On optimality conditions for some nonsmooth optimization problems over L^p spaces

J. V. Outrata* and W. Römisch†

Abstract. The paper deals with the minimization of an integral functional over an L^p space subject to various types of constraints. For such optimization problems new necessary optimality conditions are derived, based on several concepts of nonsmooth analysis. In particular, we employ the generalized differential calculus of Mordukhovich and the fuzzy calculus of proximal subgradients. The results are specialized to nonsmooth two-stage and multistage stochastic programs.

Keywords: Normal integrand, integral functional, normal cone, subdifferential, fuzzy calculus, coderivative, stochastic programming, two-stage, multistage

2000 Mathematics Subject Classification: 49J52, 49K45, 90C15

Running Title: On nonsmooth optimization problems over L^p spaces

1 Introduction

In [20, Chapter 14] the authors propose an effective treatment of a class of optimization problems with integral objectives. Let $(\Omega, \mathcal{S}, \mu)$ be a positive complete measure space with $\mu(\Omega) < \infty$. The approach of [20] can well be applied, under some assumptions, to the minimization of certain integral functionals subject to the constraints

$$\begin{aligned} & x(s) \in \Gamma(s) \quad \text{for a. e. } s \in \Omega, \\ & x \in L^p(\Omega; I\!R^n), \end{aligned} \tag{1.1}$$

where $\Gamma[\Omega \rightsquigarrow I\!R^n]$ is a given multifunction. The key idea consists in the interchange of minimization and integration by which one converts the original infinite-dimensional problem to a family of standard finite-dimensional programs, parameterized by $s \in \Omega$. We respect to optimality conditions, these finite-dimensional programs can be treated in different ways. In Section 3 we apply to this purpose the tools of the generalized differential calculus of Mordukhovich ([10, 11, 12, 13, 14]) and obtain in this way sharp optimality conditions for the original problem. We study also a specific structure of Γ and the properties of the respective Karush–Kuhn–Tucker (KKT) mappings (measurability, integrability). In the second part of Section 3 we apply the same technique to a two-stage stochastic program. This leads to substantially sharper optimality conditions in comparison with [6] and [21], where the conditions have been derived on the basis of the generalized differential calculus of Clarke ([3]).

*Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 18208 Prague, Czech Republic (Email: outrata@utia.cas.cz)

†Humboldt-University Berlin, Institute of Mathematics, 10099 Berlin, Germany (Email: romisch@mathematik.hu-berlin.de)

Unfortunately, the situation substantially changes whenever our optimization problem contains, besides the constraints (1.1), also a non-pointwise constraint. The interchange of minimization and integration is then generally not possible and so the derivation of convenient pointwise optimality conditions becomes substantially more difficult. Such conditions have been derived in [7] in terms of the Clarke generalized differential calculus under some additional assumptions imposed on the non-pointwise constraint. Trying to weaken these additional assumptions as much as possible, we have confined ourselves to the Hilbert space $L^2(\Omega; I\!\!R^n)$ and employed the so-called fuzzy calculus of proximal subdifferentials. This approach leads to fuzzy optimality conditions without any constraint qualifications coupling the pointwise and non-pointwise constraints. The satisfaction of such constraint qualifications happens namely to be the most serious hurdle on the way to classical KKT conditions, at least in the reflexive L^p spaces. The obtained fuzzy optimality conditions have the desired pointwise nature and are in a certain sense the sharpest possible. In the second part of Section 4 we specialize these conditions to nonsmooth multistage stochastic programs.

The Appendix contains three statements from the generalized differential calculus of Mordukhovich which play an essential role in the developments of Section 3. For further useful results of this kind the interested reader is referred to the cited works of this author.

The following notation is employed: $\mathrm{cl}\, A$ is the closure of a set A and $I\!\!B(a;\rho)$ is the closed ball with the center at a and radius ρ. If $a = 0$ and $\rho = 1$, we write simply $I\!\!B$. $\overline{I\!\!R}$ is the extended real line and, for a function $f[X \to \overline{I\!\!R}]$, $\mathrm{epi}\, f$ denotes its epigraph and $\overline{\partial} f(x)$ denotes the Clarke subdifferential of f at x. If F maps $I\!\!R^n$ into $I\!\!R^m$, then $\overline{\partial} F(x)$ is the Clarke generalized Jacobian of F at x. Analogously, $\overline{N}_A(x)$ denotes the Clarke normal cone to A at x. For a multifunction $\Phi[X \rightsquigarrow Y]$, $\mathrm{gph}\, \Phi = \{(x, y) \mid y \in \Phi(x)\}$. $\delta_C(\cdot)$ denotes the indicatory functional of a set C and $\mathrm{dist}(x|C)$ is the distance from x to C. $|\cdot|$ is a norm in $I\!\!R^n$ whereas $\|\cdot\|$ is the norm in a considered function space.

2 Problem formulation and preliminaries

Consider a function f mapping $\Omega \times I\!\!R^n$ into $\overline{I\!\!R}$. Following [20] f is called a *normal integrand*, provided its epigraphical mapping

$$\mathcal{E}_f(s) := \mathrm{epi}\, f(s, \cdot) = \{(y, \alpha) \in I\!\!R^n \times I\!\!R \mid f(s, y) \le \alpha\}$$

is closed-valued and measurable. We recall that a multifunction $\Theta[\Omega \rightsquigarrow I\!\!R^n]$ is *measurable* if for every closed set $\mathcal{O} \subset I\!\!R^n$ the set $\Theta^{-1}(\mathcal{O})$ is measurable, i. e., $\Theta^{-1}(\mathcal{O}) \in \mathcal{S}$. In particular, the set $\mathrm{dom}\, \Theta = \Theta^{-1}(I\!\!R^n)$ must be measurable. For a comprehensive treatment of measurable multifunctions we refer the reader e. g. to [1] or [20]. In what follows we will adopt the notation of [20], where $\int_\Omega f(s, x(s))\, \mu(\mathrm{d}s)$ is denoted by $I_f[x]$. The next section concerns essentially the optimization problems of the form

$$\begin{array}{ll} \text{minimize} & I_f[x] \\ \text{subject to} & x(s) \in \Gamma(s) \quad \text{for a. e. } s \in \Omega \\ & x \in L^p(\Omega; I\!\!R^n), \end{array} \tag{2.1}$$

where f is a normal integrand, Γ is a closed-valued and measurable multifunction and $1 \le p \le \infty$. In Section 4 we add to the constraints of (2.1) another one, namely

$$x \in C,$$

where C is a closed subset of $L^2(\Omega; I\!R^n)$, not expressible in the pointwise form.

For the reader's convenience, we give now the definitions of those basic concepts from nonsmooth analysis which will frequently be used throughout the both subsequent sections.

Consider an arbitrary set $\Pi \subset I\!R^p$.

Definition 2.1. Let $a \in \operatorname{cl} \Pi$. The cone

$$\widehat{N}_\Pi(a) := \left\{ x^* \in I\!R^p \,\middle|\, \limsup_{x \xrightarrow{\Pi} a} \frac{\langle x^*, x - a\rangle}{|x - a|} \leq 0 \right\}$$

is called the *Fréchet normal cone* to Π at a. The *limiting normal cone* (or *Mordukhovich normal cone*) to Π at a is defined by

$$N_\Pi(a) = \limsup_{a' \xrightarrow{\Pi} a} \widehat{N}_\Pi(a'). \tag{2.2}$$

The "lim sup" in (2.2) is the upper limit of multifunctions in the sense of Kuratowski-Painlevé. Generally, $N_\Pi(a)$ is a nonconvex cone, but the cone-valued multifunction $N_\Pi(\cdot)$ is upper semicontinuous at each point of $\operatorname{cl}\Pi$ (with respect to $\operatorname{cl}\Pi$). This is essential in the calculus of Mordukhovich subdifferentials and coderivatives introduced below.

Definition 2.2. Let $\varphi[I\!R^p \to \overline{I\!R}]$ be an arbitrary extended real-valued function and $a \in \operatorname{dom}\varphi$. The sets

$$\partial\varphi(a) := \{a^* \in I\!R^p \mid (a^*, -1) \in N_{\operatorname{epi}\varphi}(a, \varphi(a))\}$$

and

$$\partial^\infty\varphi(a) := \{a^* \in I\!R^p \mid (a^*, 0) \in N_{\operatorname{epi}\varphi}(a, \varphi(a))\}$$

are called the *limiting (Mordukhovich) subdifferential* and the *singular subdifferential* of φ at a.

In [10] it has been shown that

$$\overline{\partial}\varphi(a) = \operatorname{cl}\operatorname{conv}(\partial\varphi(a))$$

and for φ Lipschitz near a the closure operation is superfluous. Moreover, a lower semicontinuous function φ is Lipschitz near a if and only if $\partial^\infty\varphi(a) = \{0\}$.

