. *Let X be a normed space and $F\colon M \rightrightarrows X$ be a continuous set-valued map. Then the set-valued map* co F *is continuous as well, where "*co*" stands for the convex hull.*

Proof. Fix any $\varepsilon > 0$ and $x \in M$. Due to the continuity of F there exists $\delta > 0$ such that $d_{PH}(F(x'), F(x)) < \varepsilon$ for any $x' \in B(x, \delta)$. Choose any $x' \in B(x, \delta)$ and $y' \in \operatorname{co} F(x')$. By the definition of convex hull there exist $n \in \mathbb{N}$, $y'_i \in F(x')$, and $\alpha_i \geq 0$, $i \in \{1, \ldots, n\}$, such that

$$y' = \sum_{i=1}^{n} \alpha_i y'_i, \quad \sum_{i=1}^{n} \alpha_i = 1.$$

By Proposition 2.1 there exists $\eta < \varepsilon$ such that for any $i \in \{1, \ldots, n\}$ one can find $y_i \in F(x)$ for which $\|y'_i - y_i\| \leq \eta$. Define $y = \sum_{i=1}^{n} \alpha_i y_i$. Then $y \in \operatorname{co} F(x)$ and $\|y' - y\| \leq \eta$. Thus, for any $y' \in F(x')$ we found $y \in F(x)$ such that $\|y' - y\| \leq \eta$.

Repeating the same argument with y' and y swapped and then applying Proposition 2.1 one obtains that $d_{PH}(\operatorname{co} F(x'), \operatorname{co} F(x)) < \varepsilon$, which implies that the set-valued map $\operatorname{co} F$ is continuous. □

Proposition 2.7. *Let $F\colon M \rightrightarrows X$ be a continuous set-valued map. Then the map* $\operatorname{cl} F$ *is continuous as well, where* cl *is the closure operator.*

Proof. The validity of this proposition follows from the triangle inequality for the Pompeiu-Hausdorff metric and the obvious equality $d_{PH}(A, \operatorname{cl} A) = 0$ that holds true for any nonempty set $A \subset X$. □

Let X be a real Banach space. The topological dual space of X is denoted by X^*, and the canonical duality pairing between X and X^* is denoted by $\langle \cdot, \cdot \rangle$, that is, $\langle x^*, x \rangle = x^*(x)$ for any $x \in X$ and $x^* \in X^*$. The weak topology on X is denoted by $\sigma(X, X^*)$, while the weak* topology on X^* is denoted by $\sigma(X^*, X)$.

We endow the Cartesian product $X \times Y$ of normed spaces with the norm $\|(x, y)\| = \|x\|_X + \|y\|_Y$ for any $(x, y) \in X \times Y$. As is easily seen, the spaces $(\mathbb{R} \times X)^*$ and $\mathbb{R} \times X^*$ are isomorphic and the corresponding isomorphism is continuous both in the weak and weak* topologies. In addition, the weak* topology on $(\mathbb{R} \times X)^*$ corresponds to the the product topology $\tau_{\mathbb{R}} \times \sigma(X^*, X)$ on $\mathbb{R} \times X^*$, where $\tau_{\mathbb{R}}$ is the standard topology on $\mathbb{R}$. For the sake of shortness we call this product topology weak*. Let us also note that subsets of $\mathbb{R} \times X^*$ are weak* compact if and only if they are norm bounded and weak* closed (see, e.g. [15, Theorem 2.1]). We denote the closure in the weak* topology by cl^*.

Proposition 2.8. *Let X be a reflexive Banach space, $Q \subseteq X$ be a nonempty set, and $F\colon Q \rightrightarrows \mathbb{R} \times X^*$ be a set-valued map with bounded convex values. If F is continuous, then* $\operatorname{cl}^* F$ *is continuous as well.*

Proof. One can readily check that the reflexivity of X implies that the space $\mathbb{R} \times X$ is reflexive as well. Hence taking into account the fact that for convex sets the closure in the strong topology coincides with the closure in the weak topology one can conclude that for any convex subset $K \subset (\mathbb{R} \times X)^*$ its closure in the strong topology $\operatorname{cl} K$ coincides with $\operatorname{cl}^* K$. Therefore $\operatorname{cl} F(x) = \operatorname{cl}^* F(x)$ for any $x \in Q$, which by Proposition 2.7 implies that $\operatorname{cl}^* F$ is continuous □

3. Hypodifferential calculus for nonsmooth convex functions

In this section we present a detailed theory of *hypodifferentials* of nonsmooth convex functions defined on a real Banach space. Our aim is to develop a calculus of such approximations and show that hypodifferentials possess many useful properties in the general nonsmooth case (such as continuity) that subdifferentials do not have.

3.1. Hypodifferentials of convex functions

Let X be a real Banach space and $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a closed convex function. Denote by $\operatorname{dom} f = \{x \in X \mid f(x) \in \mathbb{R}\}$ the effective domain of f.

Definition 3.1. The function f is called *hypodifferentiable* at a point $x \in \operatorname{dom} f$, if there exists a weak* compact set $\underline{d}f(x) \subset \mathbb{R} \times X^*$ such that

$$\lim_{\alpha \to +0} \frac{1}{\alpha} \Big| f(x + \alpha \Delta x) - f(x) - \max_{(a,x^*) \in \underline{d}f(x)} (a + \alpha \langle x^*, \Delta x \rangle) \Big| = 0, \quad \max_{(a,x^*) \in \underline{d}f(x)} a = 0 \tag{1}$$

for any $\Delta x \in X$. The set $\underline{d}f(x)$ is called a *hypodifferential* of f at x.

Remark 3.2. It should be noted the maximum in $\max_{(a,x^*) \in \underline{d}f(x)} (a + \langle x^*, y \rangle)$ is attained for any $y \in X$, since the set $\underline{d}f(x) \subset \mathbb{R} \times X^*$ is weak* compact. □

Remark 3.3. Let us point out how the definition of hypodifferential transforms under various assumptions on the space X:

(1) If X is a reflexive Banach space, then a hypodifferential of f at $x \in \operatorname{dom} f$ is a norm closed and bounded subset $\underline{d}f(x) \subseteq \mathbb{R} \times X^*$ satisfying equalities (1).

(2) If X is a Hilbert space, then taking into account the Riesz representation theorem it is natural to define a hypodifferential of f as a weakly compact (or, equivalently, norm closed and bounded) set $\underline{d}f(x) \subset \mathbb{R} \times X$ such that for any $\Delta x \in X$ one has

$$\lim_{\alpha \to +0} \frac{1}{\alpha} \Big| f(x + \alpha \Delta x) - f(x) - \max_{(a,v) \in \underline{d}f(x)} (a + \langle v, x \rangle) \Big| = 0, \quad \max_{(a,v) \in \underline{d}f(x)} a = 0, \tag{2}$$

where $\langle \cdot, \cdot \rangle$ is the inner product in H.

(3) if $X = \mathbb{R}^d$, then a hypodifferential of f at $x \in \operatorname{dom} f$ is a compact subset $\underline{d}f(x) \subset \mathbb{R}^{d+1}$ satisfying equalities (2) with $\langle \cdot, \cdot \rangle$ being the inner product in $\mathbb{R}^d$. □

Unlike derivative/gradient in the smooth case and subdifferential in the nonsmooth case, hypodifferential of a convex function is not uniquely defined even in the smooth case. In particular, if f is Gâteaux differentiable at a point x and $f'(x)$ is its Gâteaux derivative at this point, then both $\{(0, f'(x))\}$ and $\operatorname{co}\{(0, f'(x))), (a, x^*)\}$ for any $a < 0$ and $x^* \in X^*$ are hydifferentials of f at x.

It turns out that closed convex functions are hypodifferentiable in the interior $\operatorname{int} \operatorname{dom} f$ of their effective domain. Moreover, one can give a complete characterization of all hypodifferentials of f at any given point.

Denote by $\operatorname{core} Q$ the algebraic interior (radial kernel) of a set $Q \subset X$, that is, the set of all $x \in Q$ such that for any $y \in X$ one can find $t_y > 0$ such that $x + ty \in Q$ for any $t \in [0, t_y]$.

Theorem 3.4. *The following statements hold true:*

(1) *f is hypodifferentiable at every point $x \in \operatorname{int}\operatorname{dom} f$;*

(2) *if f is hypodifferentiable at a point $x \in \operatorname{dom} f$, then $x \in \operatorname{core}\operatorname{dom} f$, f is subdifferentiable at this point, the subdifferential $\partial f(x)$ is weak* compact, and the directional derivative $f'(x, \cdot)$ of f at x is a proper closed convex function;*

(3) *if f is hypodifferentiable at a point $x \in \operatorname{dom} f$, then a weak* compact set $H \subset \mathbb{R} \times X^*$ is a hypodifferential of f at x if and only if it satisfies the following three conditions:*

   (a) *$a \leq 0$ for any $(a, x^*) \in H$;*

   (b) *there exists $x^* \in X^*$ such that $(0, x^*) \in H$;*

   (c) *$\{x^* \in X^* \mid (0, x^*) \in H\} = \partial f(x)$.*

Proof. For the sake of convenience, we divide the proof of the theorem into three parts corresponding to the three statements of the theorem.

Part 1. Fix any $x \in \operatorname{int}\operatorname{dom} f$. Let us show that f is hypodifferentiable at x. Indeed, by [25, Corollary I.2.5] the function f is continuous in a neighbourhood of x, since X is a Banach space, f is a closed convex function, and $x \in \operatorname{int}\operatorname{dom} f$. Therefore f is subdifferentiable at x by [25, Proposition I.5.2] and $\partial f(x)$ is a weak* compact set by [55, Theorem 2.4.9]. Moreover, by [55, Theorem 2.4.9] the function f is directionally differentiable at x and for its directional derivative at this point, $f'(x, \cdot)$, the following equalities hold true:

$$f'(x, u) = \lim_{\alpha \to +0} \frac{f(x + \alpha u) - f(x)}{\alpha} = \max_{x^* \in \partial f(x)} \langle x^*, u \rangle \quad \forall u \in X.$$

Define $\underline{d} f(x) = \{0\} \times \partial f(x)$. Then the set $\underline{d} f(x)$ is weak* compact, the first equality in (1) is satisfied for any $\Delta x \in X$, and $\max_{(a, x^*) \in \underline{d} f(x)} a = 0$. Thus, the function f is hypodifferentiable at x and the set $\underline{d} f(x)$ is its hypodifferential at this point.

Part 2. Suppose now that f is hypodifferentiable at a point $x \in \operatorname{dom} f$ and $\underline{d} f(x)$ is its hypodifferential at this point. Introduce the function

$$\Phi(u) = \max_{(a, x^*) \in \underline{d} f(x)} (a + \langle x^*, u \rangle) \quad \forall u \in X.$$

Note that from the definition of hypodifferential it follows that $\Phi(0) = 0$ and $\operatorname{dom} \Phi = X$. Moreover, Φ is a closed convex function as the supremum of a family of affine function. Therefore, Φ is continuous on X, subdifferentiable at the origin, $\partial \Phi(0)$ is a weak* compact set, Φ is directionally differentiable at the origin, and

$$\Phi'(0; u) = \lim_{\alpha \to +0} \frac{\Phi(\alpha u)}{\alpha} = \max_{x^* \in \partial \Phi(0)} \langle x^*, u \rangle \quad \forall u \in X.$$

Hence taking into account the first equality in (1) one gets that for any $u \in X$

$$f'(x, u) := \lim_{\alpha \to +0} \frac{f(x + \alpha u) - f(x)}{\alpha} = \lim_{\alpha \to +0} \frac{\Phi(\alpha u)}{\alpha} = \max_{x^* \in \partial \Phi(0)} \langle x^*, u \rangle. \tag{3}$$

Consequently, $\operatorname{dom} f'(x,\cdot) = X$, which implies that $x \in \operatorname{core}\operatorname{dom} f$, and $f'(x,\cdot)$ is a proper closed convex function. Moreover, by [55, Theorem 2.4.4, part (i) and Corollary 2.4.15] one has

$$\partial f(x) = \partial f'(x,\cdot)(0) = \partial\Phi(0). \tag{4}$$

Thus, f is subdifferentiable at x and $\partial f(x) = \partial\Phi(0)$ is a weak* compact set.

Part 3. Suppose finally that f is hypodifferentiable at a point $x \in \operatorname{dom} f$ and $\underline{d}f(x)$ is its hypodifferential at this point. Then by definition one has

$$\max\{a \mid (a, x^*) \in \underline{d}f(x)\} = 0$$

or, equivalently, $a \leq 0$ for any $(a, x^*) \in \underline{d}f(x)$ and there exists $x^* \in X^*$ such that $(0, x^*) \in \underline{d}f(x)$. Note that by the theorem on the subdifferential of an infinite family of convex functions [31, Theorem 4.2.3] one has

$$\partial\Phi(0) = \left\{x^* \in X^* \;\middle|\; (0, x^*) \in \underline{d}f(x)\right\}.$$

Hence with the use of (4) one gets that $\{x^* \in X^* \mid (0, x^*) \in \underline{d}f(x)\} = \partial f(x)$.

Let now $H \subset \mathbb{R} \times X^*$ be a weak* compat set satisfying the assumptions of the theorem. Let us show that H is a hypodifferential of f at x. Indeed, introduce the function

$$\Phi_H(u) = \max_{(a,x^*)\in H} (a + \langle x^*, u\rangle) \quad \forall u \in X.$$

Then $\Phi_H(0) = 0$, $\operatorname{dom}\Phi_H = X$ due to the fact that H is weak* compact, and Φ_H is a closed convex function as the maximum of a family of affine functions. Therefore, Φ_H is continuous on X, subdifferentiable and directionally differentiable at the origin, and

$$\Phi'_H(0;u) = \lim_{\alpha\to+0} \frac{\Phi_H(\alpha u)}{\alpha} = \max_{x^*\in\partial\Phi_H(0)} \langle x^*, u\rangle \quad \forall u \in X.$$

Moreover, applying [31, Theorem 4.2.3] and the assumptions on the set H from the formulation of the theorem one gets

$$\partial\Phi_H(0) = \left\{x^* \in X^* \;\middle|\; (0, x^*) \in H\right\} = \partial f(x).$$

Consequently, taking into account equalities (3) and (4) from the second part of the proof one can conclude that

$$\lim_{\alpha\to+0} \frac{f(x+\alpha u) - f(x)}{\alpha} = \lim_{\alpha\to+0} \frac{\Phi_H(\alpha u)}{\alpha} \quad \forall u \in X,$$

which obviously yeilds

$$\lim_{\alpha\to+0} \frac{1}{\alpha}\left|f(x+\alpha u) - f(x) - \max_{(a,x^*)\in H} (a + \langle x^*, u\rangle)\right| = 0$$

for any $u \in X$. Therefore, H is a hypodifferential of f at x. □

Corollary 3.5. *For the function f to be hypodifferentiable at a point $x \in \operatorname{dom} f$ it is necessary and sufficient that $x \in \operatorname{core}\operatorname{dom} f$.*

Proof. The necessity of the condition $x \in \operatorname{core}\operatorname{dom} f$ for the hypodifferentiability of f at $x \in \operatorname{dom} f$ follows directly from the second statement of Theorem 3.4. The sufficiency of this condition follows from the first statement of Theorem 3.4 and the fact that $\operatorname{core}\operatorname{dom} f = \operatorname{int}\operatorname{dom} f$ by [55, Theorem 2.2.20], since f is l.s.c. and X is a Banach space. □

Corollary 3.6. *Let f be hypodifferentiable at a point $x_* \in X$ and $\underline{d}f(x_*)$ be its hypodifferential at this point. Then f attains a global minimum at x_* if and only if $0 \in \underline{d}f(x_*)$.*

Thus, in a sense, hypodifferential can be viewed as an extension of the subdifferential mapping into the space $\mathbb{R}\times X^*$. Such extensions can be constructed in many different ways. As we will see below, there is a natural way to construct such extensions that allows one to recover many standard properties of the gradient of a convex function in the nonsmooth case that subdifferentials do not have.

Since a hypodifferential is not uniquely defined, we restrict our consideration to a special class of hypodifferential maps of nonsmooth convex functions that are especially convenient to work with.

Definition 3.7. The function f is called *continuously* hypodifferentiable on a set $Q \subseteq \operatorname{int}\operatorname{dom} f$, if f is hypodifferentiable at every point of this set and there exists a hypodifferential mapping $\underline{d}f\colon Q \rightrightarrows \mathbb{R} \times X^*$ that is continuous with respect to the Pompeiu-Hausdorff distance. The function f is said to be continuously hypodifferentiable near a point $x \in \operatorname{int}\operatorname{dom} f$, if f is continuously hypodifferentiable in a neighbourhood of x.

Definition 3.8. Let $\underline{d}f\colon Q \rightrightarrows \mathbb{R} \times X^*$ be a hypodifferential mapping of f on a set $Q \subseteq \operatorname{dom} f$. The hypodifferential mapping $\underline{d}f(\cdot)$ is called *consistent* on the set Q, if for any $x \in Q$ one has

$$f(y) - f(x) \geq a + \langle x^*, y - x\rangle \quad \forall y \in X,\ (a, x^*) \in \underline{d}f(x). \tag{5}$$

Remark 3.9. (i) Note that a hypodifferential map $\underline{d}f(\cdot)$ is consistent on a set $Q \subseteq \operatorname{dom} f$ if and only if

$$f(y) - f(x) \geq \max_{(a,x^*)\in\underline{d}f(x)} (a + \langle x^*, y - x\rangle) \quad \forall y \in X,\ x \in Q.$$

Roughly speaking, a consistent hypodifferential of f is a weak* compact family of affine minorants of f that defines a local approximation of f at a given point in accordance with Definition 3.1.

