Codifferentials and Quasidifferentials of the Expectation of Nonsmooth Random Integrands and Two-Stage Stochastic Programming

M.V. Dolgopolik

April 23, 2024

Abstract

This work is devoted to an analysis of exact penalty functions and optimality conditions for nonsmooth two-stage stochastic programming problems. To this end, we first study the co-/quasi-differentiability of the expectation of nonsmooth random integrands and obtain explicit formulae for its co- and quasidifferential under some natural assumptions on the integrand. Then we analyse exact penalty functions for a variational reformulation of two-stage stochastic programming problems and obtain sufficient conditions for the global exactness of these functions with two different penalty terms. In the end of the paper, we combine our results on the co-/quasi-differentiability of the expectation of nonsmooth random integrands and exact penalty functions to derive optimality conditions for nonsmooth two-stage stochastic programming problems in terms of codifferentials.

1 Introduction

Two-stage stochastic programming is one of the basic problems of stochastic optimization [3, 10] that has multiple applications in various fields, including transportation planning [2, 30], disaster management [25], optimal design of energy systems [49], resources management [27], etc. Although two-stage stochastic programming problems can be viewed as stochastic versions of bilevel optimization problems [8, 9], their stochastic nature requires a largely different approach to their solution. Optimality conditions for two-stage stochastic programming problems were obtained in [26, 36, 40, 45, 46], while numerical methods for solving various classes of two-stage stochastic programming problems were studied e.g. in [23, 28, 32, 39] (see also the references therein).

The need for computing convex or nonconvex subdifferentials of the expectation of nonsmooth random integrands arises in many areas of stochastic optimization, including two-stage stochastic programming, as well as stochastic linear complementarity problems [6], stochastic variational inequalities [7], etc. The subdifferential in the sense of convex analysis of the expectation of a convex integrand was computed in [37], while its approximations were discussed in [31]. Various approximations of the Clarke subdifferential of the expectation...
of nonsmooth random integrands were studied in [5, 47], while an outer estimate of its Mordukhovich basic subdifferential was obtained in [46]. Finally, a quasidifferential of the expectation of quasidifferentiable random integrands was computed in [29].

The main goal of this paper is to apply constructive nonsmooth analysis [12, 14, 15] to a theoretical analysis of nonsmooth two-stage stochastic programming problems. Firstly, we analyse the codifferentiability and quasidifferentiability of the expectation of nonsmooth random integrands and present explicit formulae for its codifferential and quasidifferential in the more general case and under different assumptions than in [29] (see Remark 2 for more details).

In the second part of the paper we study exact penalty functions for two-stage stochastic programming problems, reformulated as equivalent variational problems with pointwise constraints. With the use of the general theory of exact penalty functions [11, 19, 22, 34, 38, 48], we obtain sufficient conditions for the global exactness of penalty functions for two-stage stochastic programming with two different types of penalty terms. The use of penalty terms of the first type leads to much less restrictive assumptions on constraints of the second stage problem, while the second type of penalty terms is more convenient for applications. In particular, it allows one to reformulate two-stage stochastic programming problems, whose second stage problem has DC (Difference-of-Convex) objective function and DC constraints, as equivalent unconstrained DC optimization problems and apply the well-developed apparatus of DC optimization to find their solutions (cf. analogous results for bilevel programming problems in [33, 42]). Let us also note that exact penalty functions for single-stage stochastic programming were analysed in [24].

Finally, in the end of the paper we combine our results on quasidifferentials of the expectation of nonsmooth random integrands and exact penalty functions for two-stage stochastic programming problems to obtains necessary optimality conditions for these problems in terms of codifferentials.

The paper is organised as follows. Some auxiliary definitions and facts from constructive nonsmooth analysis, that are necessary for understanding the paper, are collected in Section 2. Codifferentiability and quasidifferentiability of the expectation of nonsmooth random integrands is studied in Section 3 while Section 4 is devoted to nonsmooth two-stage stochastic programming problems. Exact penalty functions for such problems are analysed in Subsection 4.1 while optimality conditions for these problems in terms of codifferentials are derived in Subsection 4.2.

2 Preliminaries

Let us introduce the notation and briefly recall several definitions from nonsmooth analysis that will be used throughout the article. For more details in the finite dimensional case see [12, 14, 15]. The infinite dimensional case was studied in [16, 18, 20].

Let $X$ be a real Banach space. Denote by $X^*$ its topological dual, and by $(\cdot, \cdot)$ the duality pairing between $X$ and $X^*$. The weak* topology on $X^*$ is denoted by $w^*$ or $\sigma(X^*, X)$ depending on the context. Denote also by $\tau_\mathbb{R}$ the canonical topology of the real line $\mathbb{R}$. Let finally $U \subset X$ be an open set.
Definition 1. A function $f: U \to \mathbb{R}$ is called codifferentiable at a point $x \in U$, if there exists a pair of convex subsets $\partial f(x), \overline{\partial} f(x) \subset \mathbb{R} \times X^*$ that are compact in the topological product $(\mathbb{R} \times X^*, \tau_{\mathbb{R}} \times \tau_{X^*})$, satisfy the equality

$$\max_{(a,x^*) \in \partial f(x)} a = \min_{(b,y^*) \in \overline{\partial} f(x)} b = 0,$$

and for any $\Delta x \in X$ satisfy the following condition:

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha} \left| f(x + \alpha \Delta x) - f(x) - \max_{(a,x^*) \in \partial f(x)} \left( a + \langle x^*, \alpha \Delta x \rangle \right) - \min_{(b,y^*) \in \overline{\partial} f(x)} \left( b + \langle y^*, \alpha \Delta x \rangle \right) \right| = 0$$

The pair $Df(x) = [\partial f(x), \overline{\partial} f(x)]$ is called a codifferential of $f$ at $x$, the set $\partial f(x)$ is referred to as a hypodifferential of $f$ at $x$, while the set $\overline{\partial} f(x)$ is called a hyperdifferential of $f$ at $x$.

Remark 1. (i) In the case when $X = \mathbb{R}^d$, a codifferential $Df(x)$ is a pair of convex compact subsets of $\mathbb{R} \times \mathbb{R}^d = \mathbb{R}^{d+1}$ satisfying the equalities from the previous definition. In addition, if $X$ is a Hilbert space, then it is natural to suppose that a codifferential $Df(x)$ is a pair of convex weakly compact subsets of the space $\mathbb{R} \times X$.

(ii) Note that a codifferential is not uniquely defined. In particular, one can easily verify that for any compact convex subset $C$ of the space $(\mathbb{R} \times X^*, \tau_{\mathbb{R}} \times \tau_{X^*})$ the pair $[\partial f(x) + C, \overline{\partial} f(x) - C]$ is a codifferential of $f$ at $x$ as well.

Definition 2. A function $f: U \to \mathbb{R}$ is called continuously codifferentiable at a point $x \in U$, if $f$ is codifferentiable at every point in a neighbourhood of $x$ and there exists a codifferential mapping $Df(\cdot) = [\partial f(\cdot), \overline{\partial} f(\cdot)]$, defined in a neighbourhood of $x$ and such that the multifunctions $\partial f(\cdot)$ and $\overline{\partial} f(\cdot)$ are continuous in Hausdorff metric at $x$.

The class of continuously codifferentiable at a given point (or on a given set) functions is closed under addition, multiplication, composition with continuously differentiable functions, as well as pointwise maximum and minimum of finite families of functions. Moreover, any convex function is continuously codifferentiable in a neighbourhood of any given point from the interior of its effective domain, and any DC function (i.e. a function that can be represented as the difference of convex functions) is continuously codifferentiable in a neighbourhood of any given point. Numerous examples of continuously codifferentiable functions, as well as main rules of codifferential calculus can be found in [12, 14, 15, 18, 20].

Definition 3. A function $f: U \to \mathbb{R}$ is called quasidifferentiable at a point $x \in U$, if $f$ is directionally differentiable at $x$ and its directional derivative $f'(x, \cdot)$ at this point can be represented as the difference of sublinear functions or, equivalently, if there exists a pair $\partial f(x), \overline{\partial} f(x) \subset X^*$ of compact weak$^*$ compact sets such that

$$f'(x, h) = \max_{x^* \in \partial f(x)} \langle x^*, h \rangle + \min_{y^* \in \overline{\partial} f(x)} \langle y^*, h \rangle \quad \forall h \in X.$$
Just like codifferential, a quasidifferential is not uniquely defined. Here we only mention that a function $f$ is codifferentiable at a point $x$ iff $f$ is quasidifferentiable at $x$ and one can easily compute a quasidifferential of $f$ at $x$ from its codifferential at this point and vice versa. Namely, if $Df(x)$ is a codifferential of $f$ at $x$, then the pair $\mathcal{D} f(x) = \{ \partial f(x), \overline{\partial f(x)} \}$ with

$$\partial f(x) = \left\{ x^* \in X^* \mid (0, x^*) \in \partial f(x) \right\}, \quad \overline{\partial f(x)} = \left\{ y^* \in X^* \mid (0, y^*) \in \overline{\partial f(x)} \right\}$$

is a quasidifferential of $g$ at $x$. Conversely, if $\mathcal{D} f(x)$ is a quasidifferential of $f$ at $x$, then the pair $\{ \{0\} \times \partial f(x), \{0\} \times \overline{\partial f(x)} \}$ is a codifferential of $f$ at $x$ (see, e.g. [14, 20]). Below we consider only quasidifferentials of the form $\mathcal{D} f(x)$, that is, we suppose that if a codifferentiable function $f$ and its codifferential $Df(x)$ are given, then $\mathcal{D} f(x)$ is a quasidifferential of $f$ of the form $\mathcal{D} f(x)$.

Let us finally recall one auxiliary definition from set-valued analysis that will be used later (see, e.g. [1, Sect. 8.2] for more details). Let $X$ and $Y$ be metric spaces and $(\Omega, \mathfrak{F}, \mu)$ be a measure space. A set-valued mapping $F: X \times \Omega \Rightarrow Y$, $F = F(x, \omega)$ is called a *Carathéodory map* if for every $x \in X$ the multifunction $F(x, \cdot)$ is measurable and for a.e. $\omega \in \Omega$ the multifunction $F(\cdot, \omega)$ is continuous.

### 3 Codifferentials of the Expectation of Nonsmooth Random Integrands

Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and suppose that a nonsmooth function $f: \mathbb{R}^d \times \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}$, $f = f(x, y, \omega)$ is given. In this section we study the codifferentiability of the nonsmooth integral functional

$$\mathcal{I}(x, y) = \mathbb{E}[f(x, y, \cdot)] := \int_{\Omega} f(x, y(\omega), \omega) dP(\omega),$$

where $x \in \mathbb{R}^d$ is a parameter and $y \in L^p(\Omega, \mathfrak{F}, P; \mathbb{R}^m)$ with $1 < p \leq +\infty$ is an $m$-dimensional random vector. Although the case $p = 1$ can be included into the general theory under some additional assumptions, we exclude it for the sake of simplicity, since the proofs of the main results below are much more cumbersome in the case $p = 1$, than in the case $1 < p \leq +\infty$.

Denote by $p' \in [1, +\infty)$ the conjugate exponent of $p$, i.e. $1/p + 1/p' = 1$, and let $|\cdot|$ be the Euclidean norm in $\mathbb{R}^m$. Let us impose some assumptions on the integrand $f$ that, as we will show below, ensure that the functional $\mathcal{I}$ is correctly defined and codifferentiable.

Namely, we will suppose that for a.e. $\omega \in \Omega$ and for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^m$ the function $f$ is codifferentiable jointly in $x$ and $y$, that is, there exists a pair of compact convex sets $\mathcal{D}_{x,y} f(x, y, \omega), \overline{\mathcal{D}}_{x,y} f(x, y, \omega) \subset \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^m$ such that

$$\Phi_f(x, y, \omega; 0, 0) = \Psi_f(x, y, \omega; 0, 0) = 0,$$

and for all $(\Delta x, \Delta y) \in \mathbb{R}^d \times \mathbb{R}^m$ one has

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} \left| f(x + \alpha \Delta x, y + \alpha \Delta y, \omega) - f(x, y, \omega) \right. - \left. \Phi_f(x, y, \omega; \alpha \Delta x, \alpha \Delta y) - \Psi_f(x, y, \omega; \alpha \Delta x, \alpha \Delta y) \right| = 0,$$
where
\[
\Phi_f(x, y, \omega; \Delta x, \Delta y) = \max_{(a, v_x, v_y) \in \mathcal{D}_{x,y} f(x, y, \omega)} (a + \langle v_x, \Delta x \rangle + \langle v_y, \Delta y \rangle)
\]
\[
\Psi_f(x, y, \omega; \Delta x, \Delta y) = \min_{(b, w_x, w_y) \in \mathcal{D}_{x,y} f(x, y, \omega)} (b + \langle w_x, \Delta x \rangle + \langle w_y, \Delta y \rangle).
\]

The pair \(D_{x,y} f(x, y, \omega) = [\mathcal{D}_{x,y} f(x, y, \omega), \mathcal{D}_{x,y}^* f(x, y, \omega)]\) is called a codifferential of \(f\) in \((x, y)\).