Definition 2.3. Let $\Phi[I\!R^p \rightsquigarrow I\!R^q]$ be a multifunction and $b \in \Phi(a)$. The multifunction $D^*\Phi(a, b)\,[I\!R^q \to I\!R^p]$ defined by

$$D^*\Phi(a, b)\,(b^*) := \{a^* \in I\!R^p \mid (a^*, -b^*) \in N_{\operatorname{gph}\Phi}(a, b)\}\,, \quad b^* \in I\!R^q$$

is called the *coderivative* of Φ at (a, b). If Φ is single-valued, one uses the notation $D^*\Phi(a)\,(b^*)$.

In Hilbert spaces a useful concept of normality is provided by the following construction. Consider a Hilbert space H and its arbitrary subset, denoted again by Π.

Definition 2.4. ([18]) Let $a \in \operatorname{cl}\Pi$. The vector $x^* \in H$ is called a *proximal normal direction* to Π at a, provided there exists $k = k(x^*, a) \geq 0$ such that

$$\langle x^*, x - a \rangle \leq k \|x - a\|^2 \quad \text{for all } x \in \Pi.$$

The set of all proximal normal directions to Π at a is termed the *proximal normal cone* to Π at a and is denoted by $N^P_\Pi(a)$.

It is easy to see that for $H = \mathbb{R}^n$ one has $N^P_\Pi(a) \subset \widehat{N}_\Pi(a)$. As explained e.g. in [9], however, the limiting normal cone has been originally defined via the proximal normal cone and (2.2) holds true with $\widehat{N}_\Pi(a')$ replaced by $N^P_\Pi(a')$.

On the basis of the proximal cone one can introduce the proximal subdifferential in the standard way.

Definition 2.5. ([18]) Let $\varphi[H \to \overline{\mathbb{R}}]$ be lower semicontinuous (lsc) at $a \in \operatorname{dom}\varphi$. The set

$$\partial_P \varphi(a) := \left\{ a^* \in H \middle| (a^*, -1) \in N^P_{\operatorname{epi}\varphi}(a, \varphi(a)) \right\}$$

is called the *proximal subdifferential* of φ at a.

Differently from the limiting subdifferential, $\partial_P \varphi(a)$ is convex, but not necessarily closed. It can well be empty, even for φ being Lipschitz near a. The advantage of $\partial_P \varphi(a)$ over some similar constructions like the Fréchet or Dini subdifferentials consists above all in handling integral functionals. For example in $L^2(\Omega; \mathbb{R}^n)$ one has the implication

$$v \in \partial_P I_f[x] \Longrightarrow v(s) \in \partial_P f(s, x(s)) \quad \text{for a. e. } s \in \Omega, \tag{2.3}$$

where the subdifferential of f is computed in the second argument only (cf. [4, 8]).

3 Pointwise constraints

This section is devoted to optimization problems over L^p spaces in which one has to do solely with various pointwise constraints. Our workhorse is the interchange of minimization and integration as stated e.g. in [20, Theorem 14.60]. So we start with the optimization problem (2.1) under the assumptions posed in the previous section and introduce the *essential integrand* $\widetilde{f}[\Omega \times \mathbb{R}^n \to \overline{\mathbb{R}}]$ by

$$\widetilde{f}(s, x) := f(s, x) + \delta_{\Gamma(s)}(x). \tag{3.1}$$

In this way (2.1) amounts to the minimization of $I_{\widetilde{f}}$ over $L^p(\Omega; \mathbb{R}^n)$.

Theorem 3.1. Let $I_{\widetilde{f}}$ be proper on $L^p(\Omega; \mathbb{R}^n)$, i. e., $I_{\widetilde{f}}[x] > -\infty$ for all $x \in L^p(\Omega; \mathbb{R}^n)$ and $I_{\widetilde{f}}[x_0] < \infty$ for some $x_0 \in L^p(\Omega; \mathbb{R}^n)$.

Then the following statements are equivalent:

(i) $\widehat{x} \in \operatorname{argmin}_{x \in L^p(\Omega; \mathbb{R}^n)} I_{\widetilde{f}}[x]$;

(ii) $\widehat{x}(s) \in \operatorname{argmin}_{y \in \mathbb{R}^n} \widetilde{f}(s, y)$ for a. e. $s \in \Omega$.

Proof. The assumptions posed on f and Γ imply that $\widetilde{f}$ is a normal integrand. Indeed, for the epigraphical mapping $\mathcal{E}_{\widetilde{f}}$ one has

$$\mathcal{E}_{\widetilde{f}}(s) = \operatorname{epi} \widetilde{f}(s,\cdot) = \operatorname{epi} f(s,\cdot) \cap (\Gamma(s) \times \mathbb{R}_+).$$

Since intersections and products preserve measurability ([20, Proposition 14.11]), the normality of $\widetilde{f}$ is implied by the normality of f and the closed-valuedness and measurability of Γ. The L^p spaces are decomposable [20, Definition 14.59] and so the mentioned Theorem 14.60 from [20] can be specialized to the above form. □

Remark. In most cases f is a Carathéodory integrand (i. e. $f(s,y)$ is measurable in s for each y and continuous in y for each s) and all problem constraints are comprised in Γ. The considered structure enables, however, to distinguish between "implicit constraints", expressed via f and "explicit constraints" modeled by Γ.

3.1 Optimality conditions

Theorem 3.1 enables us to invoke Theorem A.1 and formulate immediately the optimality conditions for (2.1).

Theorem 3.2. Let $\widehat{x}$ be a (local) solution of (2.1) and let the assumptions of Theorem 3.1 be fulfilled. Then, for a. e. $s \in \Omega$ there exist a vector $\widehat{v}_s^*$ and a real $\lambda_s \geq 0$, not both simultaneously equal to zero, such that

$$(\widehat{v}_s^* - \lambda_s) \in N_{\operatorname{epi} f(s,\cdot)}(\widehat{x},\ f(s,\widehat{x}(s)))\,, \quad -\widehat{v}_s^* \in N_{\Gamma(s)}(\widehat{x}(s)). \tag{3.2}$$

If the constraint qualification

$$\partial^\infty f\,(s,\widehat{x}(s)) \cap \Big(-N_{\Gamma(s)}(\widehat{x}(s))\Big) = \{0\} \quad \text{for a. e. } s \in \Omega \tag{3.3}$$

is fulfilled, then for a. e. $s \in \Omega$ one has $\lambda_s > 0$ and

$$0 \in \partial f\,(s,\widehat{x}(s)) + N_{\Gamma(s)}(\widehat{x}(s)). \tag{3.4}$$

The subdifferentials in (3.3), (3.4) concern the function $f(s,\cdot)$. The form of the above conditions is illustrated by the following simple academic example.

Example 3.1. Consider the optimization problem with the constraint set of a shape similar to the "Mercedes star":

$$\begin{aligned}
&\text{minimize} \quad \int_0^1 (|x_1(s)| + |x_2(s)|)\,\mathrm{d}s \\
&\text{subject to} \\
&\qquad x(s) \in \{y \in \mathbb{R}^2 \mid k(s) - |y_1| - y_2 = 0\} \cup \{y \in \mathbb{R}^2 \mid y_1 = 0,\ y_2 \geq k(s)\} \\
&\qquad\qquad\qquad\qquad \text{a. e. in } [0,1], \\
&\qquad x \in L^2(0,1;\mathbb{R}^2),
\end{aligned} \tag{3.5}$$

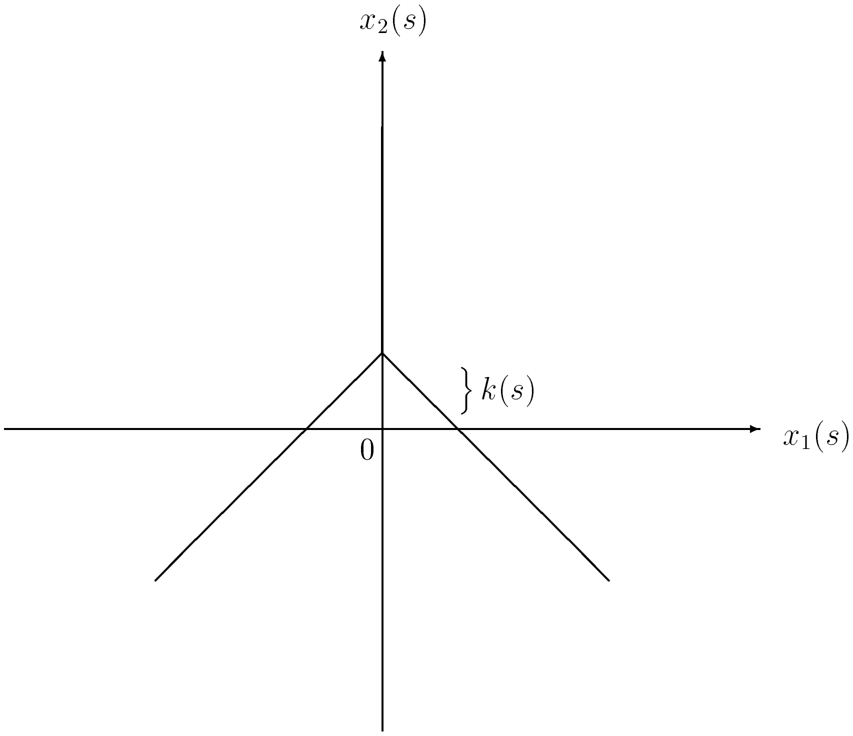

Figure 1: $\Gamma(s)$ from Example 3.1.

where $k \in L^2(0,1;\mathbb{R}^2)$ is a given function. We note that $\widehat{x}(\cdot) = (0, \max\{0, k(\cdot)\})$ is a solution of (3.5). It is not difficult to check that all assumptions of Theorem 3.1 are fulfilled.