(ii) In the author's earlier paper on the method of hypodifferential descent [18], consistent hypodifferentials were called amenable. However, the term *consistent* seems to be more suitable. □

Let us give a simple example illustrating the definition of consistent and continuous hypodifferential mappings.

Example 3.10. Let $X = \mathbb{R}$ and $f(x) = |x|$. Then for any $x, \Delta x \in \mathbb{R}$ one has

$$f(x+\Delta x)-f(x) = \max\{x+\Delta x, -x-\Delta x\}-|x| = \max\{x-|x|+\Delta x, -x-|x|-\Delta x\},$$

which implies that the set-valued map $\underline{d}f(x) := \operatorname{co}\{(x - |x|, 1), (-x - |x|, -1)\}$ is a hypodifferential mapping of f on $\mathbb{R}$.

Observe that for any $y, x \in \mathbb{R}$ one has

$$|y| - |x| \geq y - |x| = (x - |x|) + (y - x) = a + v(y - x) \tag{6}$$

for $(a, v) = (x - |x|, 1) \in \underline{d}f(x)$. Similarly, for any $y, x \in \mathbb{R}$ one has

$$|y| - |x| \geq -y - |x| = (-x - |x|) + (-1)(y - x) = a + v(y - x)$$

for $(a, v) = (-x - |x|, -1) \in \underline{d}f(x)$. Summing up inequality (6) multiplied by $\alpha \in (0, 1)$ and the inequality above multiplied by $(1-\alpha)$ one obtains that inequality (5) holds true for all $x \in \mathbb{R}$. Thus, $\underline{d}f(\cdot)$ is a consistent hypodifferential mapping on $\mathbb{R}$. Let us check that it is continuous.

Indeed, fix any $x, y \in \mathbb{R}$. By the definition of convex hull for any $(a, v) \in \underline{d}f(x)$ one has $(a, v) = ((2\alpha - 1)x - |x|, 2\alpha - 1)$ for some $\alpha \in [0, 1]$. Similarly, for any $(b, w) \in \underline{d}f(y)$ one has $(b, w) = ((2\beta - 1)y - |y|, 2\beta - 1)$. Therefore,

$$\begin{aligned}
&\max_{(a,v)\in\underline{d}f(x)} \operatorname{dist}((a, v), \underline{d}f(y)) \\
&\quad = \max_{\alpha\in[0,1]} \min_{\beta\in[0,1]} \Big(\big|(2\alpha - 1)x - |x| - ((2\beta - 1)y - |y|)\big| + \big|(2\alpha - 1) - (2\beta - 1)\big| \Big).
\end{aligned}$$

Setting $\beta = \alpha$ one gets

$$\begin{aligned}
\max_{(a,v)\in\underline{d}f(x)} \operatorname{dist}((a, v), \underline{d}f(y)) &\leq \max_{\alpha\in[0,1]} \Big|(2\alpha - 1)x - |x| - ((2\alpha - 1)y - |y|)\Big| \\
&\leq \max_{\alpha\in[0,1]} \big(|2\alpha - 1| + 1\big)|x - y| = 2|x - y|.
\end{aligned}$$

Swapping x and y one can conclude that $d_{PH}(\underline{d}f(x), \underline{d}f(y)) \leq 2|x - y|$, that is, the hypodifferential mapping $\underline{d}f$ is continuous (furthermore, globally *Lipschitz* continuous). □

Comparing inequality (5) with the inequality from the definition of ε-subdifferential one obtains the following result.

Proposition 3.11. *Let a hypodifferential map $\underline{d}f(\cdot)$ of the function f be consistent on a set $Q \subseteq H$. Then for any $x \in Q$ and $(a, x^*) \in \underline{d}f(x)$ one has $x^* \in \partial_\varepsilon f(x)$ for any $\varepsilon \geq -a$.*

Remark 3.12. From Theorem 3.4 it follows that in the general case the inclusion

$$\Big\{ x^* \in X^* \;\Big|\; (a, x^*) \in \underline{d}f(x),\ -a \leq \varepsilon \Big\} \subseteq \partial_\varepsilon f(x)$$

can be strict. This inclusion turns into an equality in the case when $\underline{d}f(\cdot)$ is a so-called *global* hypodifferential map of the function f (see [19]), that is,

$$f(y) - f(x) = \max_{(a,v)\in\underline{d}f(x)} (a + \langle x^*, y - x\rangle) \quad \forall y, x \in X.$$

Below we will provide some nontrivial examples of nonsmooth convex functions whose naturally defined hypodifferential mappings are global. Here we only note that the hypodifferential of the function $f(x) = |x|$ from Example 3.10 is global. □

Let us point out a useful property of convex functions with bounded hypodifferential mapping that can be proved directly with the use of standard results on convex functions [55, Theorem 2.4.13] (see also [44, Theorem 24.7]) and Theorem 3.4, but is also a particular case of [17, Corollary 2].

Proposition 3.13. *Let $\underline{d}f$ be a hypodifferential map of the function f on a convex set $Q \subseteq \operatorname{dom} f$. Suppose that $L = \sup\{\|x^*\| \mid (0, x^*) \in \underline{d}f(x), x \in Q\} < +\infty$. Then f is Lipschitz continuous on Q with constant L.*

3.2. Hypodifferential calculus

In this section we present main calculus rules for consistent continuous hypodifferential maps of nonsmooth convex functions. Recall that $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is a closed convex function.

Proposition 3.14. *Let f be a continuously Gâteaux differentiable function on a set $Q \subseteq \operatorname{dom} f$. Then the map $\underline{d}f(\cdot) = \{(0, f'(\cdot))\}$ is a consistent continuous hypodifferential mapping of f on Q.*

Theorem 3.15. (Maximum over a compact set) *Let T be a compact topological space and a function $f\colon X \times T \to \mathbb{R} \cup \{+\infty\}$ be given. Suppose that the following assumptions hold true:*

(1) *for any $x \in X$ the map $t \mapsto f(x,t)$ is continuous;*

(2) *for any $t \in T$ the map $f_t(\cdot) = f(\cdot, t)$ is a closed convex function;*

(3) *the set $\operatorname{int} \bigcap_{t\in T} \operatorname{dom} f_t$ is nonempty and $Q \subseteq \operatorname{int} \bigcap_{t\in T} \operatorname{dom} f_t$ is a given nonempty set;*

(4) *for any $t \in T$ there exists a consistent hypodifferential mapping $\underline{d}f_t$ of the function f_t on Q;*

(5) *the set-valued map $(x,t) \mapsto \underline{d}f_t(x)$ is continuous on $Q \times T$.*

Then the function $f(\cdot) = \sup_{t\in T} f(\cdot, t)$ is hypodifferentiable on Q and the set-valued map

$$\underline{d}f(\cdot) = \operatorname{cl}^* \operatorname{co}\left\{(f(\cdot,t) - f(\cdot), 0) + \underline{d}f_t(\cdot) \;\middle|\; t \in T\right\} \tag{7}$$

is a consistent hypodifferential of f on Q. If, in addition, either X is reflexive or the set $\operatorname{cl} D(x)$, where

$$D(x) := \operatorname{co}\{(f(x,t) - f(x), 0) + \underline{d}f_t(x) \mid t \in T\}, \tag{8}$$

is weak* closed for all $x \in Q$, then the function f is continuously hypodifferentiable on Q and the hypodifferential map (7) is continuous on Q.

Proof. For the sake of convenience, we divide the proof of the theorem into several parts.

1. Hypodifferentiability. By our assumption $Q \subseteq \operatorname{int} \bigcap_{t\in T} \operatorname{dom} f_t$, which implies that $Q \subseteq \operatorname{int} \operatorname{dom} f$ due to the fact that the map $t \mapsto f(x,t)$ is continuous for any $x \in X$. Hence by Theorem 3.4 the function f is hypodifferentiable on the set Q.

2. Equality for hypodifferential. Fix any $x \in Q$. Let us prove that the set $\underline{d}f(x)$, defined according to equality (7), is a hypodifferential of f at x. To this end, we will utilize the characterization of hypodifferentials given in Theorem 3.4.

Let us first show that the set $D(x)$ (see (8)) is norm bounded. Then one can conclude that the set $\underline{d}f(x) = \mathrm{cl}^*\, D(x)$ is weak* compact.

For any $t \in T$ the hypodifferential $\underline{d}f_t(x)$ is bounded due to the fact that it is weak* compact. By the last assumption of the theorem the map $t \mapsto \underline{d}f_t(x)$ is continuous in the Pompeiu-Hausdorff distance. Therefore, for any $t \in T$ there exists a neighbourhood of $\mathcal{U}_t$ of t such that $d_{PH}(\underline{d}f_t(x), \underline{d}f_s(x)) \leq 1$ for any $s \in \mathcal{U}_t$. Therefore the set $\bigcup_{s\in\mathcal{U}_t} \underline{d}f_s(x)$ is bounded, which due to the compactness of the space T implies that the set $\bigcup_{t\in T} \underline{d}f_t(x)$ is bounded as well. In addition, the set $\{(f(x,t) - f(x), 0) \mid t \in T\}$ is bounded due to the compactness of T and continuity of the map $t \mapsto f(x,t)$. Consequently, the sets

$$\left\{(f(x,t) - f(x), 0) + \underline{d}f_t(x) \;\middle|\; t \in T\right\}$$

and $D(x)$ (see (8)) are bounded. Hence $\underline{d}f(x)$ is a weak* compact set.

Since $f(x) = \sup_{t\in T} f(x,t)$, for any $(a, x^*) \in D(x)$ one obviously has $a \leq 0$, and, furthermore,

$$\sup_{(a,x^*)\in D(x)} a \leq \sup_{t\in T}(f(x,t) - f(x)) = \sup_{t\in T} f(x,t) - f(x) = 0.$$

Consequently, $\max\{a \mid (a, x^*) \in \underline{d}f(x)\} = 0$.

Thus, if we show that

$$\partial f(x) = \left\{x^* \in X^* \;\middle|\; (0, x^*) \in \underline{d}f(x)\right\}, \tag{9}$$

then by Theorem 3.4 one can conclude that $\underline{d}f(x)$ is a hypodifferential of f at x.

Indeed, by the theorem on the subdifferential of the supremum of an infinite family of convex functions [31, Theorem 4.2.3] and Theorem 3.4 one has

$$\partial f(x) = \mathrm{cl}^* \bigcup_{t\in T(x)} \partial f_t(x) = \mathrm{cl}^* \left\{x^* \in X^* \;\middle|\; (0, x^*) \in \underline{d}f_t(x),\ t \in T(x)\right\}, \tag{10}$$

where $T(x) = \{t \in T \mid f(x,t) = f(x)\}$. To find a more convenient expression for the set on the right-hand side of equality (9), introduce the function

$$\Phi(y) = \max_{(a,x^*)\in\underline{d}f(x)} (a + \langle x^*, y\rangle) \quad \forall y \in X.$$

Clearly, Φ is a closed real-valued convex function and $\underline{d}f(x)$ is a hypodifferential of this function at zero (note that $\Phi(0) = 0$, since $\max\{a \mid (a, x^*) \in \underline{d}f(x)\} = 0$). Therefore, by Theorem 3.4 one has

$$\partial\Phi(0) = \left\{x^* \in X^* \;\middle|\; (0, x^*) \in \underline{d}f(x)\right\}. \tag{11}$$

On the other hand, observe that for any $y \in X$ one has

$$\Phi(y) = \sup_{t\in T} \Psi(y,t), \quad \Psi(y,t) := f(x,t) - f(x) + \max_{(a,x^*)\in\underline{d}f_t(x)} (a + \langle x^*, y\rangle).$$

Clearly, for any $t \in T$ the function $y \mapsto \Psi(y,t)$ is a closed real-valued convex function that is bounded on bounded sets (due to the boundedness of $\underline{d}f_t(x)$) and therefore continuous on X by [25, Corollary I.2.5].

In turn, for any $y \in X$ the function $t \mapsto \Psi(y,t)$ is, as one can readily check, continuous due to the continuity of the maps $t \mapsto f(x,t)$ and $t \mapsto \underline{d}f_t(x)$. Note finally that $\Psi(0,t) = \Phi(0)$ if and only if $t \in T(x)$, since $\Psi(0,t) = f(x,t) - f(x)$. Therefore, by the theorem on the subdifferential of the supremum of an infinite family of convex functions [31, Theorem 4.2.3] one has

$$\partial\Phi(0) = \mathrm{cl}^* \left\{ \partial_y \Psi(0,t) \;\middle|\; t \in T(x) \right\}, \tag{12}$$

where $\partial_y \Psi(0,t)$ is the subdifferential of the map $\Psi(\cdot,t)$ at the origin.

Taking into account the fact that for any $t \in T(x)$ one has

$$\Psi(y,t) = \max_{(a,x^*) \in \underline{d}f_t(x)} (a + \langle x^*, y \rangle) \quad \forall y \in X,$$

one can conclude that $\underline{d}f_t(x)$ is a hypodifferential of $\Psi(\cdot,t)$ at the origin. Hence

$$\partial_y \Psi(0,t) = \left\{ x^* \in X^* \;\middle|\; (0,x^*) \in \underline{d}f_t(x) \right\} \quad \forall t \in T(x)$$

by Theorem 3.4. Now, combining the equality above with equalities (11), (12), and (10) one obtains that equality (9) holds true. Thus, $\underline{d}f(x)$ is indeed a hypodifferential of f at x.

3. Consistency. Let us check that the hypodifferential map $\underline{d}f(\cdot)$ is consistent. Indeed, for any $x \in Q$, $t \in T$, $(a,x^*) \in \underline{d}f_t(x)$, and $y \in X$ one has

$$f(y) - f(x) \geq f(y,t) - f(x) \geq f(x,t) - f(x) + a + \langle x^*, y - x \rangle, \tag{13}$$

since $\underline{d}f_t(x)$ is a consistent hypodifferential of f_t at x and $f(y) = \sup_{t \in T} f(y,t)$.

Fix now any $x \in Q$ and $(a,x^*) \in D(x)$. Then by the definition of convex hull there exists $n \in \mathbb{N}$, $t_i \in T$, $(a_i, x_i^*) \in \underline{d}f_{t_i}(x)$, and $\alpha_i \geq 0$, $i \in \{1,\ldots,n\}$, such that

$$(a,x^*) = \sum_{i=1}^n \alpha_i (f(x,t_i) - f(x) + a_i, x_i^*), \quad \sum_{i=1}^n \alpha_i = 1.$$

Hence applying inequality (13) one gets for any $y \in X$ that

$$\begin{aligned} f(y) - f(x) &= \sum_{i=1}^n \alpha_i (f(y) - f(x)) \\ &\geq \sum_{i=1}^n \alpha_i (f(x,t_i) - f(x) + a_i + \langle x^*, y - x \rangle) = a + \langle x^*, y - x \rangle. \end{aligned} \tag{14}$$

Finally, fix any $(a,x^*) \in \underline{d}f(x)$, $y \in X$, and $\varepsilon > 0$. Since $\underline{d}f(x) = \mathrm{cl}^* D(x)$, one can find $(b,y^*) \in D(x)$ such that $|a-b| + |\langle y^* - x^*, y - x \rangle| < \varepsilon$. Consequently, with the use of inequality (14) one gets that

$$f(y) - f(x) \geq b + \langle y^*, y - x \rangle \geq a + \langle x^*, y - x \rangle - \varepsilon,$$

which implies that

$$f(y) - f(x) \geq a + \langle x^*, y - x \rangle \quad \forall y \in X$$

due to the fact that $\varepsilon > 0$ was chosen arbitrarily.

Thus, the inequality above is satisfied for all $(a, x^*) \in \underline{d}f(x)$ and $y \in X$, which means that the hypodifferential $\underline{d}f(x)$ is consistent.

4. Continuity. Let us finally prove the continuity of the hypodifferential map $\underline{d}f(\cdot)$ on Q. Indeed, as was noted above, $Q \subseteq \operatorname{int}\operatorname{dom} f$. Therefore, the function f is continuous on Q by [25, Corollary I.2.5]. Consequently, under the assumptions of the theorem the set-valued map

$$\left\{(f(\cdot,t) - f(\cdot), 0) + \underline{d}f_t(x) \;\middle|\; t \in T\right\}$$

is continuous in the Pompeiu-Hausdorff metric on Q by Proposition 2.5, which by Proposition 2.6 implies that the set-valued map $D(\cdot)$ is continuous on Q. Finally, the continuity of $\underline{d}f(\cdot) = \operatorname{cl}^* D(\cdot)$ then trivially follows from the continuity of $D(\cdot)$ in the case when $\operatorname{cl} D(\cdot)$ is weak* closed, since in this case $\underline{d}f(\cdot) = \operatorname{cl} D(\cdot)$ and $d_{PH}(\operatorname{cl} D(x), \operatorname{cl} D(y)) = d_{PH}(D(x), D(y))$ for any $x, y \in Q$. In the case when X is reflexive the continuity of $\underline{d}f(\cdot)$ follows directly from Proposition 2.8. □

Corollary 3.16. (Finite maximum) *Let $f_i\colon X \to \mathbb{R} \cup \{+\infty\}$, $i \in I = \{1, \ldots, n\}$, be closed convex functions such that $\bigcap_{i\in I} \operatorname{int}\operatorname{dom} f_i \neq \emptyset$. Suppose that there exist consistent continuous hypodifferential mappings $\underline{d}f_i(\cdot)$ of these functions on a nonempty set $Q \subseteq \bigcap_{i\in I} \operatorname{int}\operatorname{dom} f_i$. Then the function $f = \max_{i\in I} f_i$ is continuously hypodifferentiable on Q and the set-valued map*

$$\underline{d}f(\cdot) = \operatorname{co}\left\{\{(f_i(\cdot) - f(\cdot), 0)\} + \underline{d}f_i(\cdot) \;\middle|\; i \in I\right\} \tag{15}$$

is a consistent continuous hypodifferential of f on Q.