**Assumption 1.** The function \(f\) satisfies the following conditions:

1. For any \(x \in \mathbb{R}^d\) the map \((y, \omega) \mapsto f(x, y, \omega)\) is a Carathéodory function;
2. The function \(f\) satisfies the following growth condition of order \(p\): for any \(N > 0\) there exist \(C_N > 0\) and a nonnegative function \(\beta_N \in L^1(\Omega, \mathfrak{F}, P)\) such that \(|f(x, y, \omega)| \leq \beta_N(\omega) + C_N|y|^p\) for all \(x \in \mathbb{R}^d\) with \(|x| \leq N\), all \(y \in \mathbb{R}^m\), and a.e. \(\omega \in \Omega\) in the case \(1 < p < +\infty\), and \(|f(x, y, \omega)| \leq \beta_N(\omega)\) for a.e. \(\omega \in \Omega\) and all \((x, y) \in \mathbb{R}^d \times \mathbb{R}^m\) with \(\max\{|x|, |y|\} \leq N\) in the case \(p = +\infty\);
3. The multifunctions \((y, \omega) \mapsto \mathcal{D}_{x,y} f(x, y, \omega)\) and \((y, \omega) \mapsto \mathcal{D}_{x,y}^* f(x, y, \omega)\) are Carathéodory maps for any \(x \in \mathbb{R}^d\);
4. The codifferential mapping \(D_{x,y} f(\cdot)\) satisfies the following growth condition of order \(p\): for any \(N > 0\) there exist \(C_N > 0\), and nonnegative functions \(\beta_N \in L^1(\Omega, \mathfrak{F}, P)\) and \(\gamma_N \in L^p(\Omega, \mathfrak{F}, P)\) such that
   \[
   \max\{|a|, |v_x|\} \leq \beta_N(\omega) + C_N|y|^p, \quad |v_y| \leq \gamma_N(\omega) + C_N|y|^{p-1}
   \]
   for all \((a, v_x, v_y) \in \mathcal{D}_{x,y} f(x, y, \omega) \cup \mathcal{D}_{x,y}^* f(x, y, \omega)\), all \(x \in \mathbb{R}^d\) with \(|x| \leq N\), all \(y \in \mathbb{R}^m\), and a.e. \(\omega \in \Omega\) in the case \(1 < p < +\infty\), and
   \[
   \max\{|a|, |v_x|, |v_y|\} \leq \beta_N(\omega)
   \]
   for all \((a, v_x, v_y) \in \mathcal{D}_{x,y} f(x, y, \omega) \cup \mathcal{D}_{x,y}^* f(x, y, \omega)\), a.e. \(\omega \in \Omega\), and for all vectors \((x, y) \in \mathbb{R}^d \times \mathbb{R}^m\) with \(\max\{|x|, |y|\} \leq N\) in the case \(p = +\infty\).

Note that the Carathéodory and the growth conditions on the function \(f\) ensure that the value \(\mathcal{I}(x, y)\) is correctly defined and finite for all \(x \in \mathbb{R}^d\) and \(y \in L^p(\Omega, \mathfrak{F}, P; \mathbb{R}^m)\). Let \(X = \mathbb{R}^d \times L^p(\Omega, \mathfrak{F}, P; \mathbb{R}^m)\).

**Theorem 1.** Let \(1 < p \leq +\infty\) and Assumption 1 be valid. Then the functional \(\mathcal{I}\) is codifferentiable on \(\mathbb{R}^d \times L(\Omega, \mathfrak{F}, P; \mathbb{R}^m)\), and for any \((x, y)\) from this space the pair \(D\mathcal{I}(x, y) = [\mathcal{D}\mathcal{I}(x, y), \mathcal{D}\mathcal{I}^*(x, y)]\), defined as
\[
\mathcal{D}\mathcal{I}(x, y) = \left\{ (A, x^*) \in \mathbb{R} \times X^* \left| A = \mathbb{E}[a], \langle x^*, (h_x, h_y) \rangle = \langle \mathbb{E}[v_x], h_x \rangle + \int\limits_\Omega (v_y(\omega), h_y(\omega)) dP(\omega) \quad \forall (h_x, h_y) \in X, \right. \right. \left. \left. (a(\cdot), v_x(\cdot), v_y(\cdot)) \right\} \right\}
\]

(4)
and
\[ \overline{\mathcal{I}}(x, y) = \left\{ (B, y^*) \in \mathbb{R} \times X^* \right. \mid B = \mathbb{E}[b], \]
\[ \langle y^*, (h_x, h_y) \rangle = \langle \mathbb{E}[v_x], h_x \rangle + \int_\Omega (v_y(\omega), h_y(\omega)) \, dP(\omega) \quad \forall (h_x, h_y) \in X, \]
\[ (b(\cdot), w_x(\cdot), w_y(\cdot)) \text{ is a measurable selection of the map } \mathcal{L}_{x,y} f(x, y(\cdot), \cdot) \}, \]
is a codifferential of \( \mathcal{I} \) at \( (x, y) \).

The proof of Theorem 1 is similar to the proof of the codifferentiability of the mapping \( \mathcal{I}(u) = \int_\Omega f(x, u(x), \nabla u(x)) \, dx \) from the author’s papers \([17, 21]\) (here \( \Omega \subseteq \mathbb{R}^n \) is an open set and \( u \) belongs to the Sobolev space). On the other hand, Theorem 1 cannot be directly deduced from the main results of \([17, 21]\). That is why below we present a detailed proof of Theorem 1. It seems possible to prove a more general result on the codifferentiability of integral functionals defined on Banach spaces that subsumes Theorem 1 and the main results of \([17, 21]\) as particular cases. A development of such general theorem on the codifferentiability of nonsmooth integral functionals is an interesting open problem for future research.

For the sake of convenience, we divide the proof of Theorem 1 into two lemmas.

Lemma 1. Let \( 1 < p \leq +\infty \) and Assumption 1 be valid. Then for any \( (x, y) \in X \) the sets \( \mathcal{L}_{x,y} f(x, y, \cdot) \) and \( \overline{\mathcal{I}}(x, y) \) from Theorem 1 are nonempty, convex, compact in the topological product \( (\mathbb{R} \times X^*, \tau_{\mathbb{R}} \times w^*) \), and satisfy the following equalities:
\[ \max_{(A, x^*) \in \mathcal{L}_{x,y} f(x, y, \cdot)} A = \min_{(B, y^*) \in \overline{\mathcal{I}}(x, y)} B = 0. \quad (5) \]

Proof. Fix any \( (x, y) \in X \). We prove the statement of the lemma only for the hypodifferential \( \mathcal{L}_{x,y} f(x, y, \cdot) \), since the proof for the hyperdifferential \( \overline{\mathcal{I}}(x, y) \) is exactly the same.

By Assumption 1 the multifunction \( (y, \omega) \mapsto \mathcal{L}_{x,y} f(x, y, \omega) \) is a Carathéodory map. Therefore by \([11]\) Thrm. 8.2.8 the multifunction \( \mathcal{L}_{x,y} f(x, y(\cdot), \cdot) \) is measurable, which by \([11]\) Thrm. 8.1.3 implies that there exist a measurable selection \( (a(\cdot), v_x(\cdot), v_y(\cdot)) \) of this mapping. Furthermore, by the growth condition on the codifferential \( \overline{\mathcal{I}}(x, y) \) from Assumption 1 all measurable selections of the set-valued mapping \( \mathcal{L}_{x,y} f(x, y(\cdot), \cdot) \) belong to the space
\[ Y := L^1(\Omega, \mathcal{A}, P) \times L^1(\Omega, \mathcal{A}, P; \mathbb{R}^d) \times L^{\infty}(\Omega, \mathcal{A}, P; \mathbb{R}^m). \quad (6) \]

Consequently, the linear functional \( x^* \), defined as
\[ \langle x^*, (h_x, h_y) \rangle = \langle \mathbb{E}[v_x], h_x \rangle + \int_\Omega \langle v_y(\omega), h_y(\omega) \rangle \, dP(\omega) \quad \forall (h_x, h_y) \in X, \]
belongs to \( X^* \), and one can conclude that the hypodifferential \( d\mathcal{I}(x, y) \) is correctly defined and nonempty.

Denote by \( \mathcal{E}(x, y) \) the set of all measurable selections \( z(\cdot) = (a(\cdot), v_x(\cdot), v_y(\cdot)) \) of the set-valued mapping \( \mathcal{L}_{x,y} f(x, y(\cdot), \cdot) \). As was noted above, \( \mathcal{E}(x, y) \) is a subset of the space \( Y \) defined in \((6)\). For any \( z = (a, v_x, v_y) \in Y \) denote by \( \mathcal{T}(z) \) the pair \( (A, x^*) \) defined as in \((9)\). Then \( d\mathcal{I}(x, y) = \mathcal{T}(\mathcal{E}(x, y)) \).
By definition, for a.e. \( \omega \in \Omega \) the hypodifferential \( d_{x,y}f(x,y(\omega),\omega) \) is a convex set. Therefore the set of measurable selections \( \mathcal{E}(x,y) \) of the multifunction \( \frac{d_{x,y}f(x,y(\omega),\omega)}{T} \) is convex. Hence taking into account the fact that the operator \( \mathcal{T} \) is linear one obtains that the hypodifferential \( d\mathcal{T}(x,y) \) is a convex set as the image of a convex set under a linear map.

Recall that by the definition of hypodifferential one has \( a \leq 0 \) for any \((a,v_x,v_y) \in d_{x,y}f(x,y(\omega),\omega), \omega \in \Omega \). Therefore \( A \leq 0 \) for all \((A,x^*) \in d\mathcal{T}(x,y) \). On the other hand, observe that thanks to equality (1) for a.e. \( \omega \in \Omega \), one has
\[
0 \in \{ a \in \mathbb{R} \mid \exists (v_x,v_y) \in \mathbb{R}^{d+m}: (a,v_x,v_y) \in d_{x,y}f(x,y(\omega),\omega) \}.
\]

Hence by the Filippov theorem (see, e.g. [1, Thrm. 8.2.10]) there exists a measurable selection \((a_0(\cdot),v_{x0}(\cdot),v_{y0}(\cdot))\) of the set-valued map \( d_{x,y}f(x,y(\cdot),\cdot) \) such that \( a_0(\omega) = 0 \) almost surely. Consequently, for \((A_0,x_0^*) = \mathcal{T}(a_0,v_{x0},v_{y0})\) one has \( A_0 = 0 \), which implies that equality (1) holds true.

Thus, it remains to prove the compactness of the set \( d\mathcal{T}(x,y) \) in the corresponding product topology. To this end, let us verify that the set \( \mathcal{E}(x,y) \) is a weakly compact subset of the space \( Y \) defined in (6), and the operator \( \mathcal{T} \) continuously maps the space \( Y \) endowed with the weak topology to the topological product \( (\mathbb{R},\tau_\mathbb{R}) \times (X^*,w^*) \). Then one can conclude that the hypodifferential \( d\mathcal{T}(x,y) \) is compact in the corresponding product topology as a continuous image of a compact set.

We start with the proof of the continuity of the operator \( \mathcal{T} \). Let \( \mathcal{V} \) be an open subset of the product space \( (\mathbb{R},\tau_\mathbb{R}) \times (X^*,w^*) \). Let us show that its preimage \( \mathcal{U} = \mathcal{T}^{-1}(\mathcal{V}) \) under the map \( \mathcal{T} \) is weakly open in \( Y \). Indeed, fix any \((a,v_x,v_y) \in \mathcal{U} \). Then \((A,x^*) = \mathcal{T}(a,v_x,v_y) \in \mathcal{V} \), which due to the openness of the set \( \mathcal{V} \) in the corresponding topology implies that there exist \( \varepsilon > 0 \), \( n \in \mathbb{N} \), and pairs \((h_i,\xi_i) \in X, i \in I = \{1, \ldots, n\}, \) such that
\[
\mathcal{V}_\varepsilon(A,x^*) = \{(B,y^*) \in \mathbb{R} \times X^* \mid |B-A| < \varepsilon, \max_{i \in I} |\langle y^*-x^*,(h_i,\xi_i)\rangle| < \varepsilon\} \subseteq \mathcal{V}.
\]

Introduce the set
\[
\mathcal{U}_\varepsilon(a,v_x,v_y) = \{(b,w_x,w_y) \in Y \mid |\mathbb{E}(b-a)| < \varepsilon, \max_{i \in I} \left( \int_{\Omega} |w_x(\omega) - v_x(\omega), h_i| dP(\omega) \right) < \frac{\varepsilon}{2}, \max_{i \in I} \left( \int_{\Omega} |w_y(\omega) - v_y(\omega), \xi_i(\omega)| dP(\omega) \right) < \frac{\varepsilon}{2}\}.
\]

This set is neighbourhood of the point \((a,v_x,v_y) \) in the weak topology on \( Y \). Moreover, by definition \( \mathcal{T}(\mathcal{U}_\varepsilon(a,v_x,v_y)) \subseteq \mathcal{V}_\varepsilon(A,x^*) \), which implies that \( \mathcal{U}_\varepsilon(a,v_x,v_y) \subseteq \mathcal{U} \). Thus, for any point \((a,v_x,v_y) \in \mathcal{U} \) there exists a neighbourhood of this point in the weak topology contained in \( \mathcal{U} \). In other words, the set \( \mathcal{U} \) is weakly open, and one can conclude that the operator \( \mathcal{T} \) is continuous with respect to the chosen topologies.

Let us finally proof the weak compactness of the set \( \mathcal{E}(x,y) \) in the space \( Y \) defined in (6). By the Eberlein-Šmulian theorem it suffices to prove that \( \mathcal{E}(x,y) \) is weakly sequentially compact. To this end, choose any sequence \( z_n(\cdot) = (a_n(\cdot),v_{xn}(\cdot),v_{yn}(\cdot)) \in \mathcal{E}(x,y), n \in \mathbb{N} \). Let us consider two cases.
Case $p = +\infty$. By the growth condition on the codifferential $D_{x,y}f(\cdot)$ (see Assumption 1) there exists an a.e. nonnegative function $\beta \in L^1(\Omega, \mathfrak{A}, P)$ such that for a.e. $\omega \in \Omega$ one has
\[
\max \left\{ |a_n(\omega)|, |v_n(\omega)|, |v_n(\omega)| \right\} \leq \beta(\omega) \quad \forall n \in \mathbb{N}.
\]
Hence by the weak compactness criterion in $L^1$ (see, e.g. [4, Thrm. 4.7.20]) the closures of the sets $\{a_n\}_{n \in \mathbb{N}}$, $\{v_n\}_{n \in \mathbb{N}}$, and $\{v_n\}_{n \in \mathbb{N}}$ are weakly compact in the corresponding $L^1$ spaces. Therefore by the Eberlein-Šmulian theorem there exists a subsequence $z_{n_k} = (a_{n_k}, v_{n_k}, v_{n_k})$ weakly converging to some $z_*$ in $Y$. By Mazur’s lemma there exists a sequence of convex combinations $\{z_k\}$ of elements of the sequence $z_{n_k}$ strongly converging to $z_*$. Therefore, as is well known, there exists a subsequence $\{z_k\}$ converging to $z_*$ almost surely.