Further, in this simple situation the constraint qualification (3.3) is fulfilled and the normal cones $N_{\Gamma(s)}(\widehat{x}(s))$ as well as the subdifferentials $\partial f(s, \widehat{x}(s))$ can easily be computed.

$$\text{If} \quad k(s) \geq 0, \quad \text{then} \quad N_{\Gamma(s)}(\widehat{x}(s)) = \left\{ \lambda \begin{bmatrix} -1 \\ 1 \end{bmatrix} \middle| \lambda \in \mathbb{R} \right\} \cup \left\{ \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} \middle| \lambda \in \mathbb{R} \right\} \cup \left\{ \lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix} \middle| \lambda \in \mathbb{R} \right\} \text{ a.e.,}$$

$$\text{if} \quad k(s) < 0, \quad \text{then} \quad N_{\Gamma(s)}(\widehat{x}(s)) = \left\{ \lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix} \middle| \lambda \in \mathbb{R} \right\} \quad \text{a.e.}$$

Furthermore,

$$\text{if} \quad k(s) > 0, \quad \text{then} \quad \partial f(\widehat{x}(s)) = \begin{bmatrix} [-1, \ 1] \\ 1 \end{bmatrix}, \text{ a.e., and}$$

$$\text{if} \quad k(s) \leq 0, \quad \text{then} \quad \partial f(\widehat{x}(s)) = \begin{bmatrix} [-1, \ 1] \\ [-1, \ 1] \end{bmatrix}, \text{ a.e.}$$

The optimality condition (3.4) is thus evidently fulfilled, because e.g. the function

$\hat{v}^*[[0,1] \to I\!R^2]$ defined by

$$\hat{v}^*(s) = \left\langle \begin{array}{ll} \begin{bmatrix} -1 \\ 1 \end{bmatrix} & \text{if} \quad k(s) \geq 0 \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if} \quad k(s) < 0 \end{array} \right. \qquad \text{a. e. on } [0,1]$$

fulfills the relation

$$\hat{v}^*(s) \in \partial f(\hat{x}(s)) \cap \Big(-N_{\Gamma(s)}(\hat{x}(s))\Big) \quad \text{for a. e. } s \in [0,1].$$

Since the function $\hat{v}^*$ from the above example clearly belongs to $L^2(0,1;I\!R^2)$, one can think about a strengthening of the conditions in Theorem 3.2 along the lines of Theorem 2.2 in [7].

Theorem 3.3. Let $\hat{x}$ be a solution of (2.1) and let the assumptions of Theorem 3.1 be fulfilled. Further assume that $p < \infty$ and that there exists a function $k \in L^q(\Omega)$ $\left(\frac{1}{p} + \frac{1}{q} = 1\right)$ such that, for all $s \in \Omega$,

$$|f(s,y_1) - f(s,y_2)| \leq k(s)\,|y_1 - y_2| \quad \text{for all } y_1,\, y_2 \text{ in } I\!R^n. \tag{3.6}$$

Then there exists a function $\hat{v}^* \in L^q(\Omega; I\!R^n)$ such that

$$\hat{v}^*(s) \in \partial f(s,\, \hat{x}(s)) \cap \Big(-N_{\Gamma(s)}(\hat{x}(s))\Big) \quad \text{for a. e. } s \in \Omega. \tag{3.7}$$

Proof. We start with the observation that, due to (3.6), $f(s,\cdot)$ is Lipschitz on $I\!R^n$ for all $s \in \Omega$ and consequently $\partial^\infty f(s,\,\hat{x}(s)) = \{0\}$ for a. e. $s \in \Omega$. This implies that the constraint qualification (3.3) is fulfilled. Further, by virtue of (3.6), for all $s \in \Omega$ one has

$$\partial f(s,\,\hat{x}(s)) \subset \overline{\partial} f(s,\,\hat{x}(s)) \subset k(s)\, I\!B$$

so that each measurable selection of $\partial f(s,\,\hat{x}(s))$ (a. e. on Ω) belongs to $L^q(\Omega; I\!R^n)$. Thus it remains to show that there is a measurable function $\hat{v}^*[\Omega \to I\!R^n]$ satisfying condition (3.7). This follows, however, immediately from [20, Proposition 14.11] (measurability of intersections) and [20, Theorems 14.26 and 14.56] dealing with the measurability of the multifunctions $N_{\Gamma(\cdot)}(\hat{x}(\cdot))$ and $\partial f(\cdot, \hat{x}(\cdot))$, respectively. The proof is complete. □

Let us now specify the above optimality conditions for the case where

$$\Gamma(s) = \{y \in \Theta(s) \,|\, G(s,y) \in \Lambda(s)\} \tag{3.8}$$

with $G[\Omega \times I\!R^n \to I\!R^m]$ being a Carathéodory map and multifunctions $\Theta[\Omega \rightsquigarrow I\!R^n]$, $\Lambda[\Omega \rightsquigarrow I\!R^m]$ closed-valued and measurable. As shown in [1, Theorem 8.2.9], under these conditions Γ is also closed-valued and measurable.

Theorem 3.4. Let $\widehat{x}$ be a solution of (2.1), where Γ is given in the form (3.8). Further suppose that the assumptions of Theorem 3.1 are fulfilled, for a. e. $s \in \Omega$ $f(s,\cdot)$ is Lipschitz around $\widehat{x}(s)$, and the qualification conditions

$$\left.\begin{aligned} D^*G(s,\widehat{x}(s)) \circ N_{\Lambda(s)}(G(s,\widehat{x}(s))) \cap (-N_{\Theta(s)}(\widehat{x}(s))) = \{0\} \\ N_{\Lambda(s)}(G(s,\widehat{x}(s))) \cap Ker\ D^*G(s,\widehat{x}(s)) = \{0\} \end{aligned}\right\} \tag{3.9}$$

hold true. Then for a. e. $s \in \Omega$ there is a vector $\widehat{y}^*_s \in N_{\Lambda(s)}(G(s,\,\widehat{x}(s)))$ such that

$$0 \in \partial f(s,\,\widehat{x}(s)) + D^*G(s,\,\widehat{x}(s))\,(\widehat{y}^*_s) + N_{\Theta(s)}(\widehat{x}(s)). \tag{3.10}$$

Remark. Analogously to Theorem 3.2, the coderivatives in (3.9), (3.10) concern the map $G(s,\cdot)$.

Proof of Theorem 3.4. Under the qualification conditions (3.9) Theorem A.2 provides us with an upper approximation of $N_{\Gamma(s)}(\widehat{x}(s))$ in the form

$$N_{\Gamma(s)}(\widehat{x}(s)) \subset \Big\{v^* \in D^*G(s,\,\widehat{x}(s))\,(\xi) + N_{\Theta(s)}(\widehat{x}(s))\,|\,\xi \in N_{\Lambda(s)}(G(s,\,\widehat{x}(s)))\Big\}$$

for a. e. $s \in \Omega$. The result thus follows directly from Theorem 3.2. □

The mapping which assigns the vector $\widehat{y}^*_s$ to a. e. $s \in \Omega$ can be viewed as the *Karush–Kuhn–Tucker* (KKT) function of the program (2.1) with Γ given by (3.8). In this connection a natural question arises: Under what assumptions there exists a *measurable* KKT function? This question is answered in the next statement.