Proof. Put $T = \{1, \ldots, n\}$ and endow it with discrete topology. For any $x \in X$ and $i \in T$ define $f(x, i) = f_i(x)$. Then, as one can readily check, all assumptions of the previous theorem hold true. It remains to note that the set $\underline{d}f(x)$, defined according to equality (15), is weak* closed for any $x \in Q$ as the convex hull of a *finite* number of weak* compact convex sets. □

Recall that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called non-decreasing, if $f(x) \leq f(y)$ for any $x, y \in \mathbb{R}^n$ such that $x \leq y$, where the inequality $x \leq y$ is understood coordinate-wise.

Theorem 3.17. (Outer composition with differentiable function)
Let $f_i\colon X \to \mathbb{R} \cup \{+\infty\}$, $i \in I = \{1, \ldots, n\}$, be closed convex functions and there exist consistent continuous hypodifferential mappings $\underline{d}f_i(\cdot)$ of these functions on a set $Q \subseteq \bigcap_{i\in I} \operatorname{int}\operatorname{dom} f_i$. Let also $g\colon \mathbb{R}^n \to \mathbb{R}$, $g = g(y)$, be a non-decreasing convex function that is continuously differentiable in a neighbourhood of the set $f(Q)$, where $f = (f_1, \ldots, f_n)$. Then the convex function $h(\cdot) = g(f_1(\cdot), \ldots, f_n(\cdot))$ is defined in a neighbourhood of Q, continuously hypodifferentiable on Q, and the set-valued map

$$\underline{d}h(\cdot) = \sum_{i=1}^{n} \frac{\partial g(f(\cdot))}{\partial y_i} \underline{d}f_i(\cdot) \tag{16}$$

is a consistent continuous hypodifferential of h on Q (here $h(x) = +\infty$ if we have $f_i(x) = +\infty$ for some $i \in I$).

Proof. The convexity of the composition $h = g \circ f$ follows from the fact that g is non-decreasing and can be easily verified directly. Note also that the partial derivatives $\partial g(f(\cdot))/\partial y_i$ are nonnegative, since g is non-decreasing.

The fact that h is hypodifferentiable on Q and (16) is its hypodifferential mapping on this set is proved in exactly the same way as [17, Theorem 2]. The continuity of this mapping follows from Propositions 2.3 and 2.4 and the fact that f_i are continuous on Q by Proposition 3.13.

Let us finally check that the hypodifferential mapping (16) is consistent on Q. Indeed, choose any $x \in Q$ and $(a, x^*) \in \underline{d}h(x)$. Then there exists $(a_i, x_i^*) \in \underline{d}f_i(x)$, $i \in I$ such that

$$(a, x^*) = \sum_{i=1}^{n} \frac{\partial g(f(x))}{\partial y_i}(a_i, x_i^*).$$

Choose any $y \in X$. Bearing in mind the convexity of g one has

$$h(y) - h(x) = g(f(y)) - g(f(x)) \geq \sum_{i=1}^{n} \frac{\partial g(f(x))}{\partial y_i}\big(f_i(y) - f_i(x)\big).$$

Since the hypodifferentials $\underline{d}f_i$ are consistent, one has

$$f_i(y) - f_i(x) \geq a_i + \langle x_i^*, y - x\rangle \quad \forall i \in I.$$

Hence taking into account the fact that the partial derivatives $\partial g(f(x))/\partial y_i$ are nonnegative one obtains that

$$h(y) - h(x) \geq \sum_{i=1}^{n} \frac{\partial g(f(x))}{\partial y_i}\big(a_i + \langle x_i^*, y - x\rangle\big) = a + \langle x^*, y - x\rangle,$$

which completes the proof. □

Remark 3.18. The theorem above can be slightly generalised in the following way. Namely, suppose that $K \subset \mathbb{R}^n$ is a convex cone such that if $x \in K$ and $x \leq y$, then $y \in K$ (in particular, K can be the nonnegative orthant). Then under the assumption that f maps X to K one can suppose that g is defined and nondecreasing only on K. □

Corollary 3.19. (Conic combination) *Let $f_i \colon X \to \mathbb{R} \cup \{+\infty\}$, $i \in I = \{1, \ldots, n\}$, be closed convex functions and there exist consistent continuous hypodifferential mappings $\underline{d}f_i(\cdot)$ of these functions on a set $Q \subseteq \bigcap_{i\in I} \operatorname{dom} f_i$. Then for any $\lambda_i \geq 0$, $i \in I$, the function $f = \sum_{i\in I} \lambda_i f_i$ is continuously hypodifferentiable on Q and the set-valued map*

$$\underline{d}f(\cdot) = \sum_{i\in I} \lambda_i \underline{d}f_i(\cdot)$$

is a consistent continuous hypodifferential of f on Q.

Corollary 3.20. *Let there exist a consistent continuous hypodifferential mapping $\underline{d}f(\cdot)$ of f on a set $Q \subseteq \operatorname{int}\operatorname{dom} f$. Then the functions $\max\{0, f(\cdot)\}^p$, $p > 1$, and $\exp(f(\cdot))$ are continuously hypodifferentiable on Q and the set-valued maps*

$$\max\{0, f(\cdot)\}^{p-1}\underline{d}f(\cdot), \quad \exp(f(\cdot))\underline{d}f(\cdot)$$

are their consistent continuous hypodifferential mappings on the set Q.

The following result can be readily verified directly

Proposition 3.21. (Composition with affine map) *Let Y be a real Banach space, $A\colon Y \to X$ be a continuous linear operator, $b \in X$ be fixed, and there exist a consistent continuous hypodifferential map $\underline{d}f(\cdot)$ of f on a set $Q \subseteq \operatorname{int}\operatorname{dom} f$. Then the function $g(y) = f(Ay + b)$, $y \in Y$, is continuously hypodifferentiable on the set $A^{-1}(Q - b)$ and the set-valued map*

$$\underline{d}g(\cdot) = \Big\{(a, A^*x^*) \in \mathbb{R} \times Y^* \;\Big|\; (a, x^*) \in \underline{d}f(\cdot)\Big\}, \tag{17}$$

is a consistent continuous hypodifferential map of g on $A^{-1}(Q - b)$, where A^* is the adjoint operator of A.

3.3. Examples

Let us give several nontrivial examples of hypodifferentials of nonsmooth convex functions. It should be noted that the hypodifferentials computed in all examples below are *global* (see Remark 3.12).

Example 3.22. (Polyhedral function) Let X be the space $\mathbb{R}^d$ equipped with the Euclidean norm $|\cdot|$. Let also $f\colon \mathbb{R}^d \to \mathbb{R}$ be a polyhedral convex function, that is, the epigraph of f is the intersection of a finite family of closed half-spaces. Then there exist $n \in \mathbb{N}$, $a_i \in \mathbb{R}$ and $v_i \in \mathbb{R}^d$, $i \in I = \{1, \ldots, n\}$, such that

$$f(x) = \max_{i \in I}(a_i + \langle v_i, x\rangle) \quad \forall x \in \mathbb{R}^d$$

(see [44, Section 19]). Consequently, for any $x, \Delta x \in \mathbb{R}^d$ one has

$$\begin{aligned} f(x + \Delta x) - f(x) &= \max_{i \in I}\big(a_i + \langle v_i, x\rangle - f(x) + \langle v_i, \Delta x\rangle\big) \\ &= \max_{(a,v) \in \underline{d}f(x)}(a + \langle v, x\rangle), \end{aligned} \tag{18}$$

where

$$\underline{d}f(\cdot) = \operatorname{co}\Big\{\big(a_i + \langle v_i, \cdot\rangle - f(\cdot), v_i\big) \in \mathbb{R}^{d+1} \;\Big|\; i \in I\Big\}. \tag{19}$$

Clearly, $\underline{d}f(x)$ is a convex compact set and $\max\{a \mid (a, v) \in \underline{d}f(x)\} = 0$ for any $x \in \mathbb{R}^d$. Therefore, f is hypodifferentiable on $\mathbb{R}^d$ and $\underline{d}f(\cdot)$ is its hypodifferential map on $\mathbb{R}^d$ that is consistent and global due to equality (18).

Let us show that the set-valued map $\underline{d}f(\cdot)$ is continuous. Indeed, fix any $x, y \in \mathbb{R}^d$ and choose any $(a, v) \in \underline{d}f(x)$. Then by definition there exist $\alpha_i \geq 0$, $i \in I$, such that

$$(a, v) = \sum_{i=1}^{n} \alpha_i(a_i + \langle v_i, x\rangle - f(x), v_i), \quad \sum_{i=1}^{n} \alpha_i = 1.$$

Define

$$b = \sum_{i=1}^{n} \alpha_i(a_i + \langle v_i, y\rangle - f(y)).$$

Then $(b, v) \in \underline{d}f(y)$ and taking into account the fact that f is globally Lipschitz continuous with Lipschitz constant $L = \max_{i \in I}|v_i|$ (see Proposition 3.13) one obtains

$$\begin{aligned}\operatorname{dist}((a,v),\underline{d}f(y)) &\le \big|(a,v)-(b,v)\big| = |a-b| \\ &\le \sum_{i=1}^{n} \alpha_i\big(|\langle v_i, x-y\rangle| + |f(x)-f(y)|\big) \\ &\le \max_{i\in I}|v_i||x-y| + L|x-y| = 2L|x-y|.\end{aligned}$$

Taking the supremum over all $(a,v)\in \underline{d}f(x)$ and then repeating the same argument with x and y swapped one obtains that

$$d_{PH}\big(\underline{d}f(x),\underline{d}f(y)\big) \le 2\Big(\max_{i\in I}|v_i|\Big)|x-y| \quad \forall x,y\in\mathbb{R}^d.$$

Thus, f is continuously hypodifferentiable on $\mathbb{R}^d$ and the set-valued map (19) is its consistent continuous hypodifferential map on $\mathbb{R}^d$.

Example 3.23. (Sublinear function) Let $f\colon X\to\mathbb{R}$ be a closed sublinear function. Then the subdifferential $\partial f(0)$ is weak* compact and

$$f(x) = \max_{x^*\in\partial f(0)}\langle x^*, x\rangle \quad \forall x\in X$$

(see, e.g. [55, Theorem 2.4.9]). Consequently, for any $x,\Delta x\in X$ one has

$$f(x+\Delta x)-f(x) = \max_{x^*\in\partial f(0)}\big(\langle x^*,x\rangle - f(x) + \langle x^*,\Delta x\rangle\big) = \max_{(a,x^*)\in\underline{d}f(x)}(a+\langle x^*,x\rangle), \quad (20)$$

where
$$\underline{d}f(\cdot) = \Big\{(\langle x^*,\cdot\rangle - f(\cdot), x^*)\in\mathbb{R}\times X^* \;\Big|\; x^*\in\partial f(0)\Big\} \quad (21)$$

Observe that for any $x\in X$ one has

$$\max_{(a,x^*)\in\underline{d}f(x)} a = \max_{x^*\in\partial f(0)}(\langle x^*,x\rangle - f(x)) = f(x)-f(x) = 0.$$

Moreover, $\underline{d}f(x) = \mathcal{T}(\partial f(0))$ with

$$\mathcal{T}(x^*) = \Big(\langle x^*,x\rangle - f(x), x^*\Big)\in\mathbb{R}\times X^* \quad \forall x^*\in X^*.$$

Clearly, $\mathcal{T}$ is an affine map, which implies that the set $\underline{d}f(x)$ is convex. Moreover, as one can readily check, $\mathcal{T}$ continuously maps X^* endowed with the weak* topology to $\mathbb{R}\times X^*$ equipped with the weak* topology. Therefore the set $\underline{d}f(x)$ is weak* compact as the continuous image of a compact set. Thus, f is hypodifferentiable on X and the set-valued map (21) is its consistent hypodifferential mapping on X that is, furthermore, global (see equality (20)).

Let us prove that the map $\underline{d}f$ is continuous. Indeed, note that f is globally Lipschitz continuous with Lipschitz constant $L = \sup\{\|x^*\| \mid x^*\in\partial f(0)\}$ (see Proposition 3.13). Fix any $x,y\in X$ and choose $(a,x^*)\in\underline{d}f(x)$. Then by definion $x^*\in\partial f(0)$ and $a = \langle x^*,x\rangle - f(x)$. Define $b = \langle x^*,y\rangle - f(y)$. Then $(b,x^*)\in\underline{d}f(y)$ and

$$\begin{aligned}\operatorname{dist}\big((a,x^*),\underline{d}f(y)\big) &\le \|(a,x^*)-(b,x^*)\| = |a-b| \\ &\le |\langle x^*,x-y\rangle| + |f(x)-f(y)| \le 2L\|x-y\|\end{aligned}$$

Hence taking the supremum over all $(a, x^*) \in \underline{d}f(x)$ and then repeating the same argument with x and y swapped one obtains that

$$d_{PH}(\underline{d}f(x), \underline{d}f(y)) \leq 2\Big(\sup_{x^* \in \partial f(0)} \|x^*\|\Big)\|x - y\|.$$

Thus, f is continuously hypodifferentiable and the set-valued map (21) is its consistent continuous hypodifferential map on X. □

Example 3.24. (Norm of an affine map) Suppose that Y is a real Banach space, $A\colon X \to Y$ is a bounded linear operator, and $f(x) = \|Ax + b\|$ for some $b \in Y$. Then by the previous example and Proposition 3.21 the function f is continuously hypodifferentiable on X and the set-valued map

$$\underline{d}f(x) = \left\{\Big(\langle x^*, Ax + b\rangle - \|Ax + b\|, A^*x^*\Big) \in \mathbb{R} \times X^* \;\Big|\; \|x^*\| \leq 1\right\} \quad \forall x \in X$$

is its consistent continuous hypodifferential mapping on X. Moreover, this hypodifferential map is global, and arguing as in the previous example one can readily check that

$$d_{PH}\big(\underline{d}f(x), \underline{d}f(y)\big) \leq 2\|A\|\|x - y\| \quad \forall x, y \in X. \qquad \square$$

Example 3.25. (Maximal eigenvalue) Let X be the space $\mathbb{S}^\ell$ of all real symmetric matrices of order ℓ equipped with the inner product $\langle A, B\rangle = \mathrm{Tr}(AB)$ and the corresponding norm (here $\mathrm{Tr}(A)$ is the trace of a square matrix A). Let $f(A) = \lambda_{\max}(A)$, $A \in \mathbb{S}^\ell$, be the maximal eigenvalue function. Clearly, f is a closed sublinear function. Taking into the well-known representation

$$\lambda_{\max}(A) = \max_{v \in \mathbb{B}_\ell} \langle v, Av\rangle \quad \forall A \in \mathbb{S}^\ell$$

one can readily check that $\partial f(0) = \{v^T v \mid v \in \mathbb{B}_\ell\}$, where $\mathbb{B}_\ell$ is the unit ball in $\mathbb{R}^\ell$. Therefore, by Example 3.23 the function f is continuously hypodifferentiable on $\mathbb{S}^\ell$ and the set-valued map

$$\underline{d}f(A) = \left\{(\langle v, Av\rangle - \lambda_{\max}(A), v^T v) \in \mathbb{R} \times \mathbb{S}^\ell \;\Big|\; v \in \mathbb{B}_\ell\right\} \quad \forall A \in \mathbb{S}^\ell$$

is its consistent continuous hypodifferential mapping on X that is, furthermore, global. In addition, one can easily check that

$$d_{PH}\big(\underline{d}f(A), \underline{d}f(B)\big) \leq 2\|A - B\| \quad \forall A, B \in \mathbb{S}^\ell.$$

Example 3.26. (Distance to a cone) Let $K \subset X$ be a closed convex cone and $f(x) = \mathrm{dist}(x, K)$ for all $x \in X$. Bearing in mind the fact that K is a convex cone one can easily verify that f is a closed sublinear function. By [4, Example 2.130] one has

$$\partial f(0) = \big\{x^* \in K^* \;\big|\; \|x^*\| \leq 1\big\}, \quad K^* = \left\{x^* \in X^* \;\Big|\; \langle x^*, x\rangle \leq 0 \,\forall x \in K\right\}$$

(here K^* is the polar cone of K).