Note that due to the convexity of $E(x,y)$ one has $\hat{z}_k \subset E(x,y)$, that is, $\hat{z}_k(\omega) \in d_{x,y}f(x,y(\omega),\omega)$ for a.e. $\omega \in \Omega$ and all $k \in \mathbb{N}$. Hence taking into account the fact that by definition the hypodifferential $d_{x,y}f(x,y(\omega),\omega)$, $\omega \in \Omega$, is a closed set, one obtains that $z_*(\omega) \in d_{x,y}f(x,y(\omega),\omega)$ for a.e. $\omega \in \Omega$. Thus, $z_* \in E(x,y)$, and the set $E(x,y)$ is weakly sequentially compact, which completes the proof.

Case $p < +\infty$. By the growth condition on the codifferential $D_{x,y}f(\cdot)$ (see Assumption 1) there exist $C > 0$ and a.e. nonnegative functions $\beta \in L^1(\Omega, \mathfrak{A}, P)$ and $\gamma \in L^p(\Omega, \mathfrak{A}, P)$ such that for a.e. $\omega \in \Omega$ and all $n \in \mathbb{N}$ one has
\[
\max \{ |a_n(\omega)|, |v_n(\omega)| \} \leq \beta(\omega) + C|y(\omega)|^p, \quad |v_n(\omega)| \leq \gamma(\omega) + C|y(\omega)|^{p-1}.
\]
Observe that the right-hand side of the first inequality belongs to $L^1(\Omega, \mathfrak{A}, P)$, while the right-hand side of the second one belongs to $L^p(\Omega, \mathfrak{A}, P)$. Thus, the sequence $\{v_n\}$ is norm-bounded in $L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m)$, which due to the reflexivity of this space (note that $1 < p' < +\infty$, since $1 < p < +\infty$) implies that there exists a weakly convergent subsequence $\{v_{n_k}\}$. In turn, the existence of weakly convergence subsequences of the sequences $\{a_n\}$ and $\{v_n\}$ follows from the weak compactness criterion in $L^1$ (see [4, Thrm. 4.7.20]).

Thus, there exists a subsequence $\{z_{n_k}\}$ weakly converging to some $z_* \in Y$. Now, applying Mazur’s lemma and arguing precisely in the same way as in the case $p = +\infty$ one can prove the weak compactness of the set $E(x,y)$.

Denote by $\| \cdot \|_p$ the standard norm on $L^p(\Omega, \mathfrak{A}, P)$.

**Lemma 2.** Let $1 < p \leq +\infty$, Assumption 1 be valid, and the sets $\partial I(x,y)$ and $\partial I(x,y)$ be defined as in Theorem [1]. Then for any $(x,y) \in X$ and $(\Delta x, \Delta y) \in X$ one has
\[
\lim_{\alpha \to +0} \frac{1}{\alpha} \left| I(x + \alpha \Delta x, y + \alpha \Delta y) - I(x,y) - \max_{(A,x^*) \in \partial I(x,y)} (A + (x^*, \alpha(\Delta x, \Delta y))) \right| - \min_{(B,y^*) \in \partial I(x,y)} (B + (y^*, \alpha(\Delta x, \Delta y))) = 0.
\]

**Proof.** Fix any $(x,y) \in X$ and $(\Delta x, \Delta y) \in X$, and choose an arbitrary sequence $\{a_n\} \subset (0, +\infty)$ converging to zero. For a.e. $\omega \in \Omega$ and $n \in \mathbb{N}$ denote
\[
f_n(\omega) = \frac{1}{\alpha_n} \left( f(x + \alpha_n \Delta x, y(\omega) + \alpha_n \Delta y(\omega)), \omega) - f(x, y(\omega), \omega) - \Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) - \Psi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) \right),
\]

8
where the functions $\Phi_f$ and $\Psi_f$ are defined in (3). By the definition of codifferential-
ity the sequence $f_n$ converges to zero almost surely. Our aim is to prove that
each term in the definition of $f_n$ belongs to $L^1(\Omega, \mathcal{F}, P)$ and there exists
an a.e. nonnegative function $\rho \in L^1(\Omega, \mathcal{F}, P)$ such that $|f_n| \leq \rho$ almost surely.
Then by Lebesgue’s dominated convergence theorem $\mathbb{E}[|f_n|] \to 0$ as $n \to \infty$. Hence
integrating each term in the definition of $f_n$ separately one obtains that
\[
\lim_{n \to \infty} \frac{1}{\alpha_n} [\mathcal{I}(x + \alpha_n \Delta x, y + \alpha_n \Delta y) - \mathcal{I}(x, y)] \\
- \int_\Omega \Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) \, dP(\omega) \\
- \int_\Omega \Psi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) \, dP(\omega) = 0.
\]
Let us check that
\[
\int_\Omega \Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) \, dP(\omega)
= \max_{(A, x^*') \in d_x f(x, y(\cdot), \cdot)} (A + \langle x^*, \alpha_n (\Delta x, \Delta y) \rangle) \tag{8}
\]
(a similar equality for the min terms involving the hyperdifferentials can be verified in the same way). Then one obtains the desired result.

Indeed, by definition (see (3)) for any measurable selection $(\alpha(\cdot), v_x(\cdot), v_y(\cdot))$
of the set-valued mapping $d_x f(x, y(\cdot), \cdot)$ one has
\[
\Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) \geq \alpha(\omega) + \langle v_x(\omega), \alpha_n \Delta x \rangle + \langle v_y(\omega), \alpha_n \Delta y(\omega) \rangle,
\]
which implies that
\[
\int_\Omega \Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) \, dP(\omega) \geq \max_{(A, x^*) \in d_x f(x, y(\cdot), \cdot)} (A + \langle x^*, \alpha_n \Delta x \rangle)
\]
(see (3)). On the other hand, for a.e. $\omega \in \Omega$ one has
\[
\Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega))
\in \left\{ \alpha + \langle v_x, \alpha_n \Delta x \rangle + \langle v_y, \alpha_n \Delta y(\omega) \rangle \mid \langle a, v_x, v_y \rangle \in d_x f(x, y(\omega), \omega) \right\}.
\]
Consequently, by the Filippov theorem (see, e.g. [11], Thm. 8.2.10) there exists
a measurable selection $(\alpha_0(\cdot), v_{x0}(\cdot), v_{y0}(\cdot))$ of the multifunction $d_x f(x, y(\cdot), \cdot)$
such that
\[
\Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) = \alpha_0(\omega) + \langle v_{x0}(\omega), \alpha_n \Delta x \rangle + \langle v_{y0}(\omega), \alpha_n \Delta y(\omega) \rangle
\]
for a.e. $\omega \in \Omega$. Hence for the corresponding pair $(A_0, x^*_0) = \mathcal{T}(\alpha_0, v_{x0}, v_{y0})$
(see the proof of Lemma [11]), that by definition belongs to $d_x f(x, y)$, one has
\[
\int_\Omega \Phi_f(x, y(\omega), \omega; \alpha_n \Delta x, \alpha_n \Delta y(\omega)) \, dP(\omega) = A_0 + \langle x^*_0, \alpha_n (\Delta x, \Delta y) \rangle,
\]
and therefore equality (8) holds true.
Thus, it remains to show that Lebesgue’s dominated convergence theorem is applicable to the sequence \( \{ f_n \} \). Indeed, the first two terms in the definition of \( f_n \) (see (7)) belong to \( L^1(\Omega, \mathcal{F}, P) \) by virtue of the first two parts of Assumption 1. Let us check that these terms are dominated by a Lebesgue integrable function independent of \( n \).

By the mean value theorem for codifferentiable functions [20, Prp. 2] for any \( n \in \mathbb{N} \) and for a.e. \( \omega \in \Omega \) there exist \( \alpha_n(\omega) \in (0, \alpha) \) and

\[
(0, v_{x_n}(\omega), v_{y_n}(\omega)) \in \mathcal{D}_{x,y} f(x + \alpha_n(\omega)\Delta x, y(\omega) + \alpha_n(\omega)\Delta y(\omega), \omega),
\]

\[
(0, w_{x_n}(\omega), w_{y_n}(\omega)) \in \mathcal{D}_{x,y} f(x + \alpha_n(\omega)\Delta x, y(\omega) + \alpha_n(\omega)\Delta y(\omega), \omega)
\]
such that

\[
\frac{1}{\alpha_n} \left( f(x + \alpha_n\Delta x, y(\omega) + \alpha_n\Delta y(\omega), \omega) - f(x, y(\omega), \omega) \right) = \langle v_{x_n}(\omega) + w_{x_n}(\omega), \Delta x \rangle + \langle v_{y_n}(\omega) + w_{y_n}(\omega), \Delta y(\omega) \rangle. \tag{9}
\]

Put \( \alpha_* = \max_{n \in \mathbb{N}} \alpha_n \). By the growth condition on the codifferential \( D_{x,y} f \) (see Assumption 1) there exist \( C_N > 0 \) and nonnegative functions \( \beta_N \in L^1(\Omega, \mathcal{F}, P) \) and \( \gamma_N \in L^p(\Omega, \mathcal{F}, P) \) (here \( N = |x| + \alpha_*|\Delta x| \)) such that

\[
\max \{ |v_{x_n}(\omega)|, |w_{x_n}(\omega)| \} \leq \beta_N(\omega) + C_N |y(\omega) + \alpha_n(\omega)\Delta y(\omega)|^p
\]

\[
\leq \beta_N(\omega) + C_N 2^p \left( |y(\omega)|^p + |\alpha_n(\omega)|^\beta \Delta y(\omega)|^p \right),
\]

\[
\max \{ |v_{y_n}(\omega)|, |w_{y_n}(\omega)| \} \leq \gamma_N(\omega) + C_N |y(\omega) + \alpha_n(\omega)\Delta y(\omega)|^{p-1}
\]

for a.e. \( \omega \in \Omega \) and all \( n \in \mathbb{N} \) in the case \( 1 < p < +\infty \), and there exists \( \beta_N \in L^1(\Omega, \mathcal{F}, P) \) (here \( N = \max\{ |x| + \alpha_*|\Delta x|, \|y\|_\infty + \alpha_*\|\Delta y\|_\infty \} \)) such that

\[
\max \{ |v_{x_n}(\omega)|, |w_{x_n}(\omega)|, |v_{y_n}(\omega)|, |w_{y_n}(\omega)| \} \leq \beta_N(\omega)
\]

for a.e. \( \omega \in \Omega \) and all \( n \in \mathbb{N} \) in the case \( p = +\infty \). Hence with the use of (9) one obtains that in the case \( p = +\infty \) the inequality

\[
\frac{1}{\alpha_n} \left| f(x + \alpha_n\Delta x, y(\omega) + \alpha_n\Delta y(\omega), \omega) - f(x, y(\omega), \omega) \right| \leq 2\beta_N(\omega)|\Delta x| + 2\beta_N(\omega)\|\Delta y\|_\infty
\]

holds true for a.e. \( \omega \in \Omega \) and all \( n \in \mathbb{N} \), which implies that the first two terms in the definition of \( f_n \) (see (7)) are dominated by a Lebesgue integrable function independent of \( n \). In the case \( p < +\infty \) one has

\[
\frac{1}{\alpha_n} \left| f(x + \alpha_n\Delta x, y(\omega) + \alpha_n\Delta y(\omega), \omega) - f(x, y(\omega), \omega) \right| \leq 2 \left( \beta_N(\omega) + C_N 2^p \left( |y(\omega)|^p + |\alpha_n|\Delta y(\omega)|^p \right) \right) |\Delta x|
\]

\[
+ 2 \left( \gamma_N(\omega) + C_N 2^{p-1} \left( |y(\omega)|^{p-1} + |\alpha_n|^{p-1}\Delta y(\omega)|^{p-1} \right) \right) |\Delta y(\omega)|
\]

The right-hand side of this inequality does not depend on \( n \) and is Lebesgue integrable, as one can easily verify with the use of Hölder’s inequality and the equality \( p'(p - 1) = p \). Thus, in the case \( p < +\infty \) the first two terms in the
definition of $f_n$ are dominated by a Lebesgue integrable function independent of $n$ as well.

Let us finally check that the third term in the definition of $f_n$, denoted by

$$\theta_n(\omega) := \frac{1}{\alpha_n} \max_{(a,v_x,v_y) \in \mathbb{D}_{x,y} f(x,y(\omega),\omega)} \left( a + \langle v_x, \alpha_n \Delta x \rangle + \langle v_y, \alpha_n \Delta y(\omega) \rangle \right)$$

(see [1]), is measurable and dominated by a Lebesgue integrable function independent of $n$. The fact that the last term (the min term) in the definition of $f_n$ is measurable and dominated by a Lebesgue integrable function independent of $n$ is proved in exactly the same way.

As was shown in the proof of Lemma [1] the set-valued mapping $\mathbb{D}_{x,y} f(x,y(\cdot),\cdot)$ is measurable. Consequently, the function $\theta_n$ is measurable by [1] Thm. 8.2.11.