Theorem 3.5. In addition to the assumptions of Theorem 3.4 suppose that $p < \infty$ and that condition (3.6) holds true. Then there exist an element $\widehat{v}^* \in L^q(\Omega;\mathbb{R}^n)$ and a measurable KKT function $s \mapsto \widehat{y}^*(s)$ such that

$$\left.\begin{aligned} \widehat{y}^*(s) \in N_{\Lambda(s)}(G(s,\widehat{x}(s))) \\ \widehat{v}^*(s) \in \partial f(s,\widehat{x}(s)) \cap [-D^*G(s,\,\widehat{x}(s))\,(\widehat{y}^*(s)) - N_{\Theta(s)}(\widehat{x}(s))] \end{aligned}\right\} \tag{3.11}$$

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

Proof. The statement of Theorem 3.5 can be reformulated in the following form: There exist mappings $\widehat{v}^*[\Omega \to \mathbb{R}^n]$, $\widehat{w}^*[\Omega \to \mathbb{R}^n]$ and $\widehat{y}^*[\Omega \to \mathbb{R}^m]$ such that

$$\begin{aligned} \widehat{v}^*(s) &\in \partial f(s,\,\widehat{x}(s)) \\ -\widehat{v}^*(s) - \widehat{w}^*(s) &\in N_{\Theta(s)}(\widehat{x}(s)) \\ \widehat{y}^*(s) &\in N_{\Lambda(s)}(G(s,\,\widehat{x}(s))) \\ (\widehat{w}^*(s), -\widehat{y}^*(s)) &\in N_{\mathrm{gph}\,G(s,\cdot)}(\widehat{x}(s),\,G(s,\,\widehat{x}(s))) \end{aligned} \tag{3.12}$$

for a. e. $s \in \Omega$. In the above relations $\mathrm{gph}\,G(s,\cdot)$ denotes the set $\{(y,z) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, z = G(s,y)\}$. Using the argumentation of Theorem 3.3, we easily infer that $\widehat{v}^* \in L^q(\Omega;\mathbb{R}^n)$

provided it is measurable. Thus it suffices to prove the existence of measurable functions $\widehat{v}^*$, $\widehat{w}^*$ and $\widehat{y}^*$ satisfying (3.12). To this purpose we introduce the function $q[\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m]$ defined by

$$q(a,b,c) := \begin{bmatrix} -a \\ a+b \\ -c \\ -b \\ +c \end{bmatrix},$$

and the multifunction $Q[\Omega \rightsquigarrow \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m]$ defined by

$$Q(s) := \partial f(s,\widehat{x}(s)) \times N_{\Theta(s)}(\widehat{x}(s)) \times N_{\Lambda(s)}(G(s,\widehat{x}(s))) \times N_{\operatorname{gph} G(s,\cdot)}(\widehat{x}(s), G(s,\widehat{x}(s))). \quad (3.13)$$

Functions $\widehat{v}^*$, $\widehat{w}^*$ and $\widehat{y}^*$ are thus selections of the multifunction H, given by

$$H(s) := \{(a,b,c) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \mid 0 \in q(a,b,c) + Q(s)\}.$$

Clearly, q is a Carathéodory function. Further, since G is a Carathéodory function and $\widehat{x}$ is in $L^p(\Omega;\mathbb{R}^n)$, we observe that the functions $G(\cdot,\widehat{x}(\cdot))$ and $(\widehat{x}(\cdot), G(\cdot,\widehat{x}(\cdot)))$ are measurable. The first three multifunctions in the carthesian product (3.13) are measurable due to the results from [20] mentioned in the proof of Theorem 3.3. To see the measurability of the 4th multifunction, note first that

$$\operatorname{gph} G(s,\cdot) = \{(y,z) \in \mathbb{R}^n \times \mathbb{R}^m \mid z - G(s,y) \in \{0\}\}.$$

Hence, Theorem 8.2.9 from [1] applies and yields the measurability of $\operatorname{gph} G(s,\cdot)$. The measurability of $N_{\operatorname{gph} G(s,\cdot)}(\widehat{x}(s),\, G(s,\, \widehat{x}(s)))$ follows then from [20, Theorem 14.26]. The product multifunction Q is thus also measurable ([20, Proposition 14.11 (d)]) and we can apply [1, Theorem 8.2.9] once more, this time to the multifunction H. In this way we have proved the measurability of H and our statement follows from the Measurable Selection Theorem. □

Remark. If $G(s,\cdot)$ is Lipschitz near $\widehat{x}(s)$ for some $s \in \Omega$, then one has

$$D^*G(s,\widehat{x}(s))(\widehat{y}^*(s)) = \partial\langle \widehat{y}^*(s), G(s,\cdot)\rangle\,(\widehat{x}(s)) \subset \left\{C(s)^{\mathrm{T}}\widehat{y}^*(s) | C(s) \in \overline{\partial} G(s,\widehat{x}(s))\right\} \quad (3.14)$$

cf. [10]. The 2nd or the 3rd set in (3.14) can sometimes be computed more easily than the coderivative $D^*G(s,\, \widehat{x}(s))\,(\widehat{y}^*(s))$.

Let us comment on the optimality conditions (3.7), (3.11). The limiting subdifferential and the limiting normal cone $N_{\Gamma(s)}(\widehat{x}(s))$ is in nonconvex situations usually much smaller than the corresponding objects of Clarke for which an analogous condition has been proved in [7]. This is strikingly illustrated in Example 3.1, where the Clarke normal cone to $\Gamma(s)$ at $\widehat{x}(s)$ is the whole space $\mathbb{R}^2$ provided $k(s) \geq 0$. Hence, our conditions are sharper. In Theorem 3.5 we have specified conditions under which there exists a measurable KKT function. It is tempting to try to ensure also the existence of an integrable KKT function. Before we formulate the respective statement, let us recall that a multifunction $\Phi[\mathbb{R}^p \rightsquigarrow$

$I\!R^q$] is *pseudo-Lipschitz* around $(a, b) \in \operatorname{gph} \Phi$, provided there exist neighborhoods $\mathcal{U}$ of a and $\mathcal{V}$ of b and a modulus $\ell \geq 0$ such that

$$\Phi(a_1) \cap \mathcal{V} \subset \Phi(a_2) + \ell |a_1 - a_2| \, I\!B \quad \text{for all } a_1, \, a_2 \in \mathcal{U}.$$

Let $G(s, \cdot)$ be Lipschitz near $\widehat{x}(s)$ for a. e. $s \in \Omega$. From Theorem A.3 it follows that the constraint qualification

$$\left.\begin{array}{c} 0 \in D^* G(s, \, \widehat{x}(s)) \, (\xi) + N_{\Theta(s)}(\widehat{x}(s)) \\ \xi \in N_{\Lambda(s)}(G(s, \, \widehat{x}(s))) \end{array}\right\} \Rightarrow \xi = 0 \tag{3.15}$$

ensures the pseudo-Lipschitz continuity of the multifunction $\Sigma(s, \cdot)$, defined by

$$\Sigma(s, u) := \{ y \in \Theta(s) \mid u + G(s, y) \in \Lambda(s) \},$$

around $(0, \widehat{x}(s))$ for a. e. $s \in \Omega$. This multifunction is employed in the next statement.

Theorem 3.6. Let $\widehat{x}$ be a solution of (2.1), where Γ is given in the form (3.8) with $G(s, \cdot)$ Lipschitz near $\widehat{x}(s)$ for a. e. $s \in \Omega$. Let $p < \infty$ and condition (3.6) be fulfilled. Further assume that for a. e. $s \in \Omega$ the constraint qualification (3.15) holds true, and that the function $\rho(\cdot)$, assigning $s \in \Omega$ the modulus of pseudo-Lipschitz continuity of Σ around $(0, \widehat{x}(s))$, belongs to $L^p(\Omega)$. Then there exists a KKT function $\widehat{u}^* \in L^1(\Omega; I\!R^m)$ such that

$$\begin{array}{lcl} \widehat{u}^*(s) & \in & N_{\Lambda(s)}(G(s, \widehat{x}(s))) \\ 0 & \in & \partial f(s, \widehat{x}(s)) + D^* G(s, \widehat{x}(s)) \, (\widehat{u}^*(s)) + N_{\Theta(s)}(\widehat{x}(s)) \end{array} \tag{3.16}$$