Hence by Example 3.23 the function f is continuously hypodifferentiable on X and the set-valued map

$$\underline{d}f(x) = \Big\{(\langle x^*, x\rangle - \operatorname{dist}(x, K), x^*) \in \mathbb{R} \times X^* \;\Big|\; x^* \in K^*,\ \|x^*\| \leq 1\Big\}$$

is its global and consistent continuous hypodifferential mapping on X. Moreover, in this case

$$d_{PH}\big(\underline{d}f(x), \underline{d}f(y)\big) \leq 2\|x - y\| \quad \forall x, y \in X.$$

Remark 3.27. Let us note that a consistent continuous hypodifferential map of the convex functional

$$F(u) = \int_\Omega f(u(x), \nabla u(x), x)\, dx \quad \forall u \in W^{1,p}(\Omega),$$

defined on the Sobolev space $W^{1,p}(\Omega)$, can be computed with the use of [16, Theorem 5.1] and [20, Theorem 3.3] under the assumptions that $f = f(u, \xi, x)$ is convex in (u, ξ) for a.e. $x \in \Omega$ and the function f along with its hypodifferential mapping satisfy some natural growth conditions. □

3.4. Lipschitzian approximations and Lipschitz continuous hypodifferentials

The Lipschitz continuity of the gradient of the objective function plays a fundamental role in the design and analysis of numerical methods of convex optimization [5, 24, 37, 49]. In the case when X is a Hilbert space and f is differentiable, the Lipschitz continuity of the gradient of f on a convex set Q implies that

$$\Big|f(y) - f(x) - \langle \nabla f(x), y - x\rangle\Big| \leq \frac{L}{2}\|y - x\|^2 \quad \forall x, y \in Q$$

(see, e.g. [37, Lemma 1.2.3]). This inequality is sometimes sufficient by itself for convergence analysis of optimization methods and is often exploited, for example, in the context of weakly convex optimization (see, e.g. [1, 6] and the references therein). In [18], a natural extension of this inequality to the case of nonsmooth hypodifferentiable convex functions was proposed and studied.

Definition 3.28. Let $Q \subset \operatorname{int}\operatorname{dom} f$ be a nonempty set and $\underline{d}f$ be a hypodifferential mapping of f on Q. This mapping is called a *Lipschitzian approximation* of f on Q with Lipschitz constant $L > 0$, if we have for all $x, y \in Q$:

$$\Big|f(y) - f(x) - \max_{(a,x^*)\in\underline{d}f(x)} \big(a + \langle x^*, y - x\rangle\big)\Big| \leq \frac{L}{2}\|x - y\|^2. \tag{22}$$

Observe that if a hypodifferential mapping $\underline{d}f(\cdot)$ is consistent on Q, then it is a Lipschitzian approximation of f on Q with Lipschitz constant $L > 0$ if and only if

$$f(y) \leq \varphi_x(y) := f(x) + \max_{(a,x^*)\in\underline{d}f(x)} \big(a + \langle x^*, y - x\rangle\big) + \frac{L}{2}\|y - x\|^2$$

for all $x, y \in Q$ (that is, the function $\varphi_x(\cdot)$ is a majorant of f on Q for any $x \in Q$).

One can also consider the property of the Lipschitz continuity of hypodifferential, which in the convex setting can be viewed as a natural extension of the Lipschitz continuity of the gradient to the nonsmooth case.

Definition 3.29. Let $Q \subset \operatorname{int}\operatorname{dom} f$ be a nonempty set and $\underline{d}f$ be a hypodifferential mapping of f on Q. This mapping is said to be *Lipschitz continuous* on Q with Lipschitz constant $L > 0$, if

$$d_{PH}\big(\underline{d}f(x), \underline{d}f(y)\big) \leq L\|x - y\| \quad \forall x, y \in Q$$

or, equivalently, $$\underline{d}f(y) \subseteq \underline{d}f(x) + L\|y - x\|\mathbb{B}_{\mathbb{R}\times X^*} \quad \forall x, y \in Q,$$

where $\mathbb{B}_{\mathbb{R}\times X^*}$ is the unit ball in $\mathbb{R}\times X^*$. If a hypodifferential map $\underline{d}f$ is both Lipschitz continuous on Q and a Lipschitzian approximation of f on Q, then it is said to have *the Lipschitz property* on Q.

The Lipschitz continuity of a hypodifferential map and the property of it being a Lipschitzian approximation seem to be closely related, as we will show below. However, it is unclear whether these properties are equivalent in the general nonsmooth case. We leave the difficult question of whether these properties are equivalent (at least under some additional assumptions) as an interesting problem for future research. Here we only prove a partial result showing that a certain upper estimate of the function

$$f(y) - f(x) - \max_{(a,x^*)\in\underline{d}f(x)} (a + \langle x^*, y - x\rangle)$$

holds true for any Lipschitz continuous hypodifferential map $\underline{d}f(\cdot)$.

Proposition 3.30. *Let $\underline{d}f$ be a consistent hypodifferential map of f that is defined and Lipschitz continuous with Lipschitz constant $L > 0$ on a convex set $Q \subseteq \operatorname{int}\operatorname{dom} f$. Then for any $x, y \in Q$ one has*

$$0 \leq f(y) - f(x) - \max_{(a,x^*)\in\underline{d}f(x)} (a + \langle x^*, y - x\rangle) \leq L\|y - x\|\big(1 + \|y - x\|\big).$$

Proof. Fix any $x, y \in Q$ and $\varepsilon > 0$. By [17, Proposition 2] there exists $z \in \operatorname{co}\{x, y\}$ and $(0, x_z^*) \in \underline{d}f(z)$ such that we have $f(y) - f(x) = \langle x_z^*, y - x\rangle$. By our assumption $d_{PH}(\underline{d}f(x), \underline{d}f(z)) \leq L\|z - x\|$. Therefore, by Proposition 2.1 there exists some $(a_0, x_0^*) \in \underline{d}f(x)$ such that $\|(0, x_z^*) - (a_0, x_0^*)\| \leq L\|z - x\| + \varepsilon$. Hence

$$\begin{aligned}
&f(y) - f(x) - \max_{(a,x^*)\in\underline{d}f(x)} (a + \langle x^*, y - x\rangle) \\
&\quad = \langle x_z^*, y - x\rangle - \max_{(a,x^*)\in\underline{d}f(x)} (a + \langle x^*, y - x\rangle) \\
&\quad \leq \langle x_z^*, y - x\rangle - a_0 - \langle x_0^*, y - x\rangle \leq \big\|(0, x_z^*) - (a_0, x_0^*)\big\|\big(1 + \|y - x\|\big) \\
&\quad \leq \Big(L\|z - x\| + \varepsilon\Big)\big(1 + \|y - x\|\big).
\end{aligned}$$

It remains to pass to the limit as $\varepsilon \to +0$ and note that $\|z - x\| \leq \|y - x\|$, since $z \in \operatorname{co}\{x, y\}$. □

Let us now turn to the analysis of existence of hypodifferentials of nonsmooth convex functions having the Lipschitz property. We start with the simplest case of global hypodifferentials. Namely, let us show that global hypodifferentials maps are, in a sense, always globally Lipschitz continuous (cf. examples in Subsection 3.3). We need an auxiliary definition to precisely formulate this result.

Definition 3.31. Let a hypodifferential map $\underline{d}f$ of f be defined on an open set $Q \subseteq \operatorname{dom} f$. This map is said to be *exact* on Q, if

$$f(y) - f(x) = \max_{(a,x^*)\in \underline{d}f(x)} (a + \langle x^*, y - x\rangle) \quad \forall y, x \in Q.$$

Proposition 3.32. *If there exists a hypodifferential map of f that is defined and exact on an open convex set $Q \subseteq \operatorname{int}\operatorname{dom} f$, then f is Lipschitz continuous on Q and there exists an exact hypodifferential map of f that is defined and has the Lipschitz property on Q.*

Proof. Let us define a new hypodifferential map $D_f(\cdot)$ of f on Q. To this end, fix any point $x_0 \in X$ and choose any $x \in Q$ and $\Delta x \in X$ such that $x + \Delta x \in Q$ (recall that Q is an open set). By applying the exactness property of the hypodifferential one gets

$$\begin{aligned} f(x + \Delta x) - f(x) &= f(x_0) - f(x) + \max_{(a,x^*)\in \underline{d}f(x_0)} (a + \langle x^*, x + \Delta x - x_0\rangle) \\ &= \max_{(a,x^*)\in D_f(x)} (a + \langle x^*, \Delta x\rangle), \end{aligned}$$

where $D_f(x) = \Big\{(a + \langle x^*, x - x_0\rangle + f(x_0) - f(x), x^*) \in \mathbb{R} \times X^* \;\Big|\; (a, x^*) \in \underline{d}f(x_0)\Big\}$.

Note that $$\begin{aligned} \max_{(a,x^*)\in D_f(x)} a &= \max_{(a,x^*)\in \underline{d}f(x_0)} \big(a + \langle x^*, x - x_0\rangle\big) + f(x_0) - f(x) \\ &= f(x) - f(x_0) + f(x_0) - f(x) = 0. \end{aligned}$$

Furthermore, the set $D_f(x)$ is convex and weak* compact, since $D_f(x)$ is the image of the convex weak* compact set $\underline{d}f(x_0)$ under the affine map

$$\mathcal{T}_x(a, x^*) = (a + \langle x^*, x - x_0\rangle + f(x_0) - f(x), x^*)$$

that is obviously continuous with respect to the weak* topology. Thus, $D_f(x)$ is a hypodifferential of f at x. Clearly, $D_f(\cdot)$ is an exact hypodifferential map of f on Q, which implies that $D_f(\cdot)$ is a Lipschitzian approximation of f on Q. Let us check that this hypodifferential map is also Lipschitz continuous on Q.

First, note that the quantity $L := \sup\{\|x^*\| \mid (a, x^*) \in D_f(x)\}$ does not depend on x by the definition of $D_f(\cdot)$. Therefore by Proposition 3.13 the function f is Lipschitz continuous on Q with Lipschitz constant L.

Fix any $x, y \in Q$. Let $(b, y^*) \in D_f(y)$. Then there exists $(a_0, x_0^*) \in \underline{d}f(x_0)$ such that $(b, y^*) = \mathcal{T}_y(a_0, x_0^*)$. Let $(a, x^*) = \mathcal{T}_x(a_0, x_0^*)$. Then $(a, x^*) \in D_f(x)$ and

$$\begin{aligned} \operatorname{dist}((b, y^*), D_f(x)) \le \|(b, y^*) - (a, x^*)\| &\le |\langle x_0^*, y - x\rangle| + |f(y) - f(x)| \\ &\le L\|y - x\| + L\|y - x\|. \end{aligned}$$

Consequently, one has $\sup\limits_{(b,y^*)\in D_f(y)} \operatorname{dist}\big((b, y^*), D_f(x)\big) \le 2L\|y - x\|$.

Now, repeating the same argument with x and y swapped one can conclude that $d_{PH}(D_f(x), D_f(y)) \le 2L\|y - x\|$. Thus, the hypodifferential map $D_f(\cdot)$ is Lipschitz continuous on Q. □

Nonetheless, as one can expect, not every convex function has a globally Lipschitz continuous hypodifferential mapping.

Proposition 3.33. *Let* $\operatorname{dom} f = X$ *and there exist a hypodifferential map of* f *defined on* X *and globally Lipschitz continuous. Then there exists* $C_1, C_2 \geq 0$ *such that* $f(x) \leq C_1 + C_2\|x\|^2$ *for all* $x \in X$.

Proof. Fix any $x_0 \in X$ and denote $\theta = \sup\{\|x^*\| \mid (a, x^*) \in \underline{d}f(x_0)\}$. Let also L be a Lipschitz constant from the definition of Lipschitz continuous hypodifferential map. By Proposition 3.30 one has

$$f(x) \leq f(x_0) + \max_{(a,x^*)\in \underline{d}f(x_0)} (a + \langle x^*, x - x_0\rangle) + L\|x - x_0\|\big(1 + \|x - x_0\|\big)$$

for any $x \in X$. Taking into account the fact that $a \leq 0$ for any $(a, x^*) \in \underline{d}f(x_0)$ and applying the obvious inequality $(a+b)^2 \leq 2a^2 + 2b^2$ one gets that

$$\begin{aligned} f(x) &\leq f(x_0) + (\theta + L)\|x - x_0\| + L\|x - x_0\|^2 \\ &\leq f(x_0) + (\theta + L)\|x_0\| + 2L\|x_0\|^2 + (\theta + L)\|x\| + 2L\|x\|^2. \end{aligned}$$

for any $x \in X$. Hence with the use of the inequality $\|x\| \leq 0.5 + 0.5\|x\|^2$ one obtains that $f(x) \leq C_1 + C^2\|x\|^2$ with

$$C_1 = f(x_0) + (\theta + L)\|x_0\| + 2L\|x_0\|^2 + \frac{1}{2}(\theta + L), \quad C_2 = 2L + \frac{1}{2}(\theta + L),$$

and the proof is complete. □

Remark 3.34. Note that if there exists an exact hypodifferential map of f on X, then with the use of the equality

$$f(x) - f(x_0) = \max_{(a,x^*)\in \underline{d}f(x_0)} (a + \langle x^*, x - x_0\rangle) \quad \forall x \in X$$

one gets that there exist $C_1, C_2 \geq 0$ such that $|f(x)| \leq C_1 + C_2\|x\|$ for all $x \in X$. □

Thus, for the convex function $f(x) = \|x\|^p$ with $p > 2$ there does not exist a globally Lipschitz continuous hypodifferential map (or a hypodifferential map that has the Lipschitz property on the entire space). However, in the case when the space X is reflexive, a consistent hypodifferential map of a real-valued convex function that has the Lipschitz property locally *always* exists. Moreover, under some mild assumptions there even exists a consistent hypodifferential map that has the Lipschitz property on any given bounded set. The proof of this result is inspired by the discussion from [10, Appendix I].

Theorem 3.35. *Let* X *be a relfexive Banach space,* $Q \subseteq \operatorname{int}\operatorname{dom} f$ *be a bounded open set, and* f *be Lipschitz continuous on* Q. *Then there exists a hypodifferential mapping of* f *that is defined, consistent, exact, and has the Lipschitz property on* Q. *In particular, the following statements hold true:*

(1) *for any* $x \in \operatorname{int}\operatorname{dom} f$ *there exists a neighbourhood* $\mathcal{U}_x$ *of* x *and a hypodifferential mapping of* f *that is defined, consistent, exact, and has the Lipschitz property on* $\mathcal{U}_x$;

(2) *for any bounded open set* $Q \subseteq \operatorname{int}\operatorname{dom} f$ *such that the subdifferential map* $\partial f(\cdot)$ *is bounded on* Q, *there exists a hypodifferential mapping of* f *that is defined, consistent, exact, and has the Lipschitz property on* Q.

Proof. For the sake of convenience, we divide the proof of the theorem into several parts. First, let us note that the existence of a hypodifferential map having the required properties in a neighbourhood of any point $x \in \operatorname{int}\operatorname{dom} f$ follows directly from [25, Corollary I.2.4]. In turn, the last statement of the theorem follows from the fact that if the subdifferential of f is bounded on a set Q, then f is Lipschitz continuous on this set (see [55, Theorem 2.4.13] and [44, Theorem 24.7]). Let us now turn to the proof of the main statement of the theorem.

1. The definition of hypodifferential. By [25, Corollary I.2.5 and Proposition I.5.2] the function f is subdifferentiable on $\operatorname{int}\operatorname{dom} f$. Let $V(\cdot)$ be a selection of the restriction of the subdifferential mapping $\partial f(\cdot)$ to the set Q. By the definition of subdifferential

$$f(y) - f(x) \geq \langle V(x), y - x\rangle \quad \forall y, x \in Q. \tag{23}$$

Note that this inequality turns into equality, if $y = x$. Therefore

$$f(y) = \max_{x \in Q} \big(f(x) + \langle V(x), y - x\rangle\big) \quad \forall y \in Q.$$

Choose any $y \in Q$, $\Delta y \in X$ such that $y + \Delta y \in Q$. Then

$$\begin{aligned} f(y + \Delta y) - f(y) &= \max_{x \in Q} \Big(f(x) - f(y) + \langle V(x), y + \Delta y - x\rangle\Big) \\ &= \max_{(a, x^*) \in \underline{d} f(y)} (a + \langle x^*, \Delta y\rangle), \end{aligned} \tag{24}$$

where $\underline{d} f(y) = \operatorname{cl}^* D(y)$ and

$$D(y) = \operatorname{co}\Big\{\big(f(x) - f(y) + \langle V(x), y - x\rangle, V(x)\big) \in \mathbb{R} \times X^* \;\Big|\; x \in Q\Big\}.$$

Note that $\max\{a \mid (a, x^*) \in D(y)\} = 0$ by inequality (23), which obviously implies that $\max\{a \mid (a, x^*) \in \underline{d} f(y)\} = 0$. Let us check that the set $D(x)$ is bounded for any $x \in Q$. Then one can conclude that $\underline{d} f(y)$ is a weak* compact set and, therefore, $\underline{d} f(y)$ is a hypodifferential of f at y that is obviously consistent and exact on Q due to equality (24).

2. Boundedness of hypodifferential. Choose any $x \in Q$ and $(a, x^*) \in D(x)$. By definition one can find $n \in \mathbb{N}$, $z_i \in Q$, $i \in I := \{1, \ldots, n\}$, $V(z_i) \in \partial f(z_i)$, $i \in I$, and $\alpha_i \geq 0$, $i \in I$, such that

$$(a, x^*) = \sum_{i=1}^{n} \alpha_i \Big(f(z_i) - f(x) + \langle V(z_i), x - z_i\rangle, V(z_i)\Big), \quad \sum_{i=1}^{n} \alpha_i = 1. \tag{25}$$

By our assumptions the set Q is bounded and f is Lipschitz continuous on Q. Therefore, there exist $C_Q > 0$ and $C_f > 0$ such that $\|z\| \leq C_Q$ and $|f(z)| \leq C_f$ for all $z \in Q$.