For any $\omega \in \Omega$ introduce the function

$$g_\omega(t) = \max_{(a,v_x,v_y) \in \mathbb{D}_{x,y} f(x,y(\omega),\omega)} \left( a + \langle v_x, t \Delta x \rangle + \langle v_y, t \Delta y(\omega) \rangle \right).$$

Observe that by the definition of codifferential $g_\omega(0) = 0$ (see Def. [1]) and for any $t, \Delta t \in \mathbb{R}$ and $\alpha > 0$ one has

$$\frac{1}{\alpha} |g_\omega(t + \alpha \Delta t) - g_\omega(t) - \max_{(a,v_y) \in \mathbb{D}_{x,y} f(x,y(\omega),\omega)} \left( a + v_y(\alpha \Delta t) \right)| = 0,$$

where

$$\mathbb{D}_{x,y} f(x,y(\omega),\omega) = \left\{ (a, v_x, v_y) \in \mathbb{R} \times \mathbb{R} \mid a = a + \langle v_x, t \Delta x \rangle + \langle v_y, t \Delta y(\omega) \rangle - g_\omega(t),\right.$$

$$\left. v_y = \langle v_x, \Delta x \rangle + \langle v_y, \Delta y(\omega) \rangle, (a, v_x, v_y) \in \mathbb{D}_{x,y} f(x,y(\omega),\omega) \right\}.$$

The set $\mathbb{D}_{x,y} f(x,y(\omega),\omega)$ is obviously convex and compact. Moreover, note that the equality $\max\{a_y \mid (a_y, v_y) \in \mathbb{D}_{x,y} f(x,y(\omega),\omega)\} = g_\omega(t) - g_\omega(0) = 0$ holds true. Thus, the function $g_\omega$ is codifferentiable at every point $t \in \mathbb{R}$, and the pair $[\mathbb{D}_{x,y} f(x,y(\omega),\omega)]$ is a codifferential of $g_\omega$ at the point $t$.

Applying the mean value theorem for codifferentiable functions [20] Prop. 2 one obtains that for any $n \in \mathbb{N}$ and for a.e. $\omega \in \Omega$ there exists $\alpha_n(\omega) \in (0, \alpha_n)$ and $(0, v_y(\omega)) \in \mathbb{D}_{x,y} f(x,y(\omega),\omega)$ such that

$$\theta_n(\omega) = \frac{1}{\alpha_n} (g_\omega(\alpha_n) - g_\omega(0)) = v_y(\omega)$$

or, equivalently, there exists $(\alpha_n(\omega), v_x(\omega), v_y(\omega)) \in \mathbb{D}_{x,y} f(x,y(\omega),\omega)$ such that

$$\theta_n(\omega) = \langle v_x(\omega), \Delta x \rangle + \langle v_y(\omega), \Delta y(\omega) \rangle \quad \forall n \in \mathbb{N}.$$
for a.e. \( \omega \in \Omega \) in the case \( p = +\infty \). The right-hand sides of these inequalities are Lebesgue integrable and do not depend on \( n \). Thus, the sequence \( \{ \theta_n \} \) is dominated by a Lebesgue integrable function, which completes the proof. 

With the use of Theorem \( \text{I} \) one can easily obtain sufficient conditions for the quasidifferentiability of the functional \( \mathcal{I} \). Recall that \( X = \mathbb{R}^d \times L^p(\Omega, \mathcal{A}, P; \mathbb{R}^m) \).

**Corollary 1.** Let \( 1 < p \leq +\infty \) and Assumption \( I \) be valid. Then the functional \( \mathcal{I} \) is quasidifferentiable on \( \mathbb{R}^d \times L^p(\Omega, \mathcal{A}, P; \mathbb{R}^m) \), and for any \((x, y)\) from this space the pair \( \partial \mathcal{I}(x, y) = [\partial^\mathcal{I}(x, y), \partial^\mathcal{I}(x, y)] \), defined as

\[
\partial^\mathcal{I}(x, y) = \left\{ x^* \in X^* \mid (x^*, (h_x, h_y)) = \langle E[v_x], h_x \rangle + \int_\Omega (v_y(\omega), h_y(\omega)) dP(\omega) \right\}
\]

\( \forall (h_x, h_y) \in X, \quad (0, v_x(\cdot), v_y(\cdot)) \) is a measurable selection of \( d_{x,y}f(x, y(\cdot), \cdot) \) and

\[
\partial^\mathcal{I}(x, y) = \left\{ y^* \in X^* \mid (y^*, (h_x, h_y)) = \langle E[w_x], h_x \rangle + \int_\Omega (w_y(\omega), h_y(\omega)) dP(\omega) \right\}
\]

\( \forall (h_x, h_y) \in X, \quad (0, w_x(\cdot), w_y(\cdot)) \) is a measurable selection of \( d_{x,y}f(x, y(\cdot), \cdot) \), is a quasidifferential of \( \mathcal{I} \) at \((x, y)\). Moreover, the following equality holds true:

\[
\mathcal{I}'(x, y; h_x, h_y) = \int_\Omega [f_0(\cdot, \omega)]'(x, y(\omega); h_x, h_y(\omega)) dP(\omega) \quad \forall (h_x, h_y) \in X.
\]

**Proof.** Applying Theorem \( \text{I} \) and the fact that any codifferentiable function \( g \) with codifferential \( Dg(x) \) is quasidifferentiable and the pair

\[
\partial g(x) = \left\{ x^* \in X^* \mid (0, x^*) \in g'(x) \right\}, \quad \partial g(x) = \left\{ y^* \in X^* \mid (0, y^*) \in g'(x) \right\}
\]

is a quasidifferential of \( g \) at \( x \) (see, e.g. \[13, \text{Thm. 2.2}\]), one obtains the required results on the quasidifferentiability of the functional \( \mathcal{I} \).

To prove equality (10), recall that the set-valued mappings \( d_{x,y}f(x, y(\cdot), \cdot) \) and \( d_{x,y}f(x, y(\cdot), \cdot) \) are measurable, as was shown in the proof of Lemma \( \text{I} \). Hence with the use of \[11, \text{Thm. 8.2.4}\] one obtains that the set-valued mappings \( d_{x,y}f(x, y(\cdot), \cdot) \) and \( d_{x,y}f(x, y(\cdot), \cdot) \), defined according to equalities (2), are measurable as well. Consequently, applying the definition of quasidifferentiability and arguing in the same way as in the proof of Lemma \( \text{I} \) (or utilising the interchangeability principle; see, e.g. \[35, \text{Thm. 14.60}\]) one gets that

\[
\mathcal{I}'(x, y; h_x, h_y) = \int_\Omega \left( \max_{(v_x, v_y) \in d_{x,y}f(x, y(\omega), \omega)} \langle v_x, h_x \rangle + \langle v_y, h_y(\omega) \rangle \right) dP(\omega) 
\]

\[
+ \min_{(w_x, w_y) \in d_{x,y}f(x, y(\omega), \omega)} \left( \langle w_x, h_x \rangle + \langle w_y, h_y(\omega) \rangle \right) dP(\omega)
\]

for all \((h_x, h_y) \in X\), which by the definition of quasidifferential of the function \( f \) implies that equality (10) holds true. 

\[ \square \]
Remark 2. In the particular case when the function \( f \) does not depend on \( y \), i.e. \( f = f(x, \omega) \), the previous corollary contains sufficient conditions for the quasidifferentiability of the function \( F(x) = \mathbb{E}[f(x, \cdot)] \). Quasidifferentiability of this function was studied in the recent paper [29] under different assumptions on the function \( f \). Namely, instead of imposing any growth conditions, in [29] it was assumed that all integrals are correctly defined and the function \( f \) is locally Lipschitz continuous in \( x \) uniformly in \( \omega \).

Let us finally show that under the assumptions of Theorem 1 the functional \( I(x, y) \) is not only codifferentiable, but also Lipschitz continuous on bounded sets.

**Corollary 2.** Let \( 1 < p \leq +\infty \) and Assumption 1 be valid. Then \( I \) is Lipschitz continuous on any bounded subset of the space \( X = \mathbb{R}^d \times L^p(\Omega, \mathcal{A}, P; \mathbb{R}^m) \).

**Proof.** With the use of the growth condition on the codifferential mapping \( D_{x,y} f(\cdot) \) from Assumption 1 one can readily verify that both multifunctions \( dI(\cdot) \) and \( d\mathcal{I}(\cdot) \) are bounded on bounded subsets of the space \( X \). Therefore by [20, Corollary 2] the functional \( I \) is Lipschitz continuous on any bounded subset of this space. \(\blacksquare\)

4 Nonsmooth Two-Stage Stochastic Programming

Let, as above, \((\Omega, \mathcal{A}, P)\) be a probability space. In this section we study a general two-stage stochastic programming problem of the form

\[
\min_{x \in A} \mathbb{E}[F(x, \omega)],
\]

where \( F(x, \omega) \) is the optimal value of the second stage problem

\[
\min_{y \in G(x, \omega)} f(x, y, \omega).
\]

Here \( A \subset \mathbb{R}^d \) is a closed set, \( f: \mathbb{R}^d \times \mathbb{R}^m \times \Omega \to \mathbb{R} \) is a Carathéodory function, and \( G: \mathbb{R}^d \times \Omega \rightrightarrows \mathbb{R}^m \) is a multifunction. We assume that \( G \) is measurable and for every \( \omega \in \Omega \) the multifunction \( G(\cdot, \omega) \) is closed.

Choose any \( 1 \leq p \leq +\infty \), and denote \( X = \mathbb{R}^d \times L^p(\Omega, \mathcal{A}, P; \mathbb{R}^m) \). By the interchangeability principle for two-stage stochastic programming (see, e.g. [40, Thrm. 2.20]), problem (11), (12) is equivalent the following variational problem with pointwise constraints:

\[
\min_{(x, y) \in X} \mathbb{E}[f(x, y(\cdot), \cdot)]
\]

subject to \( x \in A, \ y(\omega) \in G(x, \omega) \) for a.e. \( \omega \in \Omega \).

\(\text{(P)}\)

in the sense that the optimal values of these problems coincide, and if this common optimal value is finite, then for any globally optimal solution \((x_*, y_*(\cdot))\) of the problem \(\text{(P)}\) the point \(x_*\) is a globally optimal solution of problem (11) and for a.e. \( \omega \in \Omega \) the point \( y_*(\omega) \) is a globally optimal solution of the second stage problem (12). Conversely, if \( x_* \) is a globally optimal solution of problem (11) and for a.e. \( \omega \in \Omega \) the point \( y_*(\omega) \) is a globally optimal solution of problem (12)
with \( x = x^* \) such that \( y^* \in L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m) \), then \((x^*, y^*)\) is a globally optimal solution of the problem \((P)\).

Since problem \((11), (12)\) and the problem \((P)\) are equivalent, below we consider only the problem \((P)\). Our aim is to present several results on exact penalty functions for the problem \((P)\), which not only allow one to obtain optimality conditions for the original two-stage stochastic programming problem, but also can be used for design and analysis of exact penalty methods for solving problem \((11), (12)\).

4.1 Exact Penalty Functions

Fix any \( p \in [1, +\infty] \), and denote by \( I(x, y) = \int_\Omega f(x, y(\omega), \omega) dP(\omega) \) the objective function of the problem \((P)\). Below we suppose that the functional \( I \) is correctly defined on the space \( X := \mathbb{R}^d \times L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m) \) and does not take the value \(-\infty\). In particular, it is sufficient to suppose that for any \( x \in \mathbb{R}^d \) there exist \( C > 0 \) and an a.e. nonnegative function \( \beta \in L^1(\Omega, \mathfrak{A}, P) \) such that \( |f(x, y, \omega)| \leq \beta(\omega) + C|y|^p \) for a.e. \( \omega \in \Omega \) and all \( y \in \mathbb{R}^m \) in the case \( p < +\infty \), and for any \( x \in \mathbb{R}^d \) and \( N > 0 \) there exist an a.e. nonnegative function \( \beta_N \in L^1(\Omega, \mathfrak{A}, P) \) such that \( |f(x, y, \omega)| \leq \beta_N(\omega) \) for a.e. \( \omega \in \Omega \) and all \( y \in \mathbb{R}^m \) with \( |y| \leq N \).

Introduce the set \( M = \{(x, y) \in X \mid y(\omega) \in G(\omega, \omega) \text{ for a.e. } \omega \in \Omega\} \).

Then the problem \((P)\) can be rewritten as follows:

\[
\min_{(x, y) \in X} I(x, y) \quad \text{subject to} \quad (x, y) \in M \cap (A \times L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m)).
\]

Let \( \varphi : X \to [0, +\infty] \) be any function such that \( \varphi(x, y) = 0 \) iff \((x, y) \in M\), and let \( \Phi_c(x, y) = I(x, y) + c\varphi(x, y) \). The function \( \Phi_c \) is called a penalty function for the problem \((P)\) with \( c \geq 0 \) being the penalty parameter, while the function \( \varphi \) is called a penalty term for the constrain \((x, y) \in M\). Our aim is to obtain sufficient conditions for the exactness of the penalty function \( \Phi_c \).

Recall that the penalty function \( \Phi_c \) is called globally exact, if there exists \( c_\ast \geq 0 \) such that for any \( c \geq c_\ast \) the set of globally optimal solutions of the penalized problem

\[
\min_{(x, y) \in X} \Phi_c(x, y) \quad \text{subject to} \quad x \in A
\]

coincides with the set of globally optimal solutions of the problem \((P)\). The greatest lower bound of all such \( c_\ast \) is called the least exact penalty parameter of the penalty function \( \Phi_c \). One can verify that the penalty function \( \Phi_c \) is globally exact iff there exists \( c_\ast \geq 0 \) such that for any \( c \geq c_\ast \), the problem \((P)\) and problem \((13)\) have the same optimal value, and the greatest lower bound of all such \( c_\ast \) coincides with the least exact penalty parameter. See [11,19,22,34,38,48] for more details on exact penalty functions.
Let us obtain sufficient conditions for the global exactness of the penalty function \( \Phi_c \) with the penalty term \( \varphi \) defined in several different ways. To this end, we will utilise general sufficient conditions for the exactness of penalty functions in metric and normed spaces from [19,22], and the following auxiliary lemma, which is a slight generalization of [19, Prp. 3.13].