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

Proof. Under the assumptions for a. e. $s \in \Omega$ the pair $(0, \widehat{x}(s))$ is a (local) solution of the optimization problem (in variables (u, v))

$$\begin{array}{ll} \text{minimize} & f(s, v) + R(s) \, |u| \\ \text{subject to} & \\ & (u, v) \in \operatorname{gph} \Sigma(s, \cdot) \end{array} \tag{3.17}$$

cf. [23, Lemma 3.1], provided the penalty parameter $R(s)$ is sufficiently large. Writing down the optimality conditions for (3.17), we obtain the existence of a pair $(\widehat{u}^*(s), -\widehat{v}^*(s))$ $\in N_{\operatorname{gph} \Sigma(s, \cdot)}(0, \widehat{x}(s))$ such that

$$0 \in \partial f(s, \widehat{x}(s)) - \widehat{v}^*(s)$$

for a. e. $s \in \Omega$. Thus, $\widehat{u}^*(s) \in D^* \Sigma(s, 0, \widehat{x}(s)) \, (-\widehat{v}^*(s))$ and, by invoking (A.3), we infer that

$$\widehat{u}^*(s) \in N_{\Lambda(s)}(G(s, \widehat{x}(s))),$$

and

$$-\widehat{v}^*(s) \in D^* G(s, \widehat{x}(s)) \, (\widehat{u}^*(s)) + N_{\Theta(s)}(\widehat{x}(s)).$$

It follows that $\widehat{u}^*$ is a KKT function and $\widehat{v}^* \in L^q(\Omega; I\!R^n)$. Moreover, by virtue of [13, Theorem 3.2] for a. e. $s \in \Omega$ one has

$$\sup\Big\{|a| \,\Big|\, a \in D^*\Sigma(s, 0, \widehat{x}(s))\,(b)\Big\} \leq \rho(s)\,|b|$$

due to the assumed pseudo-Lipschitz continuity of Σ around $(0, \widehat{x}(s))$. Therefore, by the Hőlder inequality

$$\int_\Omega |\widehat{u}^*(s)|\,\mu(\mathrm{d}s) \leq \int_\Omega \rho(s)\,|\widehat{v}^*(s)|\,\mu(\mathrm{d}s) \leq \|\rho\|_{L^p}\,\|\widehat{v}^*\|_{L^q}$$

and we are done. □

To ensure the required properties of the modulus ρ in terms of the original problem data is, however, not an easy task.

3.2 Reduced optimization

In [20, Example 14.62] the authors consider an optimization problem in two variables in which one of them can be eliminated by considering the respective value function as integrand. Such a situation typically arises in a class of two-stage nonconvex stochastic programs and so, using again the interchange of minimization and integration together with some results of the Mordukhovich calculus, we can strengther the optimality conditions of [6]. Correspondingly, in this subsection we will be dealing with the optimization problem

$$\begin{array}{ll} \text{minimize} & h(z) + I_g[z, x] \\ \text{subject to} & \\ & z \in D \subset I\!R^m \\ & x(s) \in \Gamma(s, z) \quad \text{for a. e. } s \in \Omega, \\ & x \in L^p(\Omega; I\!R^n), \end{array} \tag{3.18}$$

where $h[I\!R^m \to I\!R]$ is locally Lipschitz, D is nonempty and closed and $\Gamma(\cdot, z)$ is closed-valued and measurable for all $z \in D$. Further, for the sake of simplicity, we will assume that g maps $\Omega \times I\!R^m \times I\!R^n$ into $I\!R$, $g(s, z, y)$ is measurable in s for each pair (z, y) and locally Lipschitz in (z, y) for each s. This implies in particular that g is a Carathéodory integrand.

In this situation the statement from [20, Example 14.62] attains the following form.

Theorem 3.7. Consider problem (3.18) and suppose in addition to the posed assumptions that for a. e. $s \in \Omega$ the essential integrand

$$\widetilde{g}(s, z, y) = g(s, z, y) + \delta_{\Gamma(s,z)}(y)$$

is level bounded in y locally uniformly in z. Furthermore, with

$$f(s, z) := \inf_{y \in I\!R^n} \widetilde{g}(s, z, y) = \inf_{y \in \Gamma(s,z)} g(s, z, y)$$

let $h + I_f$ be proper.

Then the following two statements are equivalent:

(i) $(\widehat{z}, \widehat{x}) \in D \times L^p(\Omega; I\!R^n)$ is a (local) solution of (3.18);

(ii) $\widehat{z}$ is a (local) solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & h(z) + \int_\Omega f(s,z)\,\mu(\mathrm{d}s) \\ \text{subject to} & \\ & z \in D, \end{array} \tag{3.19}$$

and

$$\widehat{x}(s) \in \operatorname{argmin}_{y \in \Gamma(s,z)} g(s,z,y) \quad \text{for a. e. } s \in \Omega.$$

Since $z \in I\!R^m$, it is clear that in Theorem 3.7 it suffices to assume the level-boundedness of $\widetilde{g}$ in y uniformly only with respect to z from a neighborhood of $\widehat{z}$. On the basis of Theorem 3.7 we can now derive the optimality conditions for problem (3.18), i. e. the counterpart of [6, Theorem 5]. In the first step we invoke [11, Theorem 4.1] and observe that under the posed assumptions

$$\partial^\infty f(s,\widehat{z}) \subset \bigcup_{y_0 \in \operatorname{argmin}_{y \in I\!R^n} \widetilde{g}(s,\widehat{z},y)} D^*\Gamma(s,\widehat{z},y_0)\,(0) \quad \text{for all } s \in \Omega. \tag{3.20}$$

Therefore, whenever the set on the right-hand side of inclusion (3.20) contains only the zero vector, the value functions $f(s,\cdot)$ are Lipschitz near $\widehat{z}$ with a Lipschitz modulus $k(s)$, $s \in \Omega$.

Theorem 3.8. Let $\widehat{z} \in D$ be a (locally) optimal value of variable z in (3.18) and let all assumptions of Theorem 3.7 be fulfilled. Further assume that $D^*\Gamma(s,\widehat{z},y_0)\,(0) = \{0\}$ for all $y_0 \in \operatorname{argmin}_{y \in I\!R^n} \widetilde{g}(s,\widehat{z},y)$, $s \in \Omega$ and that the Lipschitz modulus k is integrable. Then there exists an integrable mapping $z^*[\Omega \to I\!R^m]$ such that

$$0 \in \partial h(\widehat{z}) + \int_\Omega z^*(s)\,\mu(\mathrm{d}s) + N_D(\widehat{z}) \tag{3.21}$$

and for a. e. $s \in \Omega$

$$z^*(s) \in D^*\Gamma(s,\widehat{z},x_s^0)\,(x^*) + \Delta(s)$$

with

$$\Delta(s) = \left\{ v^* \in I\!R^m \,\middle|\, (v^*,x^*) \in \partial g(s,\widehat{z},x_s^0) \right\}$$

and $x_s^0 \in \operatorname{argmin}_{y \in I\!R^n} \widetilde{g}(s,\widehat{z},y)$.

Proof. By Theorem 3.7 it suffices to derive optimality conditions for the finite-dimensional problem (3.19), where the only difficult part represents the special integral functional $J(z) := \int_\Omega f(s,z)\,\mu(\mathrm{d}s)$. For all z_1, z_2 sufficiently close to $\widehat{z}$ one has

$$|J(z_1) - J(z_2)| \le \int_\Omega |f(s,z_1) - f(s,z_2)|\,\mu(\mathrm{d}s) \le |z_1 - z_2| \int_\Omega k(s)\,\mu(\mathrm{d}s),$$

i. e., J is Lipschitz near $\widehat{z}$ by the assumed integrability of the Lipschitz modulus k. By virtue of [3, Theorem 2.7.2] to each $\xi \in \overline{\partial} J(z)$ there exists an integrable mapping $\widetilde{z}^*[\Omega \to I\!R^m]$ such that

$$\xi = \int_\Omega \widetilde{z}^*(s)\,\mu(\mathrm{d}s)$$

and

$$\widetilde{z}^*(s) \in \overline{\partial} f(s,\widehat{z}) \quad \text{for a. e. } s \in \Omega.$$

In our situation the map $\partial f(\cdot,\widehat{z})$ is integrably bounded and

$$\overline{\partial} f(s,\widehat{z}) = \operatorname{conv} \partial f(s,\widehat{z}) \quad \text{for all } s \in \Omega.$$

The Lyapunov–Aumann Theorem implies now the existence of an integrable selection z^* such that

$$z^*(s) \in \partial f(s,\widehat{z}) \quad \text{for a. e. } s \in \Omega$$

and relation (3.21) is fulfilled. It remains to recall from [11, Theorem 4.1] that under our assumptions

$$\partial f(s,\widehat{z}) \subset \bigcup_{x_s^0 \in \operatorname{argmin}_{y\in I\!R^n} \widetilde{g}(s,\widehat{z},y)} \left\{ z_1^* + z_2^* \,\middle|\, z_1^* \in D^*\Gamma(s,\widehat{z},x_s^0)\,(x^*),\ (z_2^*,x^*) \in \partial g(s,\widehat{z},x_s^0) \right\}$$

and we are done. □

Using [5, Theorem 8] it is possible to ensure by a suitable assumption that the map $s \mapsto x_s^0$ is measurable. In such a case Theorem 3.8 provides us with optimality conditions of the same type as Theorem 5 from [6]. The conditions of Theorem 3.8 are, however, substantially sharper because the Clarke subdifferential of g is replaced by the limiting subdifferential and the Clarke normal cones to D and $\operatorname{gph}\Gamma(s)$ are replaced by the limiting normal cones. Further one could assume as in [6] that $\Gamma(s,\cdot)$ is given by parameter-dependent inequalities and derive readily the counterpart of [6, Theorem 7]. Unfortunately, this approach leading to an improvement in case of two-stage stochastic programs, could not be applied in case of multistage stochastic programs as we will see in the next section.