Let $L > 0$ be a Lipschitz constant of f on Q. Then for any $z \in Q$ and $V \in \partial f(z)$ one has

$$\langle V, y - z\rangle \leq f(y) - f(z) \leq L\|y - z\| \quad \forall y \in Q.$$

Since Q is an open set, there exists $r > 0$ such that $B(z, r) \subset Q$. Hence with the use of the inequality above one gets that

$$\langle V, y\rangle \leq L\|y\| \quad \forall y \in B(0, r),$$

which obviously implies that $\|V\| \leq L$. Bearing in mind equality (25) one has

$$\begin{aligned}\|(a, x^*)\| &\leq \sum_{i=1}^{n} \alpha_i \Big(\big| f(z_i) - f(x) + \langle V(z_i), x - z_i\rangle \big| + \|V(z_i)\| \Big) \\ &\leq \sum_{i=1}^{n} \alpha_i \Big(2C_f + 2LC_Q + L \Big) = 2C_f + 2LC_Q + L\end{aligned}$$

for any $(a, x^*) \in D(x)$. Thus, the set $D(x)$ is bounded.

3. Lipschitz continuity of the hypodifferential. As was noted in the proof of Proposition 2.8, the reflexivity of X implies that the set $\underline{d}f(x) = \mathrm{cl}^*\, D(x)$ coincides with the closure $\mathrm{cl}\, D(x)$ of $D(x)$ in the strong (norm) topology. Moreover, one obviously has $d_{PH}(\mathrm{cl}\, D(x), \mathrm{cl}\, D(y)) = d_{PH}(D(x), D(y))$ for any $x, y \in Q$, which implies that it is sufficient to show that the set-valued map $D(\cdot)$ is Lipschitz continuous on Q.

Fix any $x, y \in Q$. Let $(a, x^*) \in D(x)$. Then by definition there exist $n \in \mathbb{N}$, $z_i \in Q$, $i \in I := \{1, \ldots, n\}$, $V(z_i) \in \partial f(z_i)$, $i \in I$, and $\alpha_i \geq 0$, $i \in I$, satisfying (25). Denote

$$(b, y^*) = \sum_{i=1}^{n} \alpha_i \Big(f(z_i) - f(y) + \langle V(z_i), y - z_i\rangle, V(z_i) \Big).$$

Clearly, $(b, y^*) \in D(y)$. Therefore

$$\begin{aligned}\mathrm{dist}((a, x^*), D(y)) &\leq \|(a, x^*) - (b, y^*)\| = \Big| \sum_{i=1}^{n} \alpha_i (f(x) - f(y) + \langle V(z_i), x - y\rangle \Big| \\ &= \Big| f(x) - f(y) + \langle \widehat{V}, x - y\rangle \Big| \leq |f(x) - f(y)| + \|\widehat{V}\|\|x - y\|,\end{aligned}$$

where $\widehat{V} = \sum_{i\in I} \alpha_i V(z_i)$. Let $L > 0$ be a Lipschitz constant of f on Q. Then, as was noted above, $\|V\| \leq L$ for any $V \in \partial f(z)$ and $z \in Q$. Consequently, one has

$$\mathrm{dist}((a, x^*), D(y)) \leq 2L\|x - y\| \quad \forall (a, x^*) \in D(x).$$

Taking the supremum over all $(a, x^*) \in D(x)$ and then repeating the same argument with x and y swapped one obtains that $d_{PH}(D(x), D(y)) \leq 2L\|x - y\|$ for any $x, y \in Q$. □

Corollary 3.36. *Let X be a finite dimensional space. Then for any bounded set $Q \subseteq \mathrm{int}\,\mathrm{dom}\, f$ there exists a hypodifferential mapping of f that is defined, consistent, exact, and has the Lipschitz property on Q.*

Corollary 3.37. *The function f has an exact hypodifferential mapping on an open bounded set $Q \subseteq \mathrm{int}\,\mathrm{dom}\, f$ if and only if f is Lipschitz continuous on this set.*

Although Theorem 3.35 guarantees the existence of a consistent hypodifferential having the Lipschitz property under very mild assumptions, the hypodifferential constructed in the proof of this theorem is of theoretical value only. From the more practical perspective it is important to know how to compute such hypodifferentials for particular nonsmooth convex functions. Let us show that the calculus rules presented in Subsection 3.2 preserve the Lipschitz property and, therefore, can be used to compute consistent hypodifferentials having the Lipschitz property in various particular cases.

Theorem 3.38. *Under the assumptions of Theorem* 3.15 *the following statements hold true:*

(1) *If for any $t \in T$ the hypodifferential $\underline{d}f_t$ is a Lipschitzian approximation of $f(\cdot, t)$ on Q with Lipschitz constant $L_t > 0$ and $L = \sup_{t\in T} L_t < +\infty$, then the hypodifferential* (7) *is a Lipschitzian approximation of $f(\cdot) = \sup_{t\in T} f(\cdot, t)$ on Q with Lipschitz constant L.*

(2) *Let either X be reflexive or the set $\operatorname{cl} D(x)$ defined in* (8) *be weak* closed for any $x \in Q$. Suppose that for any $t \in T$ the function $f(\cdot, t)$ is Lipschitz continuous on Q with Lipschitz constant $L_t > 0$ and the hypodifferential map $\underline{d}f_t$ is Lipschitz continuous on Q with Lipschitz constant $K_t > 0$.*
If $L = \sup_{t\in T} L_t < +\infty$ and $K = \sup_{t\in T} K_t < +\infty$, then the hypodifferential (7) *is Lipschitz continuous on Q with Lipschitz constant $2L + K$.*

Proof. We split the proof of the theorem into two parts corresponding to the two statements.

Part 1. Fix any $x, y \in Q$. By Definition 3.28 one has

$$f(y,t) \leq f(x,t) + \max_{(a,x^*)\in \underline{d}f_t(x)} (a + \langle x^*, y - x\rangle) + \frac{L_t}{2}\|y - x\|^2.$$

Hence

$$\begin{aligned} f(y) - f(x) &= \sup_{t\in T} \big(f(y,t) - f(x)\big) \\ &\leq \sup_{t\in T} \Big(f(x,t) - f(x) + \max_{(a,x^*)\in \underline{d}f_t(x)} (a + \langle x^*, y - x\rangle) + \frac{L_t}{2}\|y - x\|^2\Big) \\ &\leq \sup_{t\in T} \Big(f(x,t) - f(x) + \max_{(a,x^*)\in \underline{d}f_t(x)} (a + \langle x^*, y - x\rangle\Big) + \frac{L}{2}\|y - x\|^2. \end{aligned}$$

One can readily verify that

$$\sup_{t\in T} \Big(f(x,t) - f(x) + \max_{(a,x^*)\in \underline{d}f_t(x)} (a + \langle x^*, y - x\rangle\Big) = \max_{(a,x^*)\in \underline{d}f(x)} (a + \langle x^*, y - x\rangle)$$

(see (7)). Therefore

$$f(y) - f(x) - \max_{(a,x^*)\in \underline{d}f(x)} (a + \langle x^*, y - x\rangle) \leq \frac{L}{2}\|y - x\|^2.$$

Applying the inequality

$$f(y,t) \geq f(x,t) + \max_{(a,x^*)\in \underline{d}f_t(x)} (a + \langle x^*, y - x\rangle) - \frac{L_t}{2}\|y - x\|^2,$$

that also follows from Definition 3.28, and repeating the same argument one gets

$$f(y) - f(x) - \max_{(a,x^*)\in \underline{d}f(x)} (a + \langle x^*, y - x\rangle) \geq -\frac{L}{2}\|y - x\|^2.$$

Consequently, $\underline{d}f$ is a Lipschitzian approximation of f on Q.

Part 2. If either X is reflexive or $\operatorname{cl} D(x)$ is weak* closed for any $x \in X$, then, as was noted several times above, $d_{PH}(\underline{d}f(x), \underline{d}f(y)) = d_{PH}(D(x), D(y))$ for any $x, y \in Q$. Therefore, it is sufficient to prove that the set-valued map $D(\cdot)$ is Lipschitz continuous on Q.

Fix any $x, y \in Q$ and $\varepsilon > 0$. Let $(a, x^*) \in D(x)$. Then by definition (see (8)) there exist $n \in \mathbb{N}$, $t_i \in T$, $i \in I := \{1, \ldots, n\}$, $(a_i, x_i^*) \in \underline{d}f_{t_i}(x)$, $i \in I$, and $\alpha_i \geq 0$, $i \in I$, such that

$$(a, x^*) = \sum_{i=1}^{n} \alpha_i (f(x, t_i) - f(x) + a_i, x_i^*), \quad \sum_{i=1}^{n} \alpha_i = 1.$$

By our assumption $d_{PH}(\underline{d}f_t(x), \underline{d}f_t(y)) \leq K_t\|x - y\|$. Hence by Proposition 2.1 there exist $(b_i, y_i^*) \in \underline{d}f_{t_i}(y)$, $i \in I$ such that $\|(a_i, x_i^*) - (b_i, y_i^*)\| \leq K_{t_i}\|x - y\| + \varepsilon$. Define

$$(b, y^*) = \sum_{i=1}^{n} \alpha_i (f(y, t_i) - f(y) + b_i, y_i^*).$$

Clearly, $(b, y^*) \in D(y)$. Therefore

$$\begin{aligned}
\operatorname{dist}((a, x^*), D(y)) &\leq \|(a, x^*) - (b, y^*)\| \\
&\leq \sum_{i=1}^{n} \alpha_i \Big(|f(x, t_i) - f(y, t_i)| + |f(x) - f(y)| + \|(a_i, x_i^*) - (b_i, y_i^*)\| \Big) \\
&\leq \sum_{i=1}^{n} \alpha_i \Big(L_t\|x - y\| + L\|x - y\| + K_{t_i}\|x - y\| + \varepsilon \Big) \leq (2L + K)\|x - y\| + \varepsilon
\end{aligned}$$

(here we used the fact that the function $f(\cdot) = \sup_{t\in T} f(\cdot, t)$ is Lipschitz continuous on Q with Lipschitz constant $L = \sup_{t\in T} L_t$). Since $(a, x^*) \in D(x)$ and $\varepsilon > 0$ were chosen arbitrarily, one obtains

$$\sup_{(a,x^*)\in D(x)} \operatorname{dist}((a, x^*), D(y)) \leq (2L + K)\|x - y\|.$$

Swapping x and y one can conclude that $d_{PH}(D(x), D(y)) \leq (2L+K)\|x-y\|$, which completes the proof. □

Theorem 3.39. *Let the assumptions of Theorem* 3.17 *be satisfied, the gradient* $\nabla g(\cdot)$ *of* g *be defined and Lipschitz continuous with Lipschitz constant* $L_g > 0$ *on the set* $\operatorname{co} f(Q)$*, and there exist* $C > 0$ *such that* $\|\nabla g(f(x))\| \leq C$ *for any* $x \in Q$*. Then the following statements hold true:*

(1) *If for any* $i \in I$ *the hypodifferential map* $\underline{d}f_i$ *is a Lipschitzian approximation of* f_i *on* Q *with Lipschitz constant* L_i *and* f_i *is Lipschitz continuous on* Q *with Lipschitz constant* K_i*, then the hypodifferential map* $\underline{d}h$ *(see* (16)*) is a Lipschitzian approximation of the function* $h(\cdot) = g(f(\cdot))$ *with Lipschitz constant* $L = C\sum_{i\in I} L_i + 2L_g(\sum_{i\in I} K_i)^2$.

(2) *Suppose that for any $i \in I$ the hypodifferential map $\underline{d}f_i$ is Lipschitz continuous on Q with Lipschitz constant $L_i > 0$ and there exists $C_i > 0$ such that $\|(a, x^*)\| \leq C_i$ for all $(a, x^*) \in \underline{d}f_i(x)$ and $x \in Q$. Then the hypodifferential map $\underline{d}h$ is Lipschitz continuous on Q with Lipschitz constant $l = C\sum_{i\in I} L_i + L_g \sum_{i\in I} C_i$.*

Proof. We split the proof of the theorem into two parts corresponding to the two statements.

Part 1. Fix any $x, y \in Q$. By the mean value theorem there exists $z \in \operatorname{co}\{f(x), f(y)\}$ such that

$$\begin{aligned} h(y) - h(x) &= g(f(y)) - g(f(x)) = \langle \nabla g(z), f(y) - f(x)\rangle \\ &= \langle \nabla g(f(x)), f(y) - f(x)\rangle + \langle \nabla g(z) - \nabla g(f(x)), f(y) - f(x)\rangle. \end{aligned}$$

Hence applying the fact that $\underline{d}f_i$ is a Lipschitzian approximation of f_i one gets

$$\begin{aligned} h(y) - h(x) \leq \sum_{i=1}^{n} \frac{\partial g(f(x))}{\partial y_i}\Big(&\max_{(a_i, x_i^*)\in \underline{d}f_i(x)} (a_i + \langle x_i^*, y - x\rangle) \\ &+ \frac{L_i}{2}\|y - x\|^2\Big) + \langle \nabla g(z) - \nabla g(f(x)), f(y) - f(x)\rangle \end{aligned}$$

(note that $\nabla g(f(x)) \geq 0$, since g is a non-decreasing function by the assumptions of Theorem 3.17). Bearing in mind the assumptions of the theorem and the definition of $\underline{d}h(x)$ (see (16))), one obtains

$$\begin{aligned} h(y) - h(x) \leq \max_{(a,x^*)\in \underline{d}h(x)} &(a + \langle x^*, y - x\rangle) + C\Big(\sum_{i=1}^{n} \frac{L_i}{2}\Big)\|y - x\|^2 \\ &+ L_g|z - f(x)| \cdot |f(y) - f(x)| \leq \max_{(a,x^*)\in \underline{d}h(x)} (a + \langle x^*, y - x\rangle) \\ &+ C\sum_{i=1}^{n} \frac{L_i}{2}\|y - x\|^2 + L_g\Big(\sum_{i=1}^{n} K_i\Big)^2\|y - x\|^2 \end{aligned}$$

(here we used the fact that $\|z - f(x)\| \leq \|f(y) - f(x)\|$, since $z \in \operatorname{co}\{f(x), f(y)\}$). Arguing in the same way and one can readily check that

$$h(y) - h(x) \geq \max_{(a,x^*)\in \underline{d}h(x)} (a + \langle x^*, y - x\rangle) - \Big(C\sum_{i=1}^{n} \frac{L_i}{2} + L_g(\sum_{i=1}^{n} K_i)^2\Big)\|y - x\|^2,$$

which implies the required result.

Part 2. Fix any $x, y \in Q$. Let $(a, x^*) \in \underline{d}h(x)$. Then by definition (see (16)) for any $i \in I$ there exists $(a_i, x_i^*) \in \underline{d}f_i(x)$ such that

$$(a, x^*) = \sum_{i=1}^{n} \frac{\partial g(f(x))}{\partial y_i}(a_i, x_i^*).$$

Choose any $\varepsilon > 0$. By our assumption $d_{PH}(\underline{d}f_i(x), \underline{d}f_i(y)) \leq L_i\|x - y\|$.

Therefore, by Proposition 2.1 for any $i \in I$ there exists $(b_i, y_i^*) \in \underline{d} f_i(y)$ such that

$$\|(a_i, x_i^*) - (b_i, y_i^*)\| \leq L_i \|x - y\| + \varepsilon.$$

Define
$$(b, y^*) = \sum_{i=1}^{n} \frac{\partial g(f(y))}{\partial y_i} (b_i, y_i^*).$$

Clearly, $(b, y^*) \in \underline{d} h(y)$, which implies that

$$\begin{aligned} \operatorname{dist}((a, x^*), \underline{d} h(y)) &\leq \|(a, x^*) - (b, y^*)\| \\ &= \Big\| \sum_{i=1}^{n} \frac{\partial g(f(x))}{\partial y_i} (a_i, x_i^*) - \sum_{i=1}^{n} \frac{\partial g(f(y))}{\partial y_i} (b_i, y_i^*) \Big\|. \end{aligned}$$

Adding and subtracting $\sum_{i \in I} \frac{\partial g(f(y))}{\partial y_i} (a_i, x_i^*)$ one gets

$$\begin{aligned} \operatorname{dist}((a, x^*), \underline{d} h(y)) &\leq \sum_{i=1}^{n} \left| \frac{\partial g(f(x))}{\partial y_i} - \frac{\partial g(f(y))}{\partial y_i} \right| \|(a_i, x_i^*)\| \\ &\quad + \sum_{i=1}^{n} \left| \frac{\partial g(f(y))}{\partial y_i} \right| \|(a_i, x_i^*) - (b_i, y_i^*)\| \\ &\leq L_g \Big(\sum_{i=1}^{n} C_i \Big) \|x - y\| + \sum_{i=1}^{n} C(L_i \|x - y\| + \varepsilon). \end{aligned}$$

Hence taking into account the facts that $(a, x^*) \in \underline{d} h(x)$ and $\varepsilon > 0$ were chosen arbitrarily one obtains

$$\sup_{(a, x^*) \in \underline{d} h(x)} \operatorname{dist}((a, x^*), \underline{d} h(y)) \leq \Big(L_g \sum_{i=1}^{n} C_i + C \sum_{i=1}^{n} L_i \Big) \|x - y\|.$$

Now, swapping x and y we arrive at the required result. □

Corollary 3.40. *Let $f_i \colon X \to \mathbb{R} \cup \{+\infty\}$, $i \in I = \{1, \ldots, n\}$, be closed convex functions and $\underline{d} f_i(\cdot)$ be their hypodifferential mappings defined on a set $Q \subseteq \bigcap_{i \in I} \operatorname{dom} f_i$. Let also $\lambda_i \geq 0$, $i \in I$, $f = \sum_{i \in I} \lambda_i f_i$, and $\underline{d} f = \sum_{i \in I} \lambda_i \underline{d} f_i$. The following statements hold true:*

(1) *If for any $i \in I$ the hypodifferential map $\underline{d} f_i$ is a Lipschitzian approximation of f_i on Q with Lipschitz constant $L_i > 0$, then the hypodifferential map $\underline{d} f$ is a Lipschitzian approximation of f on Q with Lipschitz constant $L = \sum_{i \in I} \lambda_i L_i$.*

(2) *If for any $i \in I$ the hypodifferential map $\underline{d} f_i$ is Lipschitz continuous on Q with Lipschitz constant $L_i > 0$, then the hypodifferential map $\underline{d} f$ is Lipschitz continuous on Q with Lipschitz constant $L = \sum_{i \in I} \lambda_i L_i$.*

The following result can be readily verified directly.