**Lemma 3.** Let \( Y \) be a normed space, \( F \subset Y \) be a nonempty set, and a function \( F: Y \to \mathbb{R} \cup \{+\infty\} \) be such that for any bounded set \( C \subset Y \) there exists a continuous from the right function \( \omega_C: [0, +\infty) \to [0, +\infty) \) for which

\[
|F(y_1) - F(y_2)| \leq \omega_C(\|y_1 - y_2\|) \quad \forall y_1, y_2 \in C. \tag{14}
\]

Then for any \( R > 0 \) there exists a bounded set \( C \subset Y \) such that

\[
F(y) \geq \inf_{z \in F} F(z) - \omega_C(\operatorname{dist}(y, F)) \quad \forall y \in B(0, R) = \{z \in Y \mid \|z\| \leq R\}. \tag{15}
\]

**Proof.** Denote \( F_* = \inf_{z \in F} F(z) \), and fix any \( R > 0 \) and \( z \in F \). By our assumption there exists a continuous from the right function \( \omega_C \) such that inequality \((14)\) holds true for \( C = B(0, R + \|z\|) \).

Choose any \( y \in B(0, R) \). If \( y \in F \), then inequality \((15)\) trivially holds true. Suppose now that \( y \in B(0, R) \setminus F \). Clearly, there exists a sequence \( \{y_n\} \subset F \) such that \( \|y - y_n\| \to \operatorname{dist}(y, F) \) as \( n \to \infty \), and the inequalities \( \|y - y_n\| \leq \|y - z\| \leq R + \|z\| \) and \( \|y - y_n\| \geq \|y - y_{n+1}\| \) are satisfied for all \( n \in \mathbb{N} \). By definition \( \{y_n\} \subset C \), \( y \in C \), and \( F(y_n) \geq F_* \) for all \( n \in \mathbb{N} \). Therefore, by applying inequality \((14)\) one obtains that

\[
F_* - F(y) = F_* - F(y_n) + F(y_n) - F(y) \leq F(y_n) - F(y) \leq \omega_C(\|y - y_n\|)
\]

for any \( n \in \mathbb{N} \). Hence passing to the limit as \( n \to \infty \) with the use of the fact that the function \( \omega_C \) is continuous from the right and the sequence \( \{\|y - y_n\|\} \) is non-increasing one gets that inequality \((15)\) holds true.

**Remark 3.** Note that if \( F \) is Lipschitz continuous on bounded sets, then inequality \((14)\) holds true with \( \omega_C(t) = L_C t \), where \( L_C \) is a Lipschitz constant of \( F \) on \( C \). In this case the statement of the lemma can be reformulated as follows: for any \( R > 0 \) there exists \( L > 0 \) such that \( F(y) \geq F_* - L \operatorname{dist}(y, F) \) for all \( y \in B(0, R) \). Thus, Lemma 3 provides a lower estimate of the decay of the function \( F \) relative to a given set \( F \).

We start our analysis of the exactness of the penalty function \( \Phi_c \) with the simplest case when the penalty term \( \varphi \) is defined via the distance function to the multifunction \( G \). Denote by \( \mathcal{I}_* \) the optimal value of the problem \((P)\).

**Theorem 2.** Let there exist a globally optimal solution of the problem \((P)\), the set-valued mapping \( G \) have closed images, and

\[
\varphi(x, y) = \left( \mathbb{E} \left[ \operatorname{dist}(y(\cdot), G(x, \cdot))^p \right] \right)^{1/p} \quad \forall (x, y) \in X
\]

in the case \( p < +\infty \), and \( \varphi(x, y) = \operatorname{ess sup}_{\omega \in \mathcal{G}} \operatorname{dist}(y(\omega), G(x, \omega)) \) for all \((x, y) \in X\) in the case \( p = +\infty \). Suppose also that the functional \( \mathcal{I} \) is Lipschitz continuous on bounded sets, and there exists \( c \geq 0 \) such that the set

\[
\{(x, y) \in X \mid x \in A, \Phi_c(x, y) < \mathcal{I}_* \}
\]

is bounded. Then the penalty function \( \Phi_c \) is globally exact.
Proof. Observe that the function \( \varphi \) is correctly defined for all \((x, y) \in X\), since the multifunction \( G \) is measurable. Moreover, \( \varphi \) is nonnegative, and \( \varphi(x, y) = 0 \) iff \((x, y) \in M\). Denote by \( \mathcal{F} \) the feasible set of the problem \((\mathcal{P})\). Let us show that

\[
\varphi(x, y) \geq \text{dist}\left((x, y), \mathcal{F}\right) \quad \forall x \in A, \ y \in L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m)
\]

(16)

Indeed, fix any \((x, y) \in X\) such that \( x \in A \). If \( \varphi(x, y) = +\infty \), then inequality (16) obviously holds true. Suppose now that \( \varphi(x, y) < +\infty \). Then, in particular, one has \( G(x, \omega) \neq \emptyset \) for a.e. \( \omega \in \Omega \).

By our assumptions the multifunction \( G \) is measurable and has closed images. Therefore by [11] Corr. 8.2.13 there exists a measurable selection \( z \) of the set-valued mapping \( G(x, \cdot) \) such that

\[
|y(\omega) - z(\omega)| = \text{dist}\left(y(\omega), G(x, \omega)\right) \quad \text{for a.e.} \ \omega \in \Omega.
\]

Let us check that \( z \in L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m) \). Then \((x, z) \in \mathcal{F}\) and

\[
\varphi(x, y) = \|y - z\|_p = \|(x, y) - (x, z)\| \geq \text{dist}\left((x, y), \mathcal{F}\right),
\]

that is, inequality (16) holds true.

To verify that \( z \) belongs to the space \( L^p \), observe that

\[
|z(\omega)| \leq |y(\omega)| + |z(\omega) - y(\omega)| = |y(\omega)| + \text{dist}\left(y(\omega), G(x, \omega)\right)
\]

for a.e. \( \omega \in \Omega \). The right-hand side of this inequality belongs to \( L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m) \) due to the fact that \( \varphi(x, y) < +\infty \). Therefore the function \( z \) belongs to this space as well.

Thus, inequality (16) holds true. Since the functional \( \mathcal{I} \) is Lipschitz continuous on bounded sets, by Lemma 3 for any \( R > 0 \) there exists \( L > 0 \) such that

\[
\mathcal{I}(x, y) \geq \mathcal{I}_* - L \text{dist}\left((x, y), \mathcal{F}\right) \quad \forall (x, y) \in B(0, R).
\]

Hence by [19] Prp. 3.16 and Remark 15, part (ii) the penalty function \( \Phi_c \) is globally exact. \( \Box \)

**Remark 4.** Note that by Corollary 2 the functional \( \mathcal{I} \) is Lipschitz continuous on bounded sets in the case \( p > 1 \), provided the integrand \( f \) satisfies Assumption 1. In turn, as one can readily verify, the set \( \{(x, y) \in X \mid x \in A, \ \Phi_c(x, y) < L_*\} \) is bounded for some \( \epsilon \geq 0 \), if \( 1 \leq p < +\infty \) and one of the following conditions is satisfied:

1. the set \( A \) is bounded, and the multifunction \( G \) is bounded on \( A \times \Omega \);

2. the set \( A \) is bounded, and there exist \( C > 0 \) and \( \beta \in L^1(\Omega, \mathfrak{A}, P) \) such that \( f(x, y, \omega) \geq C|y| + \beta(\omega) \) for all \((x, y) \in A \times \mathbb{R}^m \) and a.e. \( \omega \in \Omega \);

3. the multifunction \( G \) is bounded on \( A \times \Omega \), and there exist \( \beta \in L^1(\Omega, \mathfrak{A}, P) \) and a function \( \rho: [0, +\infty) \rightarrow [0, +\infty) \) such that \( \rho(t) \rightarrow +\infty \) as \( t \rightarrow +\infty \), and \( f(x, y, \omega) \geq \rho(|x|) + \beta(\omega) \) for all \((x, y) \in \mathbb{R}^{d+m} \) and a.e. \( \omega \in \Omega \);

4. there exist \( C > 0, \beta \in L^1(\Omega, \mathfrak{A}, P) \), and a function \( \rho: [0, +\infty) \rightarrow [0, +\infty) \) such that \( \rho(t) \rightarrow +\infty \) as \( t \rightarrow +\infty \), and \( f(x, y, \omega) \geq \rho(|x|) + C|y| + \beta(\omega) \) for all \((x, y) \in \mathbb{R}^{d+m} \) and a.e. \( \omega \in \Omega \);
5. \((\Omega, \mathcal{A}, P)\) is a finite probability space, and \(\min_{\omega \in \Omega} f(x, y, \omega) \to +\infty\) as \(|x| + |y| \to +\infty\).

In the case \(p = +\infty\) the set \(\{(x, y) \in X \mid x \in A, \Phi_c(x, y) < I_*\}\) is bounded, provided the first, the third or the last of the assumptions above is satisfied.

In most particular cases the feasible set \(G(x, \omega)\) of the second stage problem (12) is not defined explicitly, but rather via some constraints. As a result, one usually does not know an explicit expression for the penalty term \(\varphi\) from Theorem 2, which makes this theorem inapplicable to real-world problems, at least in a direct way. In some cases Theorem 2 can still be applied indirectly to reduce an analysis of the exactness of a penalty function for the problem \((P)\) to an analysis of constraints of the second stage problem. Let us explain this statement with the use of a simple example.

**Example 1.** Suppose that the set-valued map \(G\) is defined in the following way:

\[
G(x, \omega) = \left\{ y \in \mathbb{R}^m \mid 0 \in Q(x, y, \omega) \right\}
\]

where \(Q: \mathbb{R}^d \times \mathbb{R}^m \times \Omega \to \mathbb{R}^s\) is a multifunction with closed images. In other words, the second stage problem (12) has the form:

\[
\min_y f(x, y, \omega) \quad \text{subject to} \quad 0 \in Q(x, y, \omega).
\]

In this case it is natural to define

\[
\varphi(x, y) = \left( \mathbb{E} \left[ \text{dist}(0, Q(x, y(\cdot), \cdot))^p \right] \right)^{1/p}, \quad 1 \leq p < +\infty.
\]

Then \(\varphi(x, y) = 0\) iff \((x, y) \in M\). Suppose that there exists \(K > 0\) such that

\[
K \text{dist}(0, Q(x, y, \omega)) \geq \text{dist}(y, G(x, \omega)) \quad \forall x \in A, \; y \in \mathbb{R}^m, \; \omega \in \Omega,
\]

that is, the function \(g(y) = \text{dist}(0, Q(x, y, \omega))\) admits a global error bound uniform for all \(x \in A\) and \(\omega \in \Omega\). Then

\[
\Phi_{Kc}(x, y) = I(x, y) + Kc\varphi(x, y) \geq I(x, y) + c\psi(x, y)
\]

for all \(x \in A\) and \(y \in L^p(\Omega, \mathcal{A}, P; \mathbb{R}^m)\), where

\[
\psi(x, y) = \left( \mathbb{E} \left[ \text{dist}(y(\cdot), G(x, \cdot))^p \right] \right)^{1/p}.
\]

Therefore, as one can readily verify (cf. [19, Prp. 2.2]), under the assumptions of Theorem 2 the penalty function \(\Phi_c\) is globally exact and its least exact penalty parameter is at most \(K\) times greater than the least exact penalty parameter of the penalty function from Theorem 2.

Let us also point out two simple cases when Theorem 2 can be applied directly, that is, the cases when one can write a simple explicit expression for the penalty term \(\varphi\) from this theorem. Note that Theorem 2 can be applied directly whenever the distance from a given point \(y\) to the set \(G(x, \omega)\) is easy to compute, e.g. when the set \(G(x, \omega)\) is defined by linear or, more generally, convex quadratic constraints.
Example 2. Let \( I := \{1, \ldots, m\} \). Suppose that the set \( G(x, \omega) \) is defined by bound (box) constraints, that is,

\[
G(x, \omega) = \left\{ y = (y_1, \ldots, y_m)^T \in \mathbb{R}^m \ \big| \ a_i(x, \omega) \leq y_i \leq b_i(x, \omega), \ i \in I \right\}
\]

for some given functions \( a_i \) and \( b_i \). Let the space \( \mathbb{R}^m \) be equipped with the \( \ell^\infty \) norm. Then the penalty term \( \varphi \) from Theorem 2 has the form

\[
\varphi(x, y) = \left( \int_{\Omega} \max_{i \in I} \left\{ 0, y_i(\omega) - b_i(x, \omega), a_i(x, \omega) - y_i(\omega) \right\}^p dP(\omega) \right)^{1/p}
\]

in the case \( 1 \leq p < +\infty \).

Example 3. Let \( G(x, \omega) = B(z(x, \omega), R(x, \omega)) \) be the closed ball with centre \( z(x, \omega) \) and radius \( R(x, \omega) \). Then the penalty term \( \varphi \) from Theorem 2 has the form

\[
\varphi(x, y) = \left( \int_{\Omega} \max \left\{ 0, |y(\omega) - z(x, \omega)| - R(x, \omega) \right\}^p dP(\omega) \right)^{1/p}.
\]

in the case \( 1 \leq p < +\infty \).

Observe that the penalty terms from Theorem 2 and the examples above depend on the parameter \( p \) that defines the space in which one solves the problem \((P)\). This parameter must be chosen to satisfy the assumption of Theorem 2.

Under some additional assumptions on constraints of the second stage problem one can prove the global exactness of the penalty function \( \Phi \), with a penalty term \( \varphi \) that does not depend on \( p \). For the sake of simplicity, we will prove this result only in the case when the feasible set \( G(x, \omega) \) of the second stage problem is defined by inequality constraints, i.e. it has the form

\[
G(x, \omega) = \left\{ y \in \mathbb{R}^m \ \big| \ g_i(x, y, \omega) \leq 0, \ i \in I = \{1, \ldots, \ell\} \right\}
\]

for some functions \( g_i : \mathbb{R}^d \times \mathbb{R}^m \times \Omega \to \mathbb{R} \). Below we suppose that for each \( x \in \mathbb{R}^d \) the map \( (y, \omega) \mapsto g_i(x, y, \omega), i \in I, \) is a Carathéodory function, so that the penalty term

\[
\varphi(x, y) = \int_{\Omega} \max_{i \in I} \{0, g_i(x, y, \omega)\} \, dP(\omega) \tag{17}
\]

is correctly defined. Note that \( \varphi(x, y) = 0 \) iff \( (x, y) \in M \). We will assume that for any \( x \in \mathbb{R}^d \) and a.e. \( \omega \in \Omega \) the function \( y \mapsto g_i(x, y, \omega), i \in I, \) is quasidifferentiable and denote by \( D_yg_i(x, y, \omega) = (\partial_y g_i(x, y, \omega), \overline{\partial}_y g_i(x, y, \omega)) \) its quasidifferential. Denote also \( I(x, y, \omega) = \{i \in I \ | \ g_i(x, y, \omega) = \max_{k \in I} g_k(x, y, \omega)\} \).