4 Non-pointwise constraints

4.1 Fuzzy optimality conditions

This section deals with the optimization problem

$$\begin{array}{ll} \text{minimize} & I_f[x] \\ \text{subject to} & \\ & x(s) \in \Gamma(s) \quad \text{for a. e. } s \in \Omega, \\ & x \in C, \end{array} \tag{4.1}$$

where f and Γ fulfil the assumptions posed in connection with problem (2.1) and C is a nonempty and closed subset of $L^2(\Omega; I\!R^n)$. As pointed out in the introduction, C cannot be expressed in the pointwise form and therefore the approach of the preceding section is not applicable. Clearly, (4.1) amounts to the problem

$$\begin{array}{ll} \text{minimize} & I_{\widetilde{f}}[x] \\ \text{subject to} & \\ & x \in C, \end{array} \tag{4.2}$$

where $\widetilde{f}$ is the essential integrand introduced in (3.1). Since the sum $I_{\widetilde{f}} + \delta_C$ is lsc, it follows from the definition of the proximal subdifferential that

$$0 \in \partial_P(I_{\widetilde{f}} + \delta_C)\,(\widehat{x}), \tag{4.3}$$

whenever $\widehat{x}$ is a (local) minimum in (4.1) and $I_f[\widehat{x}] \in I\!R$. To express relation (4.3) in terms of the problem data, we invoke the the proximal variant of the weak fuzzy sum rule from [2]. We formulate this result, for the sake of simplicity, only for the sum of two functions, whereas the original statement concerns an arbitrary finite number of summands.

Theorem 4.1. Let X be a Hilbert space and f_1, f_2 $[X \to \overline{I\!R}]$ be lsc. Assume that $x^* \in \partial_P(f_1 + f_2)(x)$. Then, for each $\varepsilon > 0$ and each weak neighborhood V of 0 in X, there exist $x_1, x_2 \in I\!B(x;\varepsilon), x_1^* \in \partial_P f(x_1), x_2^* \in \partial_P f(x_2)$, such that

$$|f_n(x_n) - f_n(x)| < \varepsilon,\ n = 1,2$$

$$\|x_1 - x_2\| \max\{\|x_1^*\|, \|x_2^*\|\} < \varepsilon$$

and

$$x^* \in x_1^* + x_2^* + V.$$

In the proof one needs just to combine the ideas from the proof of [2, Theorem 2.7] with the strong proximal fuzzy sum rule in [8]. On the basis of Theorem 4.1 we obtain the following weak fuzzy optimality conditions for problem (4.1).

Theorem 4.2. Let $\widehat{x}$ be a (local) solution of problem (4.1). In addition to the posed assumptions suppose that either (A) or (A)' is satisfied where

(A) there exists a function $k \in L^2(\Omega; I\!R)$ such that, for all $s \in \Omega$,

$$|f(s,y_1) - f(s,y_2)| \le k(s)\,|y_1 - y_2| \quad \text{for all } y_1,\ y_2 \in I\!R^n;$$

(A)' for all $s \in \Omega$ the function $f(s,\cdot)$ is Lipschitz (of some rank) near each point of $I\!R^n$ and there exists a scalar $c > 0$ such that

$$\xi \in \overline{\partial} f(s,y) \Rightarrow |\xi| \le c(1 + |y|) \quad \text{for all } s \in \Omega,\ y \in I\!R^n.$$

Then to each $\varepsilon > 0$ and to each weak neighborhood V of 0 in $L^2(\Omega; I\!R^n)$ there exist functions x_1, x_2, x_3, x_1^*, x_2^*, $x_3^* \in L^2(\Omega; I\!R^n)$ such that

$$\begin{aligned}
&|I_f[x_1] - I_f[\widehat{x}]| < \varepsilon,\ x_2(s) \in \Gamma(s) \text{ for a.e. } s \in \Omega,\ x_3 \in C,\\
&\|x_1 - \widehat{x}\| \le \varepsilon,\ \|x_2 - \widehat{x}\| \le \varepsilon,\ \|x_3 - \widehat{x}\| \le \varepsilon,\\
&\left.\begin{aligned} x_1^*(s) &\in \partial_P f(s, x_1(s))\\ x_2^*(s) &\in N^P_{\Gamma(s)}(x_2(s)) \end{aligned}\right\} \text{ for a.e. } s \in \Omega\\
&x_3^* \in N^P_C(x_3)\\
&0 \in x_1^* + x_2^* + x_3^* + V.
\end{aligned} \tag{4.4}$$

Before we prove this statement, note that under (A) I_f is defined and (globally) Lipschitz on $L^2(\Omega; I\!R^n)$, whereas under (A)' I_f is Lipschitz only on bounded subsets of $L^2(\Omega; I\!R^n)$, cf. [3, Theorem 2.7.5].

Proof of Theorem 4.2. Since both functions $I_{\widetilde{f}}$ and δ_C are lsc, we are entitled to apply Theorem 4.1 to (4.3). This yields the existence of functions $\widetilde{x}$, $x_3 \in I\!B(\widehat{x}; \frac{\varepsilon}{2})$, $\widetilde{x}^* \in \partial_P I_{\widetilde{f}}[\widetilde{x}], x_3^* \in N_C^P(x_3)$ such that

$$\begin{array}{l} |I_f[\widetilde{x}] - I_f[\widehat{x}]| < \frac{\varepsilon}{2} \\ \widetilde{x}(s) \in \Gamma(s) \text{ for a.e. } s \in \Omega \\ x_3 \in C \\ 0 \in \widetilde{x}^* + x_3^* + U, \end{array} \tag{4.5}$$

where U is a weak neighborhood of 0 in $L^2(\Omega; I\!R^n)$ satisfying the inclusion

$$U + I\!B(0; \frac{\varepsilon}{2}) \subset V.$$

The first two relations in (4.5) follow from the inequality

$$|I_{\widetilde{f}}[\widetilde{x}] - I_{\widetilde{f}}[\widehat{x}]| < \frac{\varepsilon}{2}.$$

Now we apply the strong proximal sum rule [8, Theorem 2] to the relation $\widetilde{x}^* \in \partial_P(I_f + I_{\delta_\Gamma})[\widetilde{x}]$. This is possible, because I_f is Lipschitz on a neighborhood of $\widehat{x}$ so that the uniform lower semicontinuity property is satisfied. It follows the existence of functions $x_1, x_2 \in I\!B(\widetilde{x}; \frac{\varepsilon}{2}), x_1^* \in \partial_P I_f[x_1], x_2^* \in \partial_P I_f[x_2]$ such that

$$\begin{array}{l} |I_f[x_1] - I_f[\widetilde{x}]| < \frac{\varepsilon}{2} \\ x_2(s) \in \Gamma(s) \text{ for a.e. } s \in \Omega \\ \widetilde{x}^* \in x_1^* + x_2^* + I\!B(0; \frac{\varepsilon}{2}). \end{array} \tag{4.6}$$

It remains to put relations (4.5), (4.6) together and take into account the implication (2.3). □

Since V is a weak neighborhood, it is not possible to get a limiting version of (4.4) by letting $\varepsilon \to 0$. On the other hand, in Theorem 4.2 we do not have to consider any constraint qualification arising usually in optimality conditions of KKT type, and the subdifferentials and normal cone in (4.4) are in a certain sense the smallest possible.