Proposition 3.41. *Under the assumptions of Proposition* 3.21 *the following statements hold true:*

(1) *If the hypodifferential* $\underline{d}f$ *is a Lipschitzian approximation of* f *on* Q *with Lipschitz constant* L*, then the hypodifferential* $\underline{d}g$ *(see* (17)*) is a Lipschitzian approximation of* g *on* $A^{-1}(Q-b)$ *with Lipschitz constant* $L\|A\|$*.*

(2) *If the hypodifferential map* $\underline{d}f$ *is Lipschitz continuous on* Q *with Lipschitz constant* L*, then the hypodifferential* $\underline{d}g$ *is Lilpschitz continuous on* $A^{-1}(Q-b)$ *with Lipschitz constant* $L\max\{1, \|A^*\|\}$*.*

4. Applications to nonsmooth convex optimization

In this section we consider applications of the theory of hypodifferentials of nonsmooth convex function to convex optimization. Our goal is to present several different versions of the method of hypodifferential descent for nonsmooth convex minimization and estimate their rate of convergence.

4.1. The method of hypodifferential descent

For the sake of simplicity, suppose that X is a real Hilbert space, although under some additional assumptions the results presented below can be partially extended to the case of general Banach spaces. Let the space $\mathbb{R}\times X$ be equipped with the norm $\|(a,x)\| = \sqrt{|a|^2+\|x\|^2}$ for any $(a,x)\in\mathbb{R}\times X$ and $f\colon X\to\mathbb{R}$ be a closed convex function. Suppose that a hypodifferential map $\underline{d}f(\cdot)$ of this function defined on X is known. Recall that in the case of Hilbert spaces hypodifferential is a convex weakly compact (or, equivalently, convex, bounded, and closed) subset of the space $\mathbb{R}\times X$ (see Remark 3.3).

Consider the following optimization problem:

$$\min_{x\in X} \ f(x).$$

Let f_* be the optimal value of the problem. Following the ideas of Demyanov [10] one can propose a method for minimizing the function f called *the method of hypodifferential descent*, whose theoretical scheme is given in Algorithm 1.

Algorithm 1: The method of hypodifferential descent

Initialisation. Choose an initial point $x_0\in X$ and put $k=0$.

Step 1. Compute $\underline{d}f(x_k)$. Find an optimal solution (a_k,v_k) of the problem

$$\min \ a^2+\|v\|^2 \quad \text{subject to } (a,v)\in\underline{d}f(x_k). \tag{26}$$

Step 2. Choose a step size $\alpha_k>0$ and put $x_{k+1}=x_k-\alpha_k v_k$.

Step 3. Check a **stopping criterion**. If it is satisfied, **Stop**. Otherwise, put $k\leftarrow k+1$ and go to **Step 1**.

Note that if the hypodifferential $\underline{d}f(x_k)$ is a polytope, then the subproblem on Step 2 of the method of hypodifferential descent is the problem of finding an element of a polytope with the smallest norm for which there exists many efficient methods [21, 29, 52, 2]. It should also be mentioned that one can use the following inequalities

$$|f(x_{k+1}) - f(x_k)| \leq \varepsilon \quad \text{and/or} \quad \operatorname{dist}\big(0, \underline{d}f(x_k)\big) = \sqrt{a_k^2 + \|v_k\|^2} \leq \varepsilon$$

with small $\varepsilon > 0$ as a stopping criterion for the method of hypodifferential descent.

The rate of convergence $\mathcal{O}(1/k)$ of the method of hypodifferential with the step sizes α_k chosen according to the Armijo's rule was proved in [18]. Here we will prove that the same rate of convergence holds true for the version of the method with constant step size and provide a rigorous justification for the stopping criteria discussed above. Let us note that although the proof of this result is similar to the proof of [18, Theorem 3.5], there are certain significant differences between them (e.g. the fact that one needs find an upper estimate for the constant step size ensuring that the sequence $\{f(x_k)\}$ is decreasing).

Theorem 4.1. *Let the sublevel set $S(x_0) = \{x \in X \mid f(x) \leq f(x_0)\}$ be bounded. Suppose also that the hypodifferential mapping $\underline{d}f(\cdot)$ is consistent and bounded on the set $S(x_0)$ and there exists $\varepsilon > 0$ such that $\underline{d}f$ is a Lipschitzian approximation of f on the set $S_\varepsilon(x_0) = \{x \in X \mid \operatorname{dist}(x, S(x_0)) \leq \varepsilon\}$ with Lipschitz constant $L > 0$. Then there exists $\widehat{\alpha}$ such that for any $\alpha \in (0, \widehat{\alpha})$ the method of hypodifferential descent with $\alpha_k \equiv \alpha$ generates sequence $\{x_k\}$ satisfying the following conditions:*

(1) *the sequence $\{f(x_k)\}$ is strictly decreasing and converges to the global minimum of f (in particular, $|f(x_{k+1}) - f(x_k)| \to 0$ as $k \to \infty$);*

(2) *$\sum_{i=0}^{\infty} \|(a_k, v_k)\|^2 < +\infty$ and $\operatorname{dist}(0, \underline{d}f(x_k)) \to 0$ as $k \to \infty$;*

(3) *for all $k \in \mathbb{N}$ the following inequality holds true:*

$$f(x_k) - f_* \leq \frac{(f(x_0) - f_*)(1+R)^2}{(1+R)^2 + (f(x_0) - f_*)\left(\alpha - \frac{L\alpha^2}{2}\right)k} = \mathcal{O}\left(\frac{1}{k}\right) \tag{27}$$

where $R = \sup\{\|y - x_*\| \mid y \in S(x_0)\}$, x_* is any point of global minimum of f, and $f_* = f(x_*)$.

Moreover, one can define $\widehat{\alpha} = \min\{2/L, \min\{1, \varepsilon\}/C\}$, where $C > 0$ is such that $|a| \leq C$ and $\|v\| \leq C$ for any $(a, v) \in \underline{d}f(x)$ and $x \in S(x_0)$.

Proof. Note that under the assumptions of the theorem the function f attains a global minimum, since X is a Hilbert space and the sublevel set $S(x_0)$ is bounded.

Let $\{x_k\}$ be the sequence generated by the method of hypodifferential descent with $\alpha_k \equiv \alpha$ for some $\alpha > 0$. Denote $\Phi_k(y) = \max_{(a,v) \in \underline{d}f(x_k)}(a + \langle v, y\rangle)$. For the sake of convenience, we divide the proof of the theorem into several parts.

Part 1. Recall that (a_k, v_k) is an optimal solution of subproblem (26). Applying the standard optimality conditions to this subproblem one gets

$$a_k(a - a_k) + \langle v_k, v - v_k\rangle \geq 0 \quad \forall (a, v) \in \underline{d}f(x_k). \tag{28}$$

If $a_k = 0$, then bearing in mind the fact that $a \leq 0$ for any $(a, v) \in \underline{d}f(x_k)$ one obtains

$$\Phi_k(-v_k) \leq \max_{(a,v)\in \underline{d}f(x_k)} \langle v, -v_k \rangle \leq -\|v_k\|^2 = -\|(a_k, v_k)\|^2.$$

In turn, if $a_k < 0$, then dividing (28) by a_k and taking the maximum over all $(a, v) \in \underline{d}f(x_k)$ one gets

$$\Phi_k\left(\frac{1}{a_k} v_k\right) = \max_{(a,v)\in \underline{d}f(x_k)} \left(a + \frac{1}{a_k}\langle v, v_k\rangle\right) \leq \frac{1}{a_k}\|(a_k, v_k)\|^2.$$

Thus, in either case one has $\Phi_k(-\beta_k v_k) \leq -\beta_k \|(a_k, v_k)\|^2$, where $\beta_k = 1/|a_k|$, if $a_k < 0$, and $\beta_k = 1$, otherwise. Hence taking into account the fact that by the definition of hypodifferential $\Phi_k(0) = 0$ one obtains that

$$\Phi_k(-t\beta_k v_k) \leq t\Phi_k(-\beta_k v_k) \leq -t\beta_k \|(a_k, v_k)\|^2 \quad \forall t \in [0, 1],$$

thanks to the convexity of Φ_k, or, equivalently,

$$\Phi_k(-tv_k) \leq -t\|(a_k, v_k)\|^2 \quad \forall t \in [0, \beta_k] \tag{29}$$

for any $k \in \mathbb{N}$.

Part 2. Let us show that there exists $\widehat{\alpha} > 0$ such that for any $\alpha \in (0, \widehat{\alpha}]$ one has $\{x_k\} \subset S(x_0)$ and the sequence $\{f(x_k)\}$ is strictly decreasing. Indeed, by our assumption the hypodifferential map $\underline{d}f(\cdot)$ is bounded on $S(x_0)$. Therefore, there exists $C > 0$ such that $|a| \leq C$ and $\|v\| \leq C$ for any $(a, v) \in \underline{d}f(x)$ and $x \in S(x_0)$. Define

$$\widehat{\alpha} = \min\left\{\frac{2}{L}, \frac{\min\{1, \varepsilon\}}{C}\right\} \tag{30}$$

and let $\alpha \in (0, \widehat{\alpha})$.

Suppose that $x_k \in S(x_0)$ for some $k \in \mathbb{N}$. Let us check that $x_{k+1} \in S(x_0)$ and $f(x_{k+1}) < f(x_k)$. Indeed, since $x_k \in S(x_0)$, one has $|a_k| \leq C$ or, equivalently, $\beta_k \geq 1/C$ (one can suppose that $C \geq 1$), which implies that $\alpha_k = \alpha \leq \beta_k$. Moreover, from the definition of $\widehat{\alpha}$ it follows that $t\|v_k\| \leq \varepsilon$ for any $t \in [0, \widehat{\alpha}]$. Consequently, $x_{k+1} := x_k - \alpha v_k \in S_\varepsilon(x_0)$. Therefore, applying inequality (29) and the fact that the hypodiferential map $\underline{d}f(\cdot)$ is a Lipschitzian approximation of f on $S_\varepsilon(x_0)$ one gets that

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \max_{(a,v)\in \underline{d}f(x_k)} (a + \langle v, x_{k+1} - x_k\rangle) + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) - \alpha\|(a_k, v_k)\|^2 + \frac{L\alpha^2}{2}\|v_k\|^2 \leq f(x_k) - \left(\alpha - \frac{L\alpha^2}{2}\right)\|(a_k, v_k)\|^2. \end{aligned}$$

Hence taking into account the fact that $\alpha < 2/L$ (see (30)) one can conclude that $f(x_{k+1}) < f(x_k)$ and $x_{k+1} \in S(x_0)$. Thus, in particular, the sequence $\{f(x_k)\}$ converges.

Part 3. Denote $\Delta_k = f(x_k) - f_*$. As we have shown above, one has

$$\Delta_{k+1} - \Delta_k \leq -\eta\|(a_k, v_k)\|^2 \quad \forall k \in \mathbb{N}, \quad \eta = \alpha - \frac{L\alpha^2}{2} > 0. \tag{31}$$

Hence, in particular, $\Delta_{k+1} < \Delta_k$ and

$$f(x_{k+1}) - f(x_0) = \sum_{i=0}^{k} (\Delta_{i+1} - \Delta_i) \leq -\eta \sum_{i=0}^{k} \|(a_k, v_k)\|^2 \quad \forall k \in \mathbb{N},$$

which implies that

$$\sum_{i=0}^{\infty} \|(a_k, v_k)\|^2 \leq \frac{1}{\eta}\Big(f(x_0) - \lim_{k\to\infty} f(x_k)\Big) \leq \frac{1}{\eta}(f(x_0) - f_*).$$

Thus, the second statement of the theorem holds true.

Recall that the hypodifferential map $\underline{d}f(\cdot)$ is consistent and $(a_k, v_k) \in \underline{d}f(x_k)$. Therefore

$$\Delta_k \leq -a_k + \langle v_k, x_k - x_* \rangle \leq \|(a_k, v_k)\|\big(1 + \|x_k - x_*\|\big) \leq (1 + R)\|(a_k, v_k)\| \quad (32)$$

for any $k \in \mathbb{N}$. Combining this inequality with (31) one obtains

$$\Delta_{k+1} \leq \Delta_k - \frac{\eta}{(1+R)^2}\Delta_k^2 \quad \forall k \in \mathbb{N}.$$

Dividing this inequality by $\Delta_{k+1} \cdot \Delta_k$ and rearranging the terms one gets

$$\frac{1}{\Delta_{k+1}} \geq \frac{1}{\Delta_k} + \frac{\eta}{(1+R)^2}\frac{\Delta_k}{\Delta_{k+1}} \geq \frac{1}{\Delta_k} + \frac{\eta}{(1+R)^2} \quad \forall k \in \mathbb{N}$$

(here we used the fact that $\Delta_{k+1} < \Delta_k$, that is, $\frac{\Delta_k}{\Delta_{k+1}} > 1$). Summing up these inequalities one finally obtains

$$\frac{1}{\Delta_{k+1}} \geq \frac{1}{\Delta_0} + \frac{\eta}{(1+R)^2}(k+1) \quad \forall k \in \mathbb{N},$$

which implies that the estimate (27) holds true. □

Remark 4.2. Note that the assumptions of the previous theorem are not restrictive in the finite dimensional case, since by Corollary 3.36 a hypodifferential mapping satisfying these assumptions exists for any nonsmooth function $f\colon X \to \mathbb{R}$ having the bounded sublevel set $S(x_0)$. □

Let us show that if the hypodifferential map $\underline{d}f(\cdot)$ is *exact* (global), then one can use the first component a_k of an optimal solution of subproblem (26) on Step 1 of the method of hypodifferential descent to determine a step size that ensures the linear rate of convergence of the method.

Theorem 4.3. *Let the sublevel set $S(x_0)$ be bounded, the hypodifferential map $\underline{d}f(\cdot)$ be exact on $S_\varepsilon(x_0)$ for some $\varepsilon > 0$, and $\{x_k\}$ be the sequence generated by the method of hypodifferential descent with $\alpha_k = 1/a_k$, if $a_k < 0$, and $\alpha_k = 1$, otherwise. Then the following statements hold true:*

(1) *$a_k < 0$ for any $k \in \mathbb{N}$;*

(2) *there exists $\gamma \in (0,1)$ and $C > 0$ such that $\operatorname{dist}(0, \underline{d}f(x_k)) \leq C\gamma^k$ for all $k \in \mathbb{N}$;*

(3) *there exists $q \in (0,1)$ such that $f(x_k) - f_* \leq (f(x_0) - f_*))q^k$ for any $k \in \mathbb{N}$.*

Proof. We use the same notation as in the proof of Theorem 4.1. For the sake of convenience, we divide the proof of this theorem into several parts.

Part 1. Let us show that if $x_k \in S(x_0)$, then $a_k < 0$. Suppose by contradiction that $x_k \in S(x_0)$, but $a_k = 0$ for some $k \in \mathbb{N}$. Then from the optimality conditions (28) and the fact that the hypodifferential map $\underline{d}f(\cdot)$ is exact on $S_\varepsilon(x_0)$ it follows that

$$\begin{aligned} f(x_k - \alpha v_k) - f(x_k) = \Phi_k(-\alpha v_k) &= \max_{(a,v)\in \underline{d}f(x_k)} (a + \langle v, -\alpha v_k\rangle) \\ &\leq \alpha \max_{(a,v)\in \underline{d}f(x_k)} \langle v, -v_k\rangle \leq -\alpha \|v_k\|^2 \end{aligned} \tag{33}$$

for any $\alpha \geq 0$ such that $x_k(\alpha) := x_k - \alpha v_k \in S_\varepsilon(x_0)$. Let us show that $x_k(\alpha) \in S_\varepsilon(x_0)$ for any $\alpha \geq 0$. Then $f(x_k(\alpha)) \to -\infty$ as $\alpha \to +\infty$, which contradicts the fact that the function f attains a global minimum, since its sublevel set $S(x_0)$ is bounded.

Indeed, note that $x_k(\alpha) \in S_\varepsilon(x_0)$ for any $\alpha \in [0, \varepsilon/\|v_k\|]$. With the use of the fact that the sublevel set $S(x_0)$ is convex one can easily check that the set $S_\varepsilon(x_0)$ is convex as well. Therefore, either the ray $\{x_k - \alpha v_k \mid \alpha \geq 0\}$ is contained in $S_\varepsilon(x_0)$ or there exists $\overline{\alpha} > 0$ such that

$$\left\{x_k - \alpha v_k \;\middle|\; \alpha \in [0, \overline{\alpha}]\right\} \subset S_\varepsilon(x_0), \quad x_k - \alpha v_k \notin S_\varepsilon(x_0) \quad \forall \alpha > \overline{\alpha}.$$

Note, however, that in the latter case one has $f(x_k(\overline{\alpha})) < f(x_k) \leq f(x_0)$ due to inequality (33). Consequently, there exists $\delta > 0$ such that $f(x_k(\alpha)) \leq f(x_0)$ for any $\alpha \in [\overline{\alpha}, \overline{\alpha} + \delta]$ by virtue of the fact that f is continuous by [25, Corollary I.2.5]. Thus, the second case is impossible and we can conclude that $a_k < 0$, whenever $x_k \in S(x_0)$.