Let \((Y, d)\) be a metric space, \( K \subset Y \) be a given set, and \( g : Y \to \mathbb{R} \cup \{+\infty\} \) be a given function. Recall that for any \( y \in K \cap \text{dom} \ g \) the quantity

\[
g^+_K(y) = \liminf_{z \to y, z \in K} \frac{g(z) - g(y)}{d(z, y)}
\]

is called the rate of steepest descent of \( g \) at \( y \). If \( y \) is not a limit point of the set \( K \), then by definition \( g^+_K(y) = +\infty \). Recall also that a point \( y \in K \cap \text{dom} \ g \) is called an inf-stationary point of \( g \) on the set \( K \), if \( g^+_K(y) \geq 0 \). It should be noted that
in various particular cases this inequality is reduced to standard stationarity conditions. For example, if $Y$ is normed space, $g$ is Fréchet differentiable at a point $y \in K$, and the set $K$ is convex, then $g'_K(z) \geq 0$ if $g'(y)[z - y] \geq 0$ for all $z \in K$, where $g'(y)$ is the Fréchet derivative of $g$ at $y$. See [10,11,43,44] for more details on the rate of steepest descent and the definition of inf-stationarity.

**Theorem 3.** Let $1 \leq p < +\infty$ and the following assumptions be valid:

1. there exist a globally optimal solution of the problem (P);
2. the functional $\mathcal{I}$ is Lipschitz continuous on bounded sets;
3. the set $S_c(\gamma) = \{(x, y) \in X \mid x \in A, \Phi_c(x, y) < \gamma\}$ is bounded for some $c > 0$ and $\gamma > I_*$, where $\Phi_c$ is the penalty functions with the penalty term \[17\];
4. for any $x \in A$ there exists an a.e. nonnegative function $L(\cdot) \in L^1(\Omega, \mathfrak{A}, P)$ such that $|g_i(x, y_1, \omega) - g_i(x, y_2, \omega)| \leq L(\omega)\|y_1 - y_2\|$ for all $y_1, y_2 \in \mathbb{R}^d$, all $i \in I$ and a.e. $\omega \in \Omega$;
5. for all $i \in I$, $x \in A$, and $y \in L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m)$ the set-valued mappings $\partial_y g_i(x, y(\cdot), \cdot)$ and $\overline{\partial}_y g_i(x, y(\cdot), \cdot)$ are measurable;
6. there exists $\alpha > 0$ such that for any $(x, y) \in A \times \mathbb{R}^m$ and a.e. $\omega \in \Omega$ such that $y \notin G(x, \omega)$, and for all $i \in I(x, y, \omega)$ one can find $w_i(x, y, \omega) \in \overline{\partial}_y g_i(x, y, \omega)$ satisfying the following condition:

$$\text{dist}\left(0, \text{co}\left\{\partial_y g_i(x, y, \omega) + w_i(x, y, \omega) \mid i \in I(x, y, \omega)\right\}\right) \geq \alpha. \tag{18}$$

Then the penalty function $\Phi_c$ with the penalty term [17] is globally exact and there exists $c_* \geq 0$ such that for any $c \geq c_*$ the following statements hold true:

1. $(x_*, y_*) \in S_c(\gamma)$ is a locally optimal solution of the penalized problem [18] iff $(x_*, y_*)$ is a locally optimal solution of the problem (P);
2. $(x_*, y_*) \in S_c(\gamma)$ is an inf-stationary point of the penalty function $\Phi_c$ on the set $A \times L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m)$ iff $(x_*, y_*)$ is an inf-stationary point of the functional $\mathcal{I}$ on the feasible set $\mathcal{F}$ of the problem (P).

**Proof.** Let us show that under the assumptions of the theorem $\varphi^*(x, \cdot)(y) \leq -a$ for any $(x, y) \in X \setminus \mathcal{F}$ such that $x \in A$ and $\varphi(x, y) < +\infty$ (here $\varphi^*(x, \cdot)(y)$ is the rate of steepest descent of the function $y \mapsto \varphi(x, y)$ at the point $y$). Then applying [22] Thrm. 2 one obtains the required result.

To prove the required estimate for $\varphi^*(x, \cdot)(y)$, we first construct a descent direction for the function $\varphi$ using condition [15], and then obtain an upper estimate for the rate of steepest descent via the directional derivative of $\varphi$ along the constructed descent direction.

Fix any $(x, y) \in X \setminus \mathcal{F}$ such that $x \in A$ and $\varphi(x, y) < +\infty$. Recall that by the definition of quasidifferential one has

$$Q_i(h, \omega) = (g_i(x, \cdot, \omega))'(y(\omega), h) = \max_{v \in \partial_y g_i(x, y(\omega), \omega)} \langle v, h \rangle + \min_{w \in \overline{\partial}_y g_i(x, y(\omega), \omega)} \langle w, h \rangle \tag{19}$$
(see Def. 3). Applying Assumption 5 and Thm. 8.2.11 one obtains that the function $Q_i$ is measurable in $\omega$ for any $h \in \mathbb{R}^m$. Moreover, since in the finite dimensional case the quasidifferential is a pair of compact convex sets, the function $Q_i$ is continuous in $h$ for a.e. $\omega \in \Omega$, i.e. $Q_i$ is a Carathéodory function.

Let us now prove that the multifunction $I(\cdot) := I(x, y(\cdot), \cdot), I: \Omega \rightarrow \{1, \ldots, \ell\}$ is measurable. Indeed, by definitions for any nonempty subset $K \subseteq \{1, \ldots, \ell\}$ one has

$$I^{-1}(K) = \{ \omega \in \Omega \mid I(x, y(\omega), \omega) \cap K \neq \emptyset \}$$

$$= \{ \omega \in \Omega \mid \max_{i \in K} g_i(x, y(\omega), \omega) \geq \max_{i \in I} g_i(x, y(\omega), \omega) \}.$$

This set is measurable, since the functions $g_i(x, y(\cdot), \cdot)$ are measurable due to the fact that the maps $(y, \omega) \mapsto g_i(x, y, \omega)$ are Carathéodory functions by our assumption. Thus, for any subset $K \subseteq \{1, \ldots, s\}$ the set $I^{-1}(K)$ is measurable, that is, the set-valued map $I(\cdot)$ is measurable by definition (see, e.g. [1], Def. 8.1.1).

Introduce the set

$$E = \{ \omega \in \Omega \mid \max_{i \in I} g_i(x, y(\omega), \omega) > 0 \}.$$

Note that the set $E$ is measurable, thanks to our assumption that the mappings $(y, \omega) \mapsto g_i(x, y, \omega)$ are Carathéodory functions. Moreover, $P(E) > 0$ due to the fact that $(x, y)$ is not a feasible point of the problem (P).

Since the multifunction $I(\cdot)$ is measurable and $Q_i$ are Carathéodory functions, the set-valued mapping

$$H(\omega) := \left\{ h \in \mathbb{R}^m \mid |h| = 1, \max_{i \in I(\omega)} Q_i(h, \omega) = \min_{|z|=1} \max_{i \in I(\omega)} Q_i(z, \omega) \right\}, \quad \omega \in E$$

is measurable by Thm. 8.2.11. Furthermore, this multifunction obviously has closed images. Therefore by Thm. 8.1.3 there exists a measurable function $h_*: E \rightarrow \mathbb{R}^m$ such that $h_*(\omega) \in H(\omega)$ for all $\omega \in E$. For any $\omega \in \Omega \setminus E$ define $h_*(\omega) = 0$. Then $h_*: \Omega \rightarrow \mathbb{R}^m$ is a measurable function and, moreover, $\|h_*\|_p = P(E) > 0$.

From condition (15) and the separation theorem it follows that for any $\omega \in E$ there exists $\hat{h}(\omega) \in \mathbb{R}^m$ with $|\hat{h}(\omega)| = 1$ such that

$$\langle v, \hat{h}(\omega) \rangle \leq -a \quad \forall v \in \co \left\{ \partial g_i(x, y(\omega), \omega) + w_i(x, y(\omega), \omega) \mid i \in I(\omega) \right\}.$$

Hence with the use of (16) one obtains that $Q_i(\hat{h}(\omega), \omega) \leq -a$ for all $\omega \in E$ and $i \in I(\omega)$, which by the definition of $h_*$ implies that

$$\max_{i \in I(\omega)} Q_i(h_*(\omega), \omega) \begin{cases} \leq -a, & \text{if } \omega \in E, \\ \geq 0, & \text{if } \omega \notin E. \end{cases} \quad (20)$$

Thus, the function $h_*$ is the desired descent direction, along which we will evaluate the directional derivative of the penalty term $\varphi$. 

20
Indeed, denote $\psi(\omega, \alpha) = \max_{i \in I} \{0, g_i(x, y(\omega) + \alpha h_*(\omega), \omega)\}$ for all $\alpha > 0$ and $\omega \in \Omega$. Applying relations (20) and standard calculus rules for directional derivatives (see, e.g., [14]) one gets that

$$\lim_{\alpha \to +0} \frac{\psi(\omega, \alpha) - \psi(\omega, 0)}{\alpha} = \begin{cases} \max_{i \in I(\omega)} Q_i(h_*(\omega), \omega) \leq -a, & \text{if } \omega \in E, \\ 0, & \text{if } \omega \not\in E. \end{cases}$$

Applying Assumption [3] and the well-known fact that the maximum of a finite family of Lipschitz continuous is Lipschitz continuous (see, e.g., [13, Appendix III]) one obtains that there exists an a.e. nonnegative function $L(\cdot) \in L^1(\Omega, \mathfrak{A}, P)$ such that

$$\left| \frac{\psi(\omega, \alpha) - \psi(\omega, 0)}{\alpha} \right| \leq L(\omega)|h_*(\omega)| \leq L(\omega) \quad \forall \alpha > 0, \quad \text{a.e. } \omega \in \Omega.$$

Note also that $\psi(\cdot, 0) \in L^1(\Omega, \mathfrak{A}, P)$, since $\varphi(x, y) < +\infty$. Hence by the inequality above $\psi(\cdot, \alpha) \in L^1(\Omega, \mathfrak{A}, P)$ for all $\alpha > 0$. Consequently, applying Lebesgue’s dominated convergence theorem and the fact that $\varphi(x, y + \alpha h_*) = \mathbb{E}[\psi(\cdot, \alpha)]$ one obtains that

$$\left[\varphi(x, \cdot)\right]'(y; h_*) = \lim_{\alpha \to +0} \frac{\varphi(x, y + \alpha h_*) - \varphi(x, y)}{\alpha} = \int_E \max_{i \in I(\omega)} Q_i(h_*(\omega), \omega) \, dP(\omega) \leq -aP(E).$$

Therefore

$$\varphi^+(x, \cdot)(y) = \liminf_{\alpha \to +0} \frac{\varphi(x, z) - \varphi(x, y)}{\|z - y\|_p} \leq \liminf_{\alpha \to +0} \frac{\varphi(x, y + \alpha h_*) - \varphi(x, y)}{\alpha\|h_*\|_p} = \left[\varphi(x, \cdot)\right]'(y; h_*) \frac{\varphi(x, y)}{\|h_*\|_p} \leq \frac{aP(E)}{P(E)} = -a,$$

and the proof is complete.

**Remark 5.** (i) Note that by [35, Crl. 14.14] the multifunctions $\partial_y g_i(x, y(\cdot), \cdot)$ and $\overline{\partial}_y g_i(x, y(\cdot), \cdot)$ are measurable for any measurable function $y(\cdot)$, provided for any $\omega \in \Omega$ the mapping $\partial_y g_i(x, \cdot, \omega)$ is outer semicontinuous and the graphical mapping $\omega \mapsto \text{Graph} \partial_y g_i(x, \cdot, \omega)$ is measurable.

(ii) In the case when the functions $g_i$ are continuously differentiable in $y$, assumption [13] is satisfied if there exists $a > 0$ such that for any $(x, y) \in \mathbb{R}^{d+m}$ and a.e. $\omega \in \Omega$ such that $y \not\in G(x, \omega)$ one has

$$\text{dist} \left(0, \co \left\{ \nabla_y g_i(x, y, \omega) \mid i \in I(x, y, \omega) \right\} \right) \geq a.$$

This condition can be viewed as a uniform Mangasarian-Fromovitz constraint qualification. In turn, in the case when the functions $g_i$ are convex in $y$, assumption [13] is satisfied if there exists $a > 0$ such that for any $(x, y) \in \mathbb{R}^{d+m}$ and a.e. $\omega \in \Omega$ such that $y \not\in G(x, \omega)$ one has

$$\text{dist} \left(0, \co \left\{ \partial_y g_i(x, y, \omega) \mid i \in I(x, y, \omega) \right\} \right) \geq a.$$

where $\partial_y g_i(x, y, \omega)$ is the subdifferential of the function $g_i(x, \cdot, \omega)$ in the sense of convex analysis.