Let Ξ be a subset of an Asplund space Z. The Fréchet normal cone is defined in Z in exactly the same way as in $I\!R^p$ (Definition 2.1). The limiting normal cone to Ξ at $\bar{z}$ is then the set

$$\begin{aligned} N_\Xi(\bar{z}) \quad := \quad & \limsup_{z \to \bar{z}} \widehat{N}_\Xi(z) = \\ & \Big\{ z^* \in Z^* | \exists \text{ sequences } z_k \to \bar{z} \text{ and } z_k^* \xrightarrow{w^*} z^*, \\ & \qquad \text{with } z_k^* \in \widehat{N}_\Xi(z_k) \text{ for all } k = 1, 2 \ldots \Big\}. \end{aligned}$$

We say that Ξ is *sequentially normally compact* at $\overline{z} \in \Xi$, provided any sequence $\{(z_k, z_k^*)\}$ satisfying

$$z_k^* \in \widehat{N}_\Xi(z_k), \quad z_k \to \overline{z} \quad \text{and} \quad z_k^* \xrightarrow{w^*} 0$$

contains a subsequence $\{(z_{k'}, z_{k'}^*)\}$ with $\|z_{k'}^*\| \to 0$. To derive optimality conditions for problem (4.1) in a standard KKT form, we need to ensure the inclusion

$$N_{C \cap D}(\widehat{x}) \subset N_C(\widehat{x}) + N_D(\widehat{x}), \tag{4.7}$$

where $D := \{x \in L^2(\Omega; I\!R^n) | x(s) \in \Gamma(s) \text{ for a.e. } s \in \Omega\}$. Following [15], this can be done by requiring that either C or D is sequentially normally compact at $\widehat{x}$ and

$$N_C(\widehat{x}) \cap -N_D(\widehat{x}) = \{0\}.$$

Unfortunately, we are not able to prove the sequential normal compactness of any from the sets C, D in the applications we have in mind, in particular in multistage stochastic programs. Additionally, a pointwise description of $N_D(\widehat{x})$ in terms of Γ is, as to our knowledge, generally not available. To summarize, for problem (4.1) we dispose with the fuzzy optimality conditions stated in Theorem 4.2 which are valid under very weak conditions imposed on the problem data. If these data, however, fulfil some more restrictive assumptions, optimality conditions in the classical KKT form can be derived.

Theorem 4.3. Let $\widehat{x}$ be a (local) solution of (4.1), where C is sequentially normally compact at $\widehat{x}$ and D is regular at $\widehat{x}$ (i.e. $N_D(\widehat{x})$ equals to the negative polar of the Bouligand (contingent) cone to D at $\widehat{x}$). Further suppose that either assumption (A) or assumption (A)' from Theorem 4.2 are fulfilled and that the constraint qualification

$$N_C(\widehat{x}) \cap \{x^* \in L^2(\Omega; I\!R^n) | -x^*(s) \in N_{\Gamma(s)}(\widehat{x}(s)) \text{ a.e. on } \Omega\} = \{0\} \tag{4.8}$$

holds true. Then there exist functions $x_1^*, x_2^*, x_3^* \in L^2(\Omega; I\!R^n)$ such that

$$\left.\begin{array}{l} x_1^*(s) \in \bar{\partial} f(s, \widehat{x}(s)) \\ x_2^*(s) \in N_{\Gamma(s)}(\widehat{x}(s)) \end{array}\right\} \text{ for a.e. } s \in \Omega$$
$$x_3^*(s) \in N_C(\widehat{x})$$

and

$$0 = x_1^* + x_2^* + x_3^*.$$

Proof. Due to the regularity of D at $\widehat{x}$ one can invoke [1, Cor.8.5.2], according to which

$$N_D(\widehat{x}) = \{x^* \in L^2(\Omega; I\!R^n) | x^*(s) \in N_{\Gamma(s)}(\widehat{x}(s)) \text{ a.e. on } \Omega\}.$$

Since C is sequentially normally compact at $\widehat{x}$, the constraint qualification (4.8) ensures by virtue of [15, Prop 2.2] the inclusion (4.7). Under (A) or (A)' the objective is Lipschitz around $\widehat{x}$. Therefore,

$$0 \in \partial I_f[\widehat{x}] + N_C(\widehat{x}) + N_D(\widehat{x}).$$

It remains to observe that

$$\partial I_f[\widehat{x}] \subset \bar{\partial} I_f[\widehat{x}] \subset \{\xi \in L^2(\Omega; I\!R^n) | \xi(s) \in \bar{\partial} F(s, \widehat{x}(s)) \text{ a.e. in } \Omega\},$$

and we are done. $\square$

Remark. D is regular at $\widehat{x}$, provided Γ is convex-valued for a.e. $s \in \Omega$. Concerning the sequential normal compactness of C, the respective sufficient conditions can be found in [16] for sets with various frequently appearing structures.

A favourable situation for the construction of classical KKT optimality conditions for (4.1) arises if C is a decomposable subspace of $L^2(\Omega; I\!R^n)$ (or even $L^p(\Omega; I\!R^n), 1 \leq p \leq \infty$). In such a case the approach via extended Lipschitz integrands from [7] can be used and it is necessary to assume neither the sequential normal compactness of C nor any constraint qualification of the type (4.8). In the rest of this section we will examine the form of conditions (4.4) in case of a multistage stochastic program.

4.2 Nonsmooth multistage stochastic programs

We consider a finite horizon sequential decision process under uncertainty, in which a decision made at stage k is defined on a probability space $(\Omega, \mathcal{S}, \mu)$ and based only on information that is available at k and becomes more refined with growing k $(1 \leq k \leq K)$. More precisely, we assume that the information at k is given by a σ-algebra $\mathcal{S}_k$ and that the stochastic decision x_k at stage k varying in $I\!R^{n_k}$ is measurable with respect to $\mathcal{S}_k$. The latter property is called *nonanticipativity.* Furthermore, we assume that

$$\mathcal{S}_1 = \{\emptyset, \Omega\} \subseteq \ldots \subseteq \mathcal{S}_k \subseteq \mathcal{S}_{k+1} \subseteq \mathcal{S} \quad (k = 1, \ldots, K-1)$$

i.e., x_1 is deterministic, and, with no loss of generality, we may assume that $\mathcal{S}_K = \mathcal{S}$. We take up the classical approach of [17, 19] and formulate the sequential decision model as a mathematical program in a space of integrable functions, here, the space $L^2(\Omega; I\!R^n)$ with $n := \sum_{k=1}^{K} n_k$. The objective is given by an integral functional $I_f[x]$, where f is a normal integrand from $\Omega \times I\!R^n$ to $\overline{I\!R}$ and the constraints consist of two groups: pointwise constraints

$$\varphi(s, x(s)) \leq 0 \quad \text{for a.e. } s \in \Omega,$$

and functional (non-pointwise) constraints

$$x_k \in L^2(\Omega; I\!R^{n_k}) \qquad \text{and} \qquad x_k = I\!E[x_k|\mathcal{S}_k] \quad (k = 1, \ldots, K),$$

describing integrability and nonanticipativity of the decision x. Here, $\varphi = (\varphi^1, \ldots, \varphi^m)$ is a mapping from $\Omega \times I\!R^n$ to some Euclidean space and $I\!E[\cdot|\mathcal{S}_k]$ denotes the conditional expectation with respect to the σ-algebra $\mathcal{S}_k$ $(k = 1, \ldots, K)$. This leads to the following K-stage stochastic programming model:

$$\begin{array}{ll} \text{minimize} & I_f[x] = \int\limits_{\Omega} f(s, x(s))\mu(ds) \\ \text{subject to} & \varphi(s, x(s)) \leq 0 \quad \text{for a.e. } s \in \Omega, \\ & x \in C := \{x \in L^2(\Omega; I\!R^n) \,|\, x_k = I\!E[x_k|\mathcal{S}_k] \; k = 1, \ldots, K\}. \end{array} \tag{4.9}$$

In general, the set C forms a closed linear subspace of $L^2(\Omega; I\!R^n)$ and has the specific structure $I\!R^{n_1} \times L^2(\Omega; I\!R^{n_2})$ in the two-stage situation (i.e., $K = 2$). While the latter structure allows to reduce the model (4.9) to the model (3.18), the situation for $K > 2$ becomes quite different since C is not decomposable in general.