Part 2. Let us show that $\{x_k\} \subset S(x_0)$ and the sequence $\{f(x_k)\}$ is strictly decreasing. Indeed, let $x_k \in S(x_0)$ for some $k \in \mathbb{N}$. Then $a_k < 0$ and from the optimality conditions (28) and the convexity of Φ_k it follows that

$$\Phi_k\left(\frac{\alpha}{a_k} v_k\right) \leq \alpha \Phi_k\left(\frac{1}{a_k} v_k\right) \leq \frac{\alpha}{a_k} \|(a_k, v_k)\|^2 < 0$$

for any $\alpha \in [0, 1]$. Therefore applying the fact that the hypodifferential map $\underline{d}f(\cdot)$ is exact on $S_\varepsilon(x_0)$ one obtains that

$$f\left(x_k + \frac{\alpha}{a_k} v_k\right) = f(x_k) + \Phi_k\left(\frac{\alpha}{a_k} v_k\right) \leq f(x_k) + \frac{\alpha}{a_k} \|(a_k, v_k)\|^2$$

for any $\alpha \in [0, 1]$ such that $x_k + (\alpha/a_k) v_k \in S_\varepsilon(x_k)$. Arguing in the same way as in the first part of the proof one can show that this inclusion is satisfied for any $\alpha \in [0, 1]$. Consequently, putting $\alpha = 1$ one gets

$$f(x_{k+1}) \leq f(x_k) + \frac{1}{a_k} \|(a_k, v_k)\|^2 < f(x_k). \tag{34}$$

Therefore, $\{x_k\} \subset S(x_0)$ and the sequence $\{f(x_k)\}$ is strictly decreasing.

Part 3. From inequality (34) it follows that for any $k \in \mathbb{N}$:

$$\Delta_{k+1} - \Delta_k \leq -\frac{1}{|a_k|} \|(a_k, v_k)\|^2 \leq -\frac{1}{\|(a_k, v_k)\|} \|(a_k, v_k)\|^2 = -\|(a_k, v_k)\|. \tag{35}$$

Hence applying inequality (32) one obtains

$$\Delta_{k+1} \leq \left(1 - \frac{1}{1+R}\right) \Delta_k \quad \forall k \in \mathbb{N},$$

which implies that $\Delta_{k+1} \leq \Delta_0 q^{k+1}$ for any $k \in \mathbb{N}$, where $q = 1 - \frac{1}{1+R} \in (0,1)$. Thus, the third statement of the theorem holds true. To prove the second statement of the theorem, note that from inequality (35) it follows that

$$\operatorname{dist}(0, \underline{d}f(x_k)) = \|(a_k, v_k)\| \leq \Delta_k - \Delta_{k+1} \leq \Delta_0 q^k - \Delta_0 q^{k+1} = \Delta_0 (1-q) q^k$$

for any $k \in \mathbb{N}$. Thus, the second statement of the theorem holds true with the values $C = \Delta_0(1-q)$ and $\gamma = q$. □

Remark 4.4. (i) If under the assumptions of the previous theorem the function f is polyhedral, then by [18, Theorem 4.7] the method of hypodifferential descent with $\alpha_k = 1/a_k$ converges to a point of global minimum of f in a finite number of steps.

(ii) The previous theorem remains to hold true if the step sizes α_k are defined via the exact line search:

$$f(x_k - \alpha_k v_k) = \min_{\alpha \geq 0} f(x_k - \alpha v_k).$$

Indeed, in this case one obviously has

$$\Delta_{k+1} - \Delta_k = \min_{\alpha \geq 0} f(x_k - \alpha v_k) - f(x_k) \leq f\left(x_k + \frac{1}{a_k} v_k\right) - f(x_k) \leq \frac{1}{a_k} \|(a_k, v_k)\|^2$$

(see (34)). Then arguing in the same way as in the third part of the proof of Theorem 4.3 one can prove the linear rate of convergence of the method. □

4.2. Accelerated proximal hypodifferential descent

Let us consider a different method for minimizing nonsmooth convex functions based on the use of hypodifferentials, which we call *the proximal hypodifferential descent method*. This method is related to various proximal point methods [49] and can be viewed as a generalisation of the Pschenichnyi-Pironneau-Polak algorithm for convex minimax problems [43, Algorithm 2.4.1] (see also [37, Section 2.3.3]) to the case of general nonsmooth convex functions, as well as a particular version of the quadratic regularisation of the method of codifferential descent for nonconvex nonsmooth functions [17]. Moreover, unlike the method of hypodifferential descent from the previous section, the proximal hypodifferential descent can be applied to problems of minimizing the function f over a closed convex set $Q \subseteq X$. A theoretical scheme of this method is given in Algorithm 2.

Let us give some comments on the proximal hypodifferential descent method. Firstly, note that the point z_k is correctly defined for any $k \in \mathbb{N}$. Indeed, it is easily seen that the function $\Psi_k(x) := \Phi_k(x - x_k) + 0.5\gamma\|x - x_k\|^2$, $x \in X$, is strongly convex, where, as above,

$$\Phi_k(x) = \max_{(a,v) \in \underline{d}f(x_k)} (a + \langle v, x \rangle) \quad \forall x \in X.$$

Algorithm 2: Proximal hypodifferential descent method

Initialisation. Choose an initial point $x_0 \in Q$, a paramter $\gamma > 0$, and put $k = 0$.

Step 1. Compute $\underline{d}f(x_k)$. Find an optimal solution z_k of the problem

$$\min_{x \in Q} \max_{(a,v) \in \underline{d}f(x_k)} (a + \langle v, x - x_k \rangle) + \frac{\gamma}{2}\|x - x_k\|^2. \tag{36}$$

Step 2. Choose a step size $\alpha_k \in (0, 1]$ and put $x_{k+1} = x_k + \alpha_k(z_k - x_k)$.

Step 3. Check a **stopping criterion**. If it is satisfied, **Stop**. Otherwise, put $k \leftarrow k + 1$ and go to **Step 1**.

Moreover, this function is coercive, since $|\Phi_k(x)| \leq C_k + C_k\|x\|$ for all $x \in X$ and some $C_k > 0$ due to the fact that hypodifferential is a bounded set. Therefore, the function $\Psi_k(\cdot)$ attain a global minimum on Q at a unique point z_k (see, e.g. [25, Proposition II.1.2]).

If the set Q is polyhedral (that is, it is defined by a finite number of linear equality and inequality constraints) and the hypodifferential $\underline{d}f(x_k)$ is a polytope, then the auxiliary subproblem (36) on Step 1 of the proximal hypodifferential descent method can obviously be rewritten as an equivalent convex quadratic programming problem. Let us also point out that the following inequalities

$$\|z_k - x_k\| \leq \varepsilon \quad \text{and/or} \quad f(x_k) - f(x_{k+1}) \leq \varepsilon \tag{37}$$

with small $\varepsilon > 0$ can be used as a stopping criterion for Algorithm 2.

The correctness of the stopping criteria (37), as well as global convergence of the proximal hypodifferential descent method can be proved in exactly the same way as the analogous results for QR-MCD method from [17, Section 4], since Algorithm 2 is a particular case of QR-MCD in the convex case. In turn, the rate of convergence of the proximal hypodifferential descent method can be estimated in the strongly convex case in exactly the same way as the rate of convergence of similar methods for convex minimax problems in [37, Theorem 2.3.4] and [43, Theorem 2.4.5]. For the sake of shortness, we do not present detailed formulations and proofs of these results here. Instead, we will analyse an accelerated version of Algorithm 2 designed with the use of Nesterov's acceleration technique based on estimating sequences [37, Sections 2.2 and 2.3].

For any $x \in Q$ let $z_\gamma(x)$ be an optimal solution of the problem

$$\min_{z \in Q} \max_{(a,v) \in \underline{d}f(x)} (a + \langle v, z - x \rangle) + \frac{\gamma}{2}\|z - x\|^2.$$

Let us obtain one useful estimate.

Lemma 4.5. *Let the hypodifferential map $\underline{d}f(\cdot)$ be consistent on X and a Lipschitzian approximation of f on X with Lipschitz constant $L > 0$. Then for any $\gamma \geq L$, $x \in Q$, and $y \in X$ one has*

$$f(x) \geq f(z_\gamma(y)) + \gamma\langle y - z_\gamma(y), x - y \rangle + \frac{\gamma}{2}\|y - z_\gamma(y)\|^2.$$

Proof. Fix any $y \in X$ and denote

$$\Phi(x) = \max_{(a,v) \in \underline{d}f(y)} (a + \langle v, x - y \rangle), \quad \Psi(x) = \Phi(x) + \frac{\gamma}{2}\|x - y\|^2 \quad \forall x \in X.$$

Clearly, the function Ψ is strongly convex with modulus $\mu = \gamma$. Therefore

$$\Psi(x) \geq \Psi(z_\gamma(y)) + \frac{\gamma}{2}\|x - z_\gamma(y)\|^2 \quad \forall x \in Q,$$

(see, e.g. [51, Proposition 4.8]). Hence for any $x \in Q$ one has

$$\begin{aligned}\Phi(x) &= \Psi(x) - \frac{\gamma}{2}\|x-y\|^2 \geq \Psi(z_\gamma(y)) + \frac{\gamma}{2}\Big(\|x - z_\gamma(y)\|^2 - \|x-y\|^2\Big)\\ &= \Psi(z_\gamma(y)) + \frac{\gamma}{2}\langle y - z_\gamma(y), 2x - z_\gamma(y) - y\rangle\\ &= \Psi(z_\gamma(y)) + \frac{\gamma}{2}\langle y - z_\gamma(y), 2(x-y) + y - z_\gamma(y)\rangle\\ &= \Psi(z_\gamma(y)) + \gamma\langle y - z_\gamma(y), x - y\rangle + \frac{\gamma}{2}\|y - z_\gamma(y)\|^2.\end{aligned}$$

From the facts that $\underline{d}f(\cdot)$ is a Lipschitzian approximation of f on X with Lipschitz constant $L > 0$ and $\gamma \geq L$ it follows that

$$f(z_\gamma(y)) \leq f(y) + \Phi(z_\gamma(y)) + \frac{L}{2}\|z_\gamma(y) - y\|^2 = f(y) + \Psi(z_\gamma(y)).$$

Hence taking into account the fact that the hypodifferential map $\underline{d}f(\cdot)$ is consistent one obtains

$$\begin{aligned}f(x) &\geq f(y) + \Phi(x)\\ &\geq f(y) + \Psi(z_\gamma(y)) + \gamma\langle y - z_\gamma(y), x - y\rangle + \frac{\gamma}{2}\|y - z_\gamma(y)\|^2\\ &\geq f(z_\gamma(y)) + \gamma\langle y - z_\gamma(y), x - y\rangle + \frac{\gamma}{2}\|y - z_\gamma(y)\|^2\end{aligned}$$

for any $x \in Q$. □

Suppose that the hypodifferential map $\underline{d}f(\cdot)$ is a Lipschitzian approximation of f on X with Lipschitz constant $L > 0$. Following the derivation of Nesterov's accelerated method for convex minimax problems [37, Section 2.3.3], choose some $x_0 \in Q$ and $\gamma_0 > 0$. Let $\{y_k\} \subset Q$ and $\{\alpha_k\} \subset (0,1)$ be some sequences. For any $k \in \mathbb{N}$ define

$$\begin{aligned}\phi_0(x) &= f(x_0) + \frac{\gamma_0}{2}\|x - x_0\|^2,\\ \phi_{k+1}(x) &= (1-\alpha_k)\phi_k(x) + \alpha_k\Big(f(z_L(y_k)) + \frac{L}{2}\|y_k - z_L(y_k)\|^2\Big)\\ &\quad + \alpha_k L\langle y_k - z_L(y_k), x - y_k\rangle.\end{aligned}$$

Since ϕ_0 is a quadratic function and ϕ_{k+1} is the sum of $(1-\alpha_k)\phi_k(x)$ and an affine function, the functions ϕ_k are quadratic for all $k \in \mathbb{N}$. Moreover, one can readily check that

$$\phi_k(x) = \phi_k^* + \frac{\gamma_k}{2}\|x - v_k\|^2 \quad \forall k \in \mathbb{N}, \tag{38}$$

where
$$v_0 = x_0, \quad \phi_0^* = f(x_0), \quad \gamma_{k+1} = (1-\alpha_k)\gamma_k, \tag{39}$$

$$v_{k+1} = v_k - \frac{\alpha_k L}{\gamma_{k+1}}(y_k - z_L(y_k)), \tag{40}$$

$$\begin{aligned}\phi_{k+1}^* &= (1-\alpha_k)\phi_k^* + \alpha_k\Big(f(z_L(y_k)) + \frac{L}{2}\|y_k - z_L(y_k)\|^2\Big)\\ &\quad - \frac{\alpha_k^2 L^2}{2\gamma_{k+1}}\|y_k - z_L(y_k)\|^2 + \alpha_k L\langle y_k - z_L(y_k), v_k - y_k\rangle\end{aligned} \tag{41}$$

for any $k \in \mathbb{N}$.

With the use of the above expression for the functions ϕ_k we can derive an accelerated version of the proximal hypodifferential descent method.

Lemma 4.6. *Let $\alpha_0 \in (0,1)$, $\gamma_0 = L\alpha_0^2$ and sequences $\{x_k\}$, $\{y_k\}$, and $\{\alpha_k\}_{k\geq 1}$ be chosen according to the following equalities:*

$$x_{k+1} = z_L(y_k), \quad L\alpha_k^2 = \gamma_{k+1}, \quad y_k = \alpha_k v_k + (1-\alpha_k)x_k. \tag{42}$$

Then for all $k \in \mathbb{N}$ one has

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2, \quad \alpha_k \in (0,1), \quad \phi_k^* \geq f(x_k).$$

Proof. The equality for α_{k+1} follows directy from equalities $L\alpha_k^2 = \gamma_{k+1}$ and (39). Note also that the equation $t^2 = (1-t)\alpha_k^2$ has a unique positive solution that lies in the interval $(0,1)$, since the function $g(t) = t^2 - \alpha_k^2(1-t)$ is strictly increasing on $[0,1]$ and satisfies the equalities $g(0) = -\alpha_k^2 < 0$ and $g(1) = 1$. Therefore, $\alpha_k \in (0,1)$ for all $k \in \mathbb{N}$.

Let us prove the inequality $\phi_k^* \geq f(x_k)$ by induction in k. Suppose that $\phi_k^* \geq f(x_k)$ for some $k \in \mathbb{N}$ (recall that $\phi_0^* = f(x_0)$). Applying Lemma 4.5 with $x = x_k$, $y = y_k$, and $\gamma = L$ one gets

$$f(x_k) \geq f(z_L(y_k)) + L\langle y_k - z_L(y_k), x_k - y_k\rangle + \frac{L}{2}\|y_k - z_L(y_k))\|^2.$$

Hence with the use of (41) and inequality $\phi_k^* \geq f(x_k)$ one obtains

$$\begin{aligned}
\phi_{k+1}^* &\geq (1-\alpha_k)f(x_k) + \alpha_k\left(f(z_L(y_k)) + \frac{L}{2}\|y_k - z_L(y_k)\|^2\right) \\
&\quad - \frac{\alpha_k^2 L^2}{2\gamma_{k+1}}\|y_k - z_L(y_k)\|^2 + \alpha_k L\langle y_k - z_L(y_k), v_k - y_k\rangle \\
&\geq f(z_L(y_k)) + \left(\frac{L}{2} - \frac{\alpha_k^2 L^2}{2\gamma_{k+1}}\right)\|y_k - z_L(y_k)\|^2 \\
&\quad + (1-\alpha_k)L\left\langle y_k - z_L(y_k), x_k - y_k + \frac{\alpha_k}{(1-\alpha_k)}(v_k - y_k)\right\rangle
\end{aligned}$$

Consequently, $\phi_{k+1}^* \geq f(x_{k+1})$ due to equalities (42). □

Observe that from (42) and (40) it follows that

$$\begin{aligned}
v_{k+1} &= v_k - \frac{\alpha_k L}{\gamma_{k+1}}(y_k - z_L(y_k)) = v_k - \frac{1}{\alpha_k}(y_k - x_{k+1}) \\
&= -\frac{1}{\alpha_k}(y_k - \alpha_k v_k) + \frac{1}{\alpha_k}x_{k+1} = -\frac{1}{\alpha_k}(1-\alpha_k)x_k + \frac{1}{\alpha_k}x_{k+1} \\
&= x_k + \frac{1}{\alpha_k}(x_{k+1} - x_k).
\end{aligned}$$

Hence with the use of (42) one gets

$$\begin{aligned}
y_{k+1} &= \alpha_{k+1}v_{k+1} + (1-\alpha_{k+1})x_{k+1} \\
&= \alpha_{k+1}x_k + \frac{\alpha_{k+1}}{\alpha_k}(x_{k+1} - x_k) + (1-\alpha_{k+1})x_{k+1} \\
&= x_{k+1} + \frac{\alpha_{k+1}(1-\alpha_k)}{\alpha_k}(x_{k+1} - x_k).
\end{aligned}$$

Thus, we arrive at the following method, which we call the *accelerated proximal hypodifferential descent method*, whose theoretical scheme is given in Algorithm 3.