21
Remark 6. Let for a.e. $\omega \in \Omega$ the functions $(x,y) \mapsto f(x,y,\omega)$ and $(x,y) \mapsto g_i(x,y,\omega)$, $i \in I$, be DC (Difference-of-Convex), that is, there exist convex in $(x,y)$ functions $f_1(x,y,\omega)$, $f_2(x,y,\omega)$, $g_{i1}(x,y,\omega)$, and $g_{i2}(x,y,\omega)$ such that

$$f(x,y,\omega) = f_1(x,y,\omega) - f_2(x,y,\omega), \quad g_i(x,y,\omega) = g_{i1}(x,y,\omega) - g_{i2}(x,y,\omega)$$

for all $(x,y) \in \mathbb{R}^{d+m}$, $i \in I$, and a.e. $\omega \in \Omega$. Then the penalty function from Theorem 3 is DC as well. Namely, one has $\Phi_c(x,y) = \Phi^1_c(x,y) - \Phi^2_c(x,y)$, where

$$\Phi^1_c(x,y) = \int_\Omega \left( f_1(x,y(\omega),\omega) + c \max_{i \in I} \left\{ 0, g_{i1}(x,y(\omega),\omega) + \sum_{k \neq i} g_{k2}(x,y(\omega),\omega) \right\} \right) dP(\omega),$$

and

$$\Phi^2_c(x,y) = \int_\Omega \left( f_2(x,y(\omega),\omega) + c \sum_{i \in I} g_{i2}(x,y(\omega),\omega) \right) dP(\omega)$$

are convex functionals. Therefore with the use of Theorem 3 and well-known global optimality conditions for DC optimization problems one can easily obtain global optimality conditions for the problem $(P)$ and the original two-stage stochastic programming problem (cf. [41]). Moreover, under the assumptions of Theorem 3 one can apply well-developed methods of DC optimization to find local or global minima of the DC penalty function $\Phi_c(x,y)$, which coincide with local/global minima of the problem $(P)$. Thus, Theorem 3 opens a way for applications of DC programming algorithms to two-stage stochastic programming problems (cf. [33,42]).

4.2 Optimality Conditions

Let us finally derive optimality conditions for the problem $(P)$ in terms of codifferentials. We will derive these conditions by applying standard optimality conditions for quasidifferentiable functions to an exact penalty function for the problem $(P)$.

For the sake of shortness, we will consider only the case when the set $A$ is convex and obtain optimality conditions under the assumptions of Theorem 3. It should be noted that one can obtain such conditions under less restrictive assumptions on the functional $I$ and the penalty function $\Phi_c$, if one considers the so-called local exactness of the penalty function instead of the global one (see [11,19]). Moreover, one can significantly relax the assumptions on the constraints of the second-stage problem by considering the case $p = +\infty$ and utilising the highly nonsmooth penalty term

$$\varphi(x,y) = \operatorname{ess sup}_{\omega \in \Omega} \left\{ \max_{i \in I} \{ 0, g_i(x,y(\omega),\omega) \} \right\}. $$

However, the price one has to pay for less restrictive assumptions on constraints is the reduced regularity of Lagrange multipliers (see the theorem below). Namely, in this case one must assume that the Lagrange multipliers are just finitely additive measures.

For any convex subset $K$ of a Banach space $Y$ and any $y \in K$ denote the normal cone to the set $K$ at the point $y$ by $N_K(y) = \{ y^* \in Y^* \mid \langle y^*, z - y \rangle \leq 0 \ \forall z \in K \}$. 

22
Theorem 4. Let $1 < p < +\infty$, the set $A$ be convex, the feasible set of the second-stage problem \([12]\) have the form

$$G(x, \omega) = \left\{ y \in \mathbb{R}^m \mid g_i(x, y, \omega) \leq 0, \; i \in I = \{1, \ldots, l\} \right\}$$

for some functions $g_i : \mathbb{R}^d \times \mathbb{R}^m \times \Omega \to \mathbb{R}$, the function $f$ satisfy Assumption 1, and the functions $g_i$, $i \in I$, satisfy the same assumption. Suppose also that assumptions \([1] [2] [3] [4]\) of Theorem 3 are valid, and $(x_*, y_*)$ is a locally optimal solution of the problem $(P)$ such that $(x_*, y_*) \in S_c(\gamma)$ for some $c \geq c_*$, where $c_*$ is from Theorem 3.

Then for any measurable selection $(0, w_x(\cdot), w_y(\cdot))$ of the set-valued mapping $\mathcal{A}_{x,y}f(x_*, y_*(\cdot), \cdot)$ and any measurable selections $(0, w_{xx}(\cdot), w_{yy}(\cdot))$ of the multifunctions $\mathcal{A}_{x,y}g_i(x_*, y_*(\cdot), \cdot)$, $i \in I$, there exist $\zeta \in L^p(\Omega, \mathcal{A}, P; \mathbb{R}^d)$ and nonnegative multipliers $\lambda_i \in L^\infty(\Omega, \mathcal{A}, P)$, $i \in I$, such that $\mathbb{E}[\zeta] \in -N_A(x_*)$, $\sum_{i \in I} ||\lambda_i|| \leq c_*$, $\lambda_i(\omega)g_i(x_*, y_*(\omega), \omega) = 0$ for a.e. $\omega \in \Omega$ and all $i \in I$, and

$$(0, \zeta(\omega), 0) \in \mathcal{A}_{x,y}f(x_*, y_*(\cdot), \cdot) + (0, w_x(\cdot), w_y(\cdot)) + \sum_{i=1}^n \lambda_i(\omega) \left( \mathcal{A}_{x,y}g_i(x_*, y_*(\cdot), \cdot) + (0, w_{xx}(\cdot), w_{yy}(\cdot)) \right)$$

for a.e. $\omega \in \Omega$.

Proof. Under the assumptions of the theorem the functional $I$ is Lipschitz continuous on bounded sets by Corollary 2. Let

$$\varphi(x, y) = \int_\Omega \max_{i \in I} \{0, g_i(x, y, \omega)\} \; dP(\omega) \quad \forall (x, y) \in X.$$ 

Then by Theorem 3 the pair $(x_*, y_*)$ is a point of local minimum of the penalty function $\Phi_c$ on the set $A \times L^p(\Omega, \mathcal{A}, P)$ for any $c \geq c_*$, where $c_*$ is from Theorem 3. Thus, in particular, $(x_*, y_*)$ is a point of local minimum of the problem

$$\min_{(x,y)} J(x, y) = \int_\Omega f_0(x, y(\omega), \omega) \; dP(\omega) \quad \text{s.t.} \quad (x, y) \in A \times L^p(\Omega, \mathcal{A}, P; \mathbb{R}^m),$$

where $f_0 = f + c_\max \{0, g_i\}$. The function $f_0$ is codifferentiable in $(x, y)$, and applying codifferential calculus (see, e.g., [14]) one can compute its codifferent and verify that $f_0$ satisfies Assumption 1. Therefore by Corollary 4 the functional $J$ is directionally differentiable. Applying well-known necessary conditions for a minimum of a directionally differentiable function on a convex set (see, e.g., [14] Lemma V.1.2) and Corollary 4 one obtains that

$$J'(x_*, y_*, h_x, h_y) = \int_\Omega \left[ f_0(x_*, y_*(\omega), h_x, h_y(\omega)) \right] dP(\omega) \geq 0$$

for all $(h_x, h_y) \in (A - x_*) \times L^p(\Omega, \mathcal{A}, P; \mathbb{R}^m)$. Hence with the use of the standard calculus rules for directional derivatives (see [14] Sect. I.3) one gets that for all such $(h_x, h_y)$ the following inequality holds true:

$$J'(x_*, y_*, h_x, h_y) = \int_\Omega \left[ f(x_*, y_*(\omega), h_x, h_y(\omega)) + c_\max_{i \in I(\omega)} g_i(x_*, y_*(\omega), h_x, h_y(\omega)) \right] dP(\omega) \geq 0,$$
where \( g_0(x, y, \omega) \equiv 0 \) and
\[
\hat{I}(\omega) = \left\{ i \in I \cup \{ 0 \} \mid g_i(x_*, y_*, \omega, \omega) = \max_{i \in I} \left\{ 0, g_i(x_*, y_*, \omega, \omega) \right\} \right\}.
\]
Fix a measurable selection \((0, w_x(\cdot), w_y(\cdot))\) of the set-valued map \( \partial_x f(x_*, y_*, \cdot) \), for all \( i \in I \) fix any measurable selections \((0, w_{x_i}(\cdot), w_{y_i}(\cdot))\) of the set-valued maps \( \partial_x g_i(x_*, y_*, \cdot) \), and denote \((w_{x_0}(\cdot), w_{y_0}(\cdot)) \equiv 0\). Then by the definition of quasidifferential (Def. [3] and equality (2)) one has
\[
\int_{\Omega} \max_{(v_x, v_y) \in \partial_x f(x_*, y_*, v_\omega)} \left( (v_x + w_x(\omega), h_x) + (v_y + w_y(\omega), h_y(\omega)) \right) + c_* \max_{i \in I(\omega)} \left( (v_{x_i} + w_{x_i}(\omega), h_x) + (v_{y_i} + w_{y_i}(\omega), h_y(\omega)) \right) dP(\omega) \geq 0
\]
for all \((h_x, h_y) \in (A - x_*) \times L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m)\), where the last maximum is taken over all \((v_{x_i}, v_{y_i}) \in \partial_x g_i(x_*, y_*, \omega, \omega)\). Consequently, one has
\[
\int_{\Omega} \max_{(v_x, v_y) \in Q(\omega)} \left( (v_x, h_x) + (v_y, h_y(\omega)) \right) dP(\omega) \geq 0
\]
(21)
for all \((h_x, h_y) \in (A - x_*) \times L^p(\Omega, \mathfrak{A}, P; \mathbb{R}^m)\), where
\[
Q(\omega) = \partial_x f(x_*, y_*, \omega, \omega) + (w_x(\omega), w_y(\omega))
\]
\[
+ c_* \co \left\{ \partial_x g_i(x_*, y_*, \omega, \omega) + (w_{x_i}(\omega), w_{y_i}(\omega)) \mid i \in I(\omega) \right\}
\]
for any \( \omega \in \Omega \).

Let us show that the multifunction \( Q(\cdot) \) is measurable. Indeed, as was pointed out in the proof of Corollary [1] Assumption 1 guarantees that the set-valued mappings \( \partial_x f(x_*, y_*, \cdot) \) and \( \partial_x g_i(x_*, y_*, \cdot) \), \( i \in I \cup \{ 0 \} \), are measurable. Hence with the use of [33] Proposition 14.11, part (c) one gets that the set-valued mappings \( \partial_x f(x_*, y_*, \cdot) + (w_x(\cdot), w_y(\cdot)) \) and \( \partial_x g_i(x_*, y_*, \cdot) + (w_{x_i}(\cdot), w_{y_i}(\cdot)) \), \( i \in I \cup \{ 0 \} \), are measurable as well.

Arguing in the same way as in the proof of Theorem [4] one can easily check that the multifunction \( \hat{I}(\cdot) \) is measurable, which implies that the set-valued maps
\[
Q_i(\omega) := \begin{cases} \partial_x g_i(x_*, y_*, \omega, \omega) + (w_{x_i}(\omega), w_{y_i}(\omega)), & \text{if } i \in I(\omega), \\ \emptyset, & \text{if } i \notin I(\omega) \end{cases}
\]
are measurable for all \( i \in I \cup \{ 0 \} \). Therefore by [33] Prp. 14.11, part (b) and [1] Thrm. 8.2.2] the set-valued map
\[
\co \left( \bigcup_{i \in I \cup \{ 0 \}} Q_i(\cdot) \right) = \co \left\{ \partial_x g_i(x_*, y_*, \cdot) + (w_{x_i}(\cdot), w_{y_i}(\cdot)) \mid i \in I(\cdot) \right\}
\]
is measurable. Hence applying [33] Prp. 14.11, part (c) one finally gets that the multifunction \( Q(\cdot) \) is measurable.

Now, arguing in the same way as in the proof of Lemma [2] (or utilising the interchangeability principle; see, e.g. [33] Thrm. 14.60]) one gets that inequality (21) is satisfied iff
\[
\max_{(v_x(\omega), v_y(\cdot))} \int_{\Omega} \left( (v_x(\omega), h_x) + (v_y(\omega), h_y(\omega)) \right) dP(\omega) \geq 0
\]
(22)
for all \((h_x, h_y) \in (A - x_\ast) \times L^p(\Omega, \mathfrak{F}, P; \mathbb{R}^m)\), where the maximum is taken over all measurable selections of the multifunction \(Q(\cdot)\) (note that at least one such selection exists by [1, Thrm. 8.1.3]). From the definition of \(Q(\cdot)\) and the growth condition on the codifferentials of the functions \(f\) and \(g_i\) from Assumption 1 it follows that the set of all measurable selection of \(Q(\cdot)\) is a bounded subspace of the space \(L^1(\Omega, \mathfrak{F}, P; \mathbb{R}^d) \times L^{p'}(\Omega, \mathfrak{F}, P; \mathbb{R}^m)\). Therefore inequality (22) can be rewritten as follows:

\[
\max_{(v_1, v_2) \in Q(x_\ast, y_\ast)} \left( \langle v_1, h_x \rangle + \int_\Omega \langle v_y(\omega), h_y(\omega) \rangle \, dP(\omega) \right) \geq 0 \tag{23}
\]

for all \((h_x, h_y) \in (A - x_\ast) \times L^p(\Omega, \mathfrak{F}, P; \mathbb{R}^m)\), where

\[
Q(x_\ast, y_\ast) := \left\{ (v_1, v_2) \in \mathbb{R}^d \times L^{p'}(\Omega, \mathfrak{F}, P; \mathbb{R}^m) \left| v_1 = \mathbb{E}[v_x], \quad v_2 = v_y, \quad (v_x(\cdot), v_y(\cdot)) \text{ is a measurable selection of the map } Q(\cdot) \right\}
\]

The set \(Q(x_\ast, y_\ast)\) is bounded due to the boundedness of the set of all measurable selections of \(Q(\cdot)\). Furthermore, the set \(Q(x_\ast, y_\ast)\) is convex and closed, since by definition \(Q(\cdot)\) has closed and convex images. Therefore, \(Q(x_\ast, y_\ast)\) is a weakly compact convex subset of \(\mathbb{R}^d \times L^{p'}(\Omega, \mathfrak{F}, P; \mathbb{R}^m)\). Hence taking into account inequality (23) and applying the separation theorem one can easily check that

\[
Q(x_\ast, y_\ast) \cap \left\{ (I - N_A(x_\ast)) \times \{0\} \right\} \neq \emptyset.
\]

Consequently, by the definitions of \(Q(x_\ast, y_\ast)\) and \(Q(\cdot)\) there exists a function \(\zeta \in L^1(\Omega, \mathfrak{F}, P; \mathbb{R}^d)\) such that \(\mathbb{E}[\zeta] \in -N_A(x_\ast)\) and

\[
(\zeta(\omega), 0) \in \partial_{x_\ast, y_\ast} f(x_\ast, y_\ast(\omega), \omega) + (w_x(\omega), w_y(\omega)) + c, \co \left\{ \partial_{x_\ast, y_\ast} g_i(x_\ast, y_\ast(\omega), \omega) + (w_{x_i}(\omega), w_{y_i}(\omega)) \left| i \in I(\omega) \right\}
\]

for a.e. \(\omega \in \Omega\).

Let \(E_J = \{ \omega \in \Omega \mid I(\omega) = J \}\) for any nonempty subset \(J \subseteq I \cup \{0\}\). The sets \(E_J\) form a partition of \(\Omega\). Moreover, these sets are measurable, since the multifunction \(I(\cdot)\) is measurable.

Observe that from (24) it follows that

\[
(\zeta(\omega), 0) \in \partial_{x_\ast, y_\ast} f(x_\ast, y_\ast(\omega), \omega) + (w_x(\omega), w_y(\omega)) + c, \co \left\{ \partial_{x_\ast, y_\ast} g_i(x_\ast, y_\ast(\omega), \omega) + (w_{x_i}(\omega), w_{y_i}(\omega)) \left| i \in J \right\}
\]

for any \(\omega \in E_J\) and any nonempty \(J \subseteq I \cup \{0\}\). With the use of the Filippov theorem (see, e.g. [1, Thrm. 8.2.10]) one can readily check that the previous inclusion implies that for any nonempty \(J \subseteq I \cup \{0\}\) there exist nonnegative measurable functions \(\alpha_i(\cdot),\ i \in J\), such that \(\sum_{i \in J} \alpha_i(\omega) = 1\) and

\[
(\zeta(\omega), 0) \in \partial_{x_\ast, y_\ast} f(x_\ast, y_\ast(\omega), \omega) + (w_x(\omega), w_y(\omega)) \]

\[
+ c, \sum_{i \in J} \alpha_i(\omega) \left( \partial_{x_\ast, y_\ast} g_i(x_\ast, y_\ast(\omega), \omega) + (w_{x_i}(\omega), w_{y_i}(\omega)) \right)
\]

for any \(\omega \in E_J\) and any nonempty \(J \subseteq I \cup \{0\}\).
for a.e. \( \omega \in E_J \). For any \( i \in I \) define

\[
\lambda_i(\omega) = \begin{cases} 
  c_i \alpha_i^*(\omega), & \text{if } \omega \in E_J, \ i \in J \ (\text{or, equivalently, } i \in \hat{I}(\omega)), \\
  0, & \text{otherwise.}
\end{cases}
\]

Observe that by definition \( \lambda_i \), \( i \in I \), are nonnegative measurable functions such that \( \sum_{i \in I} \| \lambda_i \|_{\infty} \leq c_* \), and \( \lambda_i(\omega)g_i(x_*, y_*(\omega), \omega) = 0 \) for a.e. \( \omega \in \Omega \), since \( \lambda_i(\omega) = 0 \) whenever \( i \notin \hat{I}(\omega) \), i.e. \( g_i(x_*, y_*(\omega), \omega) < 0 \). Furthermore, bearing in mind the fact that \( w_{x_0}(\cdot) \equiv 0 \), \( w_{y_0}(\cdot) \equiv 0 \), and \( \partial_{x,y} g_0(x_*, y_*(\omega), \omega) \equiv \{0\} \) one gets that

\[
\left( \zeta(\omega), 0 \right) \in \partial_{x,y} f(x_*, y_*(\omega), \omega) + (w_x(\omega), w_y(\omega))
\]

\[
+ \sum_{i \in I} \lambda_i(\omega) \left( \partial_{x,y} g_i(x_*, y_*(\omega), \omega) + (w_{x_i}(\omega), w_{y_i}(\omega)) \right).
\]

for a.e. \( \omega \in \Omega \). Hence applying equality (2) we arrive at the required result. \( \square \)

Remark 7. It should be noted that with the use of the co-differential calculus one can compute a co-differential of the function \( f_0 \) from the proof of the previous theorem, apply necessary conditions for a minimum of a codifferentiable function on a convex set [17, Thrm. 2.8] to the functional \( J \), and then directly rewrite these conditions in terms of the problem \( (P) \) with the use of Theorem 1 and an explicit expression for a codifferential of \( f_0 \). However, one can check that this approach leads to more cumbersome optimality conditions than the ones from the theorem above. It is possible to verify that these conditions are equivalent, but in the author’s opinion the proof of this equivalence is more difficult than the proof of the previous theorem. That is why we chose to present a simpler, but somewhat indirect derivation of optimality conditions for the problem \( (P) \).

Remark 8. Note that in the case when the functions \( f \) and \( g_i \), \( i \in I \), are differentiable jointly in \( x \) and \( y \), the optimality conditions from Theorem 1 take the following well-known form (cf. [26, 36, 40, 45, 46]). There exist nonnegative multipliers \( \lambda_i \in L^\infty(\Omega, \mathfrak{A}, P), i \in I \), such that \( \lambda_i(\omega)g_i(x_*, y_*(\omega), \omega) = 0 \) for a.e. \( \omega \in \Omega \) and all \( i \in I \), and

\[
\left\{ \mathbb{E} \left[ \nabla_x f(x_*, y_*(\cdot), \cdot) + \sum_{i \in I} \lambda_i(\cdot) \nabla_x g_i(x_*, y_*(\cdot), \cdot) \right], x - x_* \right\} \geq 0 \quad \forall x \in A,
\]

\[
\nabla_y f(x_*, y_*(\omega), \omega) + \sum_{i \in I} \lambda_i(\omega) \nabla_y g_i(x_*, y_*(\omega), \omega) = 0 \quad \text{for a.e. } \omega \in \Omega.
\]

5 Conclusions

This work was devoted to an analysis of nonsmooth two-stage stochastic programming problems with the use of tools of constructive nonsmooth analysis [13]. In the first part of the paper, we analysed the co-/quasi-differentiability of the expectation of nonsmooth random integrands and obtained explicit formulae for its co-/quasi-differentials under some natural measurability and growth conditions on the integrand and its codifferential.

In the second part of the paper, we obtained two types of sufficient conditions for the global exactness of a penalty function for two-stage stochastic programming problems, reformulated as equivalent variational problems with pointwise
constraints. The first type of sufficient conditions is formulated for the penalty term defined via the $L^p$ norm of the distance to the feasible set of the second stage problem, while the second type of sufficient conditions is formulated for the penalty term that is independent of $p$ and is defined via the constraints of the second stage problems. Although the second type of sufficient conditions is much more restrictive than the first one, it is more convenient for applications and derivation of optimality conditions. Furthermore, as is pointed out in Remark 6, these conditions open a way for the derivation of global optimality conditions and application of DC optimization methods to two stage stochastic programming problems, whose second stage problem has DC objective function and DC constraints.

Finally, in the last part of the paper we combined our results on codifferentiability of the expectation of nonsmooth random integrands and exact penalty functions to derive optimality conditions for nonsmooth two-stage stochastic programming problems in terms of codifferentials, involving essentially bounded Lagrange multipliers.

References

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

[2] G. Barbarosoglu and Y. Arda. A two-stage stochastic programming framework for transportation planning in disaster response. J. Oper. Res. Soc., 55:43–53, 2004.

[3] J. R. Birge and F. Louveaux. Introduction to Stochastic Programming. Springer, New York, 2011.

[4] V. I. Bogachev. Measure Theory. Volume I. Springer-Verlag, Berlin, Heidelberg, 2007.

[5] J. V. Burke. The subdifferential of measurable composite max integrands and smoothing approximation. Math. Program., 181:229–264, 2020.

[6] X. Chen and M. Fukushima. Expected residual minimization method for stochastic linear complementarity problems. Math. Oper. Res., 30:916–638, 2005.

[7] X. Chen, R. J.-B. Wets, and Y. Zhang. Stochastic variational inequalities: residual minimization smoothing sample average approximations. SIAM J. Optim., 22:649–673, 2012.

[8] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova, editors. Bilevel Programming Problems. Theory, Algorithms and Applications to Energy Networks. Springer, Berlin, Heidelberg, 2015.

[9] S. Dempe and A. Zemkoho, editors. Bilevel Optimization. Advances and Next Challenges. Springer, Cham, 2020.

[10] V. F. Demyanov. Conditions for an extremum in metric spaces. J. Glob. Optim., 17:55–63, 2000.
[11] V. F. Demyanov. Nonsmooth optimization. In G. Di Pillo and F. Schoen, editors, *Nonlinear optimization. Lecture notes in mathematics*, vol. 1989, pages 55–163. Springer-Verlag, Berlin, 2010.

[12] V. F. Demyanov and L. C. W. Dixon, editors. *Quasidifferential Calculus*. Springer, Berlin, Heidelberg, 1986.

[13] V. F. Dem’yanov and V. N. Malozemov. *Introduction to Minimax*. Dover Publications, New York, 2014.

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

[15] V. F. Demyanov and A. M. Rubinov, editors. *Quasidifferentiability and Related Topics*. Kluwer Academic Publishers, Dordrecht, 2000.

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

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

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

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

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

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

[22] M. V. Dolgopolik and A. Fominyh. Exact penalty functions for optimal control problems I: main theorem and free-endpoint problems. *Optim. Control Appl. Meth.*, 40:1018–1044, 2019.

[23] C. I. Fábián and Z. Szőke. Solving two-stage stochastic programming problems with level decomposition. *Comput. Manag. Sci.*, 4:313–353, 2007.

[24] S. D. Flåm and J. Zowe. Exact penalty functions in single-stage stochastic programming. *Optim.*, 21:723–734, 1990.

[25] E. Grass and K. Fischer. Two-stage stochastic programming in disaster management: a literature survey. *Surv. Oper. Res. Manag. Sci.*, 21:85–100, 2016.

[26] J. B. Hiriart-Urruty. Conditions nécessaires d’optimalité pour un programme stochastique avec recours. *SIAM J. Control Optim.*, 16:317–329, 1978.

[27] G. H. Huang and D. P. Loucks. An inexact two-stage stochastic programming model for water resources management under uncertainty. *Civ. Eng. Environ. Syst.*, 17:95–118, 2000.
[28] H. Leövey and W. Römisch. Quasi-Monte Carlo methods for linear two-stage stochastic programming problems. Math. Program., 151:315–345, 2015.

[29] S. Lin, M. Huang, Z. Xia, and D. Li. Quasidifferentiabilities of the expectation functions of random quasidifferentiable functions. Optim., pages 1–16, 2020. DOI: 10.1080/02331934.2020.1818235.

[30] C. Liu, Y. Fan, and F. Ordóñez. A two-stage stochastic programming model for transportation network protection. Comput. Oper. Res., 36:1582–1590, 2009.

[31] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. SIAM J. Optim., 19:1574–1609, 2009.

[32] W. Oliveira, C. Sagastizábal, and S. Scheimberg. Inexact bundle methods for two-stage stochastic programming. SIAM J. Optim., 21:517–544, 2011.

[33] A. V. Orlov. On a solving bilevel D.C.-convex optimization problems. In Y. Kochetov, I. Bukadorov, and T. Gruzdeva, editors, Mathematical Optimization Theory and Operations Research. MOTOR 2020., pages 179–191. Springer, Cham, 2020.

[34] G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. SIAM J. Control Optim., 27:1333–1360, 1989.

[35] R. T. Rockafellar and R. J.-B. Wets. Variational Analysis. Springer-Verlag, Berlin, Heidelberg, 1998.

[36] R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: Kuhn-Tucker conditions. J. Math. Econ., 2:349–370, 1975.

[37] R. T. Rockafellar and R. J.-B. Wets. On the interchange of subdifferentiation and conditional expectation for convex functions. Stochastics, 7:173–182, 1982.

[38] A. Rubinov and X. Yang. Lagrange-Type Functions in Constrained Non-Convex Optimization. Kluwer Academic Publishers, Boston, 2003.

[39] A. Shapiro and T. H. de Mello. A simulation-based approach to two-stage stochastic programming with recourse. Math. Program., 81:301–325, 1998.

[40] A. Shapiro, D. Dentcheva, and A. Ruszczynski. Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia, 2014.

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

[42] A. S. Strekalovsky and A. V. Orlov. Global search for bilevel optimization with quadratic data. In S. Dempe and A. Zemkoho, editors, Bilevel Optimization, pages 313–334. Springer, Cham, 2020.

[43] A. Uderzo. On the variational behaviour of functions with positive steepest descent rate. Positivity, 19:725–745, 2015.
[44] A. Uderzo. A strong metric subregularity analysis of nonsmooth mappings via steepest displacement rate. *J. Optim. Theory Appl.*, 171:573–599, 2016.

[45] S. Vogel. Necessary optimality conditions for two-stage stochastic programming problems. *Optim.*, 16:607–616, 1985.

[46] H. Xu and J. J. Ye. Necessary optimality conditions for two-stage stochastic programs with equilibrium constraints. *SIAM J. Optim.*, 20:1685–1715, 2010.

[47] H. Xu and D. Zhang. Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. *Math. Program.*, 119:371–401, 2009.

[48] A. J. Zaslavski. *Optimization on Metric and Normed Spaces*. Springer, New York, 2010.

[49] Z. Zhou, J. Zhang, P. Liu, Z. Li, M. C. Georgiadis, and E. N. Pistikopoulos. A two-stage stochastic programming model for the optimal design of distributed energy systems. *Appl. Energy*, 103:135–144, 2013.