Algorithm 3: Accelerated proximal hypodifferential descent method

---

Initialisation. Choose an initial point $x_0 \in Q$ and parameter $\alpha_0 \in (0,1)$. Set $y_0 = x_0$ and $k = 0$.

Step 1. Compute $\underline{d}f(y_k)$. Find an optimal solution x_{k+1} of the problem

$$\min_{x \in Q} \max_{(a,v) \in \underline{d}f(y_k)} (a + \langle v, x - y_k \rangle) + \frac{L}{2}\|x - y_k\|^2.$$

Step 2. Find a solution $\alpha_{k+1} \in (0,1)$ of the equation

$$\alpha_{k+1}^2 = (1 - \alpha_{k+1})\alpha_k^2$$

and put $$y_{k+1} = x_{k+1} + \frac{\alpha_{k+1}(1 - \alpha_k)}{\alpha_k}(x_{k+1} - x_k).$$

Set $k \leftarrow k + 1$ and go to **Step 1**.

The following theorem describes the rate of convergence of this method. Let us note that its proof largely repeats the way in which the rate of convergence of Nesterov's accelerated gradient descent method is estimated [37].

Theorem 4.7. *Suppose that f attains a global minimum on the set Q and the hypodifferential map $\underline{d}f(\cdot)$ is consistent on X and a Lipschitzian approximation of f on X with Lipschitz constant $L > 0$. Let $\{x_k\}$ be the sequence generated by the accelerated proximal hypodifferential descent method. Then for any $k \in \mathbb{N}$ one has*

$$f(x_k) - f_* \leq \frac{4L}{(2\sqrt{L} + k\sqrt{\gamma_0})^2}\left(f(x_0) - f_* + \frac{\gamma_0}{2}\|x_0 - x_*\|^2\right) = \mathcal{O}\left(\frac{1}{k^2}\right),$$

where $f_* = \min_{x \in Q} f(x_*)$ and x_* is a point of global minimum of f on Q.

Proof. Define $\lambda_0 = 1$ and $\lambda_{k+1} = (1 - \alpha_k)\lambda_k$ for all $k \in \mathbb{N}$. Let us check that

$$\phi_k(x) \leq (1 - \lambda_k)f(x) + \lambda_k \phi_0(x) \quad \forall x \in X,\ k \in \mathbb{N}. \tag{43}$$

Indeed, in the case $k = 0$ the inequality is obvious. Suppose that it is satisfied for some $k \in \mathbb{N}$. Recall that

$$\begin{aligned}\phi_{k+1}(x) = (1 - \alpha_k)\phi_k(x) + \alpha_k \left(f(z_L(y_k)) + \frac{L}{2}\|y_k - z_L(y_k)\|^2\right) \\ + \alpha_k L\langle y_k - z_L(y_k), x - y_k \rangle.\end{aligned}$$

Applying Lemma 4.5 with $y = y_k$ one gets that

$$\phi_{k+1}(x) \leq (1 - \alpha_k)\phi_k(x) + \alpha_k f(x) \quad \forall x \in X.$$

Hence with the use of inequality (43) one obtains for any $x \in X$ that

$$\begin{aligned}\phi_{k+1}(x) &\leq (1 - \alpha_k)\Big((1 - \lambda_k)f(x) + \lambda_k \phi_0(x)\Big) + \alpha_k f(x) \\ &= (1 - \lambda_{k+1})f(x) + \lambda_{k+1}\phi_0(x).\end{aligned}$$

Thus, by induction inequality (43) holds true for all $k \in \mathbb{N}$.

By the construction of the method one has $f(x_k) \leq \phi_k^*$. Therefore by applying inequality (43) one obtains that

$$f(x_k) \leq \phi_k^* = \min_{x \in X} \phi_k(x) \leq \min_{x \in Q} \phi_k(x) \leq \inf_{x \in Q} \Big((1-\lambda_k) f(x) + \lambda_k \phi_0(x) \Big)$$
$$\leq (1-\lambda_k) f(x_*) + \lambda_k \phi_0(x_*)$$

or, equivalently,

$$f(x_k) - f_* \leq \lambda_k (\phi_0(x_*) - f_*) = \lambda_k \left[f(x_0) - f_* + \frac{\gamma_0}{2} \|x_0 - x_*\|^2 \right] \tag{44}$$

for any $k \in \mathbb{N}$.

Let us estimate λ_k. First, note that $\gamma_k \geq \gamma_0 \lambda_k$ for all $k \in \mathbb{N}$. Indeed, $\gamma_0 = \gamma_0 \lambda_0$, since $\lambda_0 = 1$ by definition. Suppose that the inequality holds true for some $k \in \mathbb{N}$. Then

$$\gamma_{k+1} = (1-\alpha_k)\gamma_k \geq (1-\alpha_k)\gamma_0 \lambda_k = \gamma_0 \lambda_{k+1},$$

which by induction implies that $\gamma_k \geq \gamma_0 \lambda_k$ or, equivalently, $\lambda_k \leq \gamma_k / \gamma_0$ for all $k \in \mathbb{N}$.

For any $k \in \mathbb{N}$ one has

$$\frac{1}{\sqrt{\lambda_{k+1}}} - \frac{1}{\sqrt{\lambda_k}} = \frac{\sqrt{\lambda_k} - \sqrt{\lambda_{k+1}}}{\sqrt{\lambda_k \lambda_{k+1}}} = \frac{\lambda_k - \lambda_{k+1}}{(\sqrt{\lambda_k} + \sqrt{\lambda_{k+1}})\sqrt{\lambda_k \lambda_{k+1}}}.$$

Recall that $\lambda_{k+1} = (1-\alpha_k)\lambda_k$ and $\alpha_k \in (0,1)$. Therefore

$$\frac{1}{\sqrt{\lambda_{k+1}}} - \frac{1}{\sqrt{\lambda_k}} = \frac{\lambda_k - (1-\alpha_k)\lambda_k}{(\sqrt{\lambda_k} + \sqrt{\lambda_{k+1}})\sqrt{\lambda_k \lambda_{k+1}}}$$
$$= \frac{\alpha_k}{(1+\sqrt{1-\alpha_k})\sqrt{\lambda_{k+1}}} \geq \frac{\alpha_k}{2\sqrt{\lambda_{k+1}}}.$$

Hence applying the inequality $\lambda_{k+1} \leq \gamma_{k+1}/\gamma_0$ and the equality $L\alpha_k^2 = \gamma_{k+1}$ (see Lemma 4.6) one obtains that

$$\frac{1}{\sqrt{\lambda_{k+1}}} - \frac{1}{\sqrt{\lambda_k}} \geq \frac{\alpha_k}{2} \sqrt{\frac{\gamma_0}{\gamma_{k+1}}} = \frac{1}{2}\sqrt{\frac{\gamma_0}{L}}$$

for any $k \in \mathbb{N}$, which implies that

$$\frac{1}{\sqrt{\lambda_k}} \geq 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{L}} \quad \Longleftrightarrow \quad \lambda_k \leq \frac{4L}{(2\sqrt{L} + k\sqrt{\gamma_0})^2} \quad \forall k \in \mathbb{N}.$$

Combining this inequality with (44) one obtains the required result. □

5. Conclusions

We presented a general theory of hypodifferentials of nonsmooth convex functions and its applications to nonsmooth convex optimization. The key difference between subdifferentials and hypodifferentials is the fact that, in a sense, hypodifferentials much closer resemble the gradient of a smooth convex function in its properties than

subdifferentials. In particular, hypodifferentials possess many important properties of the gradient that are widely used in convex optimization (such as continuity and Lipschitz continuity, the property of being a Lipschitzian approximation, etc.), that subdifferentials do not have. This important feature of hypodifferentials allowed us to extend some of the results on smooth convex functions/convex optimization methods to the general nonsmooth setting, which is apparently impossible to do in the context of the theory of subdifferentials. For example, as we have shown, hypodifferentials can be used to generalize accelerated gradient methods to the case of nonsmooth convex functions. Thus, from the theoretical point of view, the theory of hypodifferentials provides one with new useful tools for studying nonsmooth convex functions and problems.

Although the calculus rules and examples presented in the paper show how hypodifferentials of nonsmooth convex functions can be computed in particular cases, it must be said that hypodifferentials are often much more cumbersome and hard to deal with from the computational point of view than subdifferentials/subgradients. They provide significantly more information about the function than subdifferentials/subgradients at the expense of much higher computational cost and memory consumption. In particular, the higher rate of convergence of convex optimization methods based on hypodifferentials in comparison with standard nonsmooth convex optimization methods is achieved by a significant increase of complexity of each iteration and each oracle call, which makes them inapplicable to large scale problems. However, in some applications (especially in the case of problems of moderate dimensions), the higher rate of convergence of accelerated hypodifferential descent methods might compensate for the significantly higher complexity of each iteration to make such methods competitive in comparison with classical subgradient/bundle methods.

Acknowledgements. The author is sincerely grateful to the anonymous reviewer for carefully reading the paper, providing useful comments and suggestions, as well as for helping to clarify the formulation of Corollary 3.5.

References

[1] F. Atenas, C. Sagastizábal, P. J. S. Silva, M. Solodov: *A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods*, SIAM J. Optim. 33/1 (2023) 89–115.

[2] A. Barbero, J. López, J. R. Dorronsoro: *An accelerated MDM algorithm for SVM training*, in: *Advances in Computational Intelligence and Learning*, Proc. of the ESANN 2008 conference (2008) 421–426.

[3] J. Bolte, T. P. Nguyen, J. Peypouquet, B. W. Suter: *From error bounds to the complexity of first-order descent methods for convex functions*, Math. Program. 165 (2017) 471–507.

[4] J. F. Bonnnans, A. Shapiro: *Perturbation Analysis of Optimization Problems*, Springer, New York (2000).

[5] S. Bubeck: *Convex optimization: algorithms and complexity*, Found. Trends Mach. Learn. 8/3-4 (2015) 231–357.

[6] M. Danilova, P. Dvurechensky, A. Gasnikov, E. Gorbunov, S. Guminov, D. Kamzolov, I. Shibaev: *Recent theoretical advances in non-convex optimization*, in: *High-Dimensional Optimization and Probability*, A. Nikeghbali et al. (eds.), Springer, Cham (2022) 79–163.

[7] V. F. Demyanov; *Continuous generalized gradients for nonsmooth functions*, in: *Optimization, Parallel Processing and Applications*, A. Kurzhanski et al. (eds.), Springer, Berlin (1988) 24–27.

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

[9] V. F. Demyanov: *Smoothness of nonsmooth functions*, in: *Nonsmooth Optimization and Related Topics*, F. H. Clarke et al. (eds.), Springer, Boston (1989) 79–88.

[10] V. F. Demyanov, A. M. Rubinov: *Constructive Nonsmooth Analysis*, Peter Lang, Frankfurt (1995).

[11] V. F. Demyanov, A. M. Rubinov (eds.): *Quasidifferentiability and Related Topics*, Kluwer, Dordrecht (2000),

[12] V. F. Demyanov, G. E. Stavroulakis, L. N. Polyakova, P. D. Panagiotopoulos: *Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economics*, Kluwer, Dordrecht (1996).

[13] V. F. Demyanov, G. Sh. Tamasyan: *Exact penalty functions in isoperimetric problems*, Optim. 60 (2011) 153–177.

[14] V. F. Demyanov, G. Sh. Tamasyan: *Direct methods in the parametric moving boundary variational problem*, Numer. Funct. Anal. Optim. 35 (2014) 932–961.

[15] M. V. Dolgopolik: *Codifferential calculus in normed spaces*, J. Math. Sci. 173 (2011) 441–462.

[16] M. Dolgopolik: *Nonsmooth problems of calculus of variations via codifferentiation*, ESAIM Control Optim. Calc. Var. 20/4 (2014) 1153–1180.

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

[18] M. V. Dolgopolik: *The method of codifferential descent for convex and global piecewise affine optimization*, Optim. Methods Software 35/6 (2020) 1191–1222.

[19] M. V. Dolgopolik: *New global optimality conditions for nonsmooth DC optimization problems*, J. Glob. Optim. 76 (2020) 25–55.

[20] M. Dolgopolik: *Constrained nonsmooth problems of the calculus of variations*, ESAIM Control Optim. Calc. Var. 27 (2021) 1–35.

[21] M. V. Dolgopolik: *An efficient acceleration technique for methods for finding the nearest point in a polytope and computing the distance between two polytopes*, arXiv: 2205.04553 (2022).

[22] Y. Drori, A. B. Taylor: *Efficient first-order methods for convex minimization: a constructive approach*, Math. Program. 184 (2020) 183–220.

[23] Y. Drori, M. Teboulle: *An optimal variant of Kelley's cutting-plane method*, Math. Program. 160 (2016) 321–351.

[24] P. Dvurechensky, S. Shtern, M. Staudigl: *First-order methods for convex optimization*, EURO J. Comput. Optim. 9 (2021), art. no. 100015, 27p.

[25] I. Ekeland, R. Temam: *Convex Analysis and Variational Problems*, SIAM, Philadelphia (1999).

[26] A. V. Fominyh, V. V. Karelin, L. N. Polyakova: *Application of the hypodifferential descent method to the problem of constructing an optimal control*, Optim. Lett. 12/8 (2018) 1825–1839.

[27] A. V. Fominykh: *Method of subdifferential and hypodifferential descent in the problem of constructing an integrally constrained program control*, Autom. Remote Control. 78/4 (2017) 608–617.

[28] R. M. Freund, H. Lu: *New computational guarantees for solving convex optimization problems with first order methods, via a function growth condition measure*, Math. Program. 170 (2018) 445–477.

[29] E. G. Gilbert: *An iterative procedure for computing the minimum of a quadratic form on a convex set*, SIAM J. Control 4 (1966) 61–80.

[30] G. N. Grapiglia, Y. Nesterov: *Accelerated regularized Newton methods for minimizing composite convex functions*, SIAM J. Optim. 29/1 (2019) 77–99.

[31] A. D. Ioffe, V. M. Tihomirov: *Theory of Extremal Problems*, North-Holland, Amsterdam (1979).

[32] P. R. Johnstone, P. Mouline: *Faster subgradient methods for functions with Hölderian growth*, Math. Program. 180 (2020) 417–450.

[33] D. Kim, J. A. Fessler: *Optimized first-order methods for smooth convex minimization*, Math. Program. 159 (2016) 81–107.

[34] D. Kim, J. A. Fessler: *Adaptive restart of the optimized gradient method for convex optimization*, J. Optim. Theory Appl. 178/1 (2018) 240–263.

[35] H. Luo, L. Chen: *From differential equation solvers to accelerated first-order method for convex optimization*, Math. Program. 195 (2022) 735–781.

[36] Y. Nesterov: *A method for unconstrained convex minimization problem with the rate of convergence* $O(1/k^2)$, Dokl. Akad. Nauk USSR 269/3 (1983) 543–547.

[37] Y. Nesterov: *Introductory Lectures on Convex Optimization. A Basic Course*, Kluwer, London (2004).

[38] Y. Nesterov: *Excessive gap technique in nonsmooth convex minimization*, SIAM J. Optim. 16 (2005) 235–249.

[39] Y. Nesterov: *Smooth minimization of non-smooth functions*, Math. Program. 103 (2005) 127–152.

[40] Y. Nesterov: *Primal-dual subgradient methods for convex problems*, Math. Program. 120 (2009) 221–259.

[41] A. Neumaier: *OSGA: a fast subgradient algorithm with optimal complexity*, Math. Program. 158 (2016) 1–21.

[42] J.-P. Penot: *Calculus Without Derivatives*, Springer, New York (2013).

[43] E. Polak: *Optimization, Algorithms and Consistent Approximations*, Springer, New York (1997).

[44] R. T. Rockafellar: *Convex Analysis*, Princeton University Press, Princeton (1970).

[45] R. T. Rockafellar, R. J.-B. Wets: *Variational Analysis*, Springer, Berlin (1998).

[46] B. Shi, S. S. Du, M. I. Jordan, W. J. Su: *Understanding the acceleration phenomenon via high-resolution differential equations*, Math. Program. 195 (2022) 79–148.

[47] N. Z. Shor: *Minimization Methods for Nondifferentiable Functions*, Springer, Berlin (1985).

[48] A. B. Taylor, J. M. Hendrickx, F. Glineur: *Exact worst-case convergence rates of the proximal gradient method for composite convex minimization*, J. Optim. Theory Appl. 178 (2018) 455–476.

[49] M. Teboulle: *A simplified view of first order methods for optimization*, Math. Program. 170 (2018) 67–96.

[50] L. Thibault: *Unilateral Variational Analysis in Banach Spaces. Part I: General Theory*, World Scientific, Singapore (2023).

[51] J.-P. Vial: *Strong and weak convexity of sets and functions*, Math. Oper. Res. 8/2 (1983) 231–259.

[52] P. Wolfe: *Finding the nearest point in a polytope*, Math. Program. 11 (1976) 128–149.

[53] Y. Xu: *Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming*, SIAM J. Optim. 27/3 (2017) 1459–1484.

[54] T. Yang, Q. Lin: *RSG: beating subgradient method without smoothness and strong convexity*, J. Mach. Learn. Res. 19 (2018) 1–33.

[55] C. Zalinescu: *Convex Analysis in General Vector Spaces*, World Scientific, Singapore (2002).