Theorem 4.4. Let $\widehat{x} \in L^2(\Omega; I\!R^n)$ be a (local) solution of problem (4.9). Suppose that $\varphi(s, \cdot)$ is locally Lipschitz on $I\!R^n$ for each $s \in \Omega$ and that either assumption (A) or (A)' of Theorem 4.2 is fulfilled. Further assume that for all $x \in L^2(\Omega; I\!R^n)$ such that $x(s) \in \Gamma(s)$ for a. e. $s \in \Omega$ one has

$$0 \notin \{ \sum_{i \in I(s,x(s))} \lambda_i \xi_i \,|\, \xi_i \in \partial\varphi^i(s, x(s)) \,,\, \lambda_i \geq 0 \,,\, \sum_{i \in I(s,x(s))} \lambda_i = 1\} \quad \text{for a. e. } s \in \Omega \,, \tag{4.10}$$

where $I(s, x(s)) := \{i \in \{1, \dots, m\} \,:\, \varphi^i(s, x(s)) = 0\}$.
Then to each $\varepsilon > 0$ and to each weak neighborhood V of 0 in $L^2(\Omega; I\!R^n)$ there exist functions x_1, x_2, x_3, x^*, $x_3^* \in L^2(\Omega; I\!R^n)$ such that

(i) $|I_f[x_1] - I_f[\widehat{x}]| < \varepsilon \,,\, \varphi(s, x_2(s)) \leq 0$ for a. e. $s \in \Omega$, and $x_3 = (x_{31}, x_{32}, \dots, x_{3K})$ where x_{3k} is $\mathcal{S}_k$-measurable for $k = 1, \dots, K$;

(ii) $\|x_1 - \widehat{x}\| \leq \varepsilon$, $\|x_2 - \widehat{x}\| \leq \varepsilon$, $\|x_3 - \widehat{x}\| \leq \varepsilon$;

(iii) for a.e. $s \in \Omega$ there exist multipliers $\lambda_s \in I\!R^m_+$ such that

$$x^*(s) \in \partial_P f(s, x_1(s)) + \sum_{i=1}^m \lambda_s^i \partial\varphi^i(s, x_2(s)) \quad \text{and} \quad \lambda_s^i \varphi^i(s, x_2(s)) = 0 \; (i = 1, \dots, m) \,;$$

(iv) for $x_3^* = (x_{31}^*, x_{32}^*, \dots, x_{3K}^*)$ it holds $I\!E[x_{3k}^* | \mathcal{S}_k] = 0$ a.e. and $k = 1, \dots, K$;

(v) $0 \in x^* + x_3^* + V$.

Proof. It suffices to express the cones $N^P_{\Gamma(s)}(x_2(s))$ for a.e. $s \in \Omega$ and $N^P_C(x_3)$ from Theorem 4.2 in terms of the data of problem (4.9). To this purpose, we invoke [10, Cor. 4.4.2] according to which the inclusion

$$N^P_{\Gamma(s)}(x_2(s)) \subseteq N_{\Gamma(s)}(x_2(s)) \subseteq \bigcup \left\{ \sum_{i=1}^m \lambda^i \partial\varphi^i(s, x_2(s)) \,|\, \lambda^i \geq 0 \,,\, \lambda^i \varphi^i(s, x_2(s)) = 0 \right\}$$

holds, whenever the constraint qualification (4.10) is fulfilled.
The cone $N^P_C(x_3)$ coincides with the normal cone of convex analysis to the closed linear subspace C of $L^2(\Omega; I\!R^n)$ at x_3, i.e., with the orthogonal subspace $C^\perp$ to C. Due to the orthogonal projection property of the conditional expectation, it holds $C^\perp = \{x^* \in L^2(\Omega; I\!R^n) \,|\, I\!E[x_k^* | \mathcal{S}_k] = 0 \text{ a. e.}\}$ completing the proof. □

The usage of limiting subdifferentials in the upper approximation of $N^P_{\Gamma(s)}(x_2(s))$ makes possible to evaluate all needed subdifferentials $\partial\varphi^i$, $i \in \{1, 2, \dots, m\}$, at the points $x_2(s)$, $s \in \Omega$. One could, however, apply a relevant rule from the fuzzy calculus also to this purpose. This leads to the following statement.

Theorem 4.5. Let the assumptions of Theorem 4.4 be satisfied except condition (4.10) which has to be replaced by:

$$\liminf_{y\to x(s)} \ \mathrm{dist}(0|\partial_P\varphi^i(s,y)) > 0 \text{ for a. e. } s\in\Omega\,,\ i=1,2,\dots,m. \tag{4.11}$$

Then to each $\varepsilon > 0$ and to each weak neighborhood V of 0 in $L^2(\Omega;I\!R^n)$ there exist functions x_1, x_2, x_3, x_1^*, x_2^*, $x_3^* \in L^2(\Omega;I\!R^n)$ such that the assertions (i), (ii) and (iv) of Theorem 4.5 hold true and assertions (iii) and (v) are replaced by (iii)' and (v)' below.

(iii)' For a.e. $s\in\Omega$ one has

$$x_1^*(s)\in\partial_P f(s,x_1(s))\,,$$

and there exist vectors $y_{is}\in I\!B(x_2(s);\varepsilon)$, $i=1,2,\dots,m$, a multiplier $\lambda_s\in$ int $I\!R^m_+$ and proximal subgradients $\xi_{is}\in\partial_P\varphi^i(s,y_{is})$, $i=1,2,\dots,m$, such that

$$|\varphi^i(s,y_{is})-\varphi^i(s,x_2(s))|<\varepsilon \quad\text{and}\quad \left|x_2^*(s)-\sum_{i=1}^m\lambda_s^i\xi_{is}\right|\le\varepsilon;$$

(v)' $0\in x_1^*+x_2^*+x_3^*+V$.

In the proof it suffices to express the proximal normal cones to the corresponding sets $\{y\in I\!R^n\,|\,\varphi^i(s,y)\le 0\}$ at $y=x_2(s)$, $s\in\Omega$, on the basis of [2, Thm. 3.6], which is possible by virtue of the qualification condition (4.11). Then one applies the proximal variant of the weak fuzzy sum rule [2, Thm. 2.7] and arrives at the above result.

In the description of $N^P_{\Gamma(s)}(x_2(s))$ one works now with smaller subdifferentials (proximal instead of limiting), but they are not evaluated at $x_2(s)$, $s\in\Omega$. On the other hand, the qualification condition (4.11) concerns the single functions φ^i , $i=1,2,\dots,m$, separately and not jointly as condition (4.10).

Appendix

Consider first an abstract mathematical program of the form

$$\begin{array}{ll}\text{minimize} & \varphi(x)\\ \text{subject to} & x\in\Pi,\end{array} \tag{A.1}$$

where φ maps $I\!R^p$ into $\overline{I\!R}$ and Π is a closed subset of $I\!R^p$. In [10] the 1st-order necessary optimality conditions for problem (A.1) have been proved in the following form.

Theorem A.1. Let $\widehat{x}$ be a local solution of (A.1) and φ be lower semicontinuous in a neighborhood of $\widehat{x}$. Then there exist an element $\widehat{x}^*\in I\!R^p$ and a real $\lambda\ge 0$, not both equal to zero, such that

$$(\widehat{x}^*,-\lambda)\in N_{\mathrm{epi}\,\varphi}(\widehat{x},\,\varphi(\widehat{x}))\quad\text{and}\quad -\widehat{x}^*\in N_\Pi(\widehat{x}).$$

Under the additional condition

$$\partial^\infty \varphi(\widehat{x}) \cap (-N_\Pi(\widehat{x})) = \{0\},$$

one has $\lambda \neq 0$ and

$$0 \in \partial\varphi(\widehat{x}) + N_\Pi(\widehat{x}).$$

The generalized differential calculus of B. Mordukhovich is rather rich and enables to compute generalized normal cones or their upper approximations to a large number of sets with different structure. The next statement can easily be proved on the basis of [10, Theorem 1.2] and [12, Corollary 5.5 and Theorem 6.10].

Theorem A.2. Consider the set

$$A := \{x \in C \mid F(x) \in D\},$$

and the associated multifunction Q defined by

$$Q(y) := \{x \in C \mid y + F(x) \in D\},$$

where $F[\mathbb{R}^n \to \mathbb{R}^m]$ is continuous and C, D are closed subsets of $\mathbb{R}^n$, $\mathbb{R}^m$, respectively. Let $\overline{x} \in A$ and

$$D^*F(\overline{x}) \circ N_D(F(\overline{x})) \cap (-N_C(\overline{x})) = \{0\}. \tag{A.2}$$

Then for all $x^* \in \mathbb{R}^n$ one has

$$D^*Q(0,\overline{x})\,(x^*) \subset \left\{y^* \in N_D(F(\overline{x})) \middle| 0 \in x^* + D^*F(\overline{x})\,(y^*) + N_C(\overline{x})\right\}. \tag{A.3}$$

Furthermore, under the condition

$$N_D(F(\overline{x})) \cap Ker\ D^*\,F(\overline{x}) = \{0\}, \tag{A.4}$$

the inclusion

$$N_A(\overline{x}) \subset \{x^* \in \mathbb{R}^n \mid x^* \in D^*\,F(\overline{x})\,(\xi) + N_C(\overline{x}),\ \xi \in N_D(F(\overline{x}))\}.$$

holds true.

By combining (A.3) and [13, Theorem 3.2] we obtain the following criterion of pseudo-Lipschitz continuity of Q around $(0,\overline{x})$.

Theorem A.3. Consider the map Q from Theorem A.2 with F Lipschitz near $\overline{x}$ and assume that the constraint qualification

$$\left.\begin{array}{rcl} 0 & \in & D^*\,F(\overline{x})\,(\xi) + N_C(\overline{x}) \\ \xi & \in & N_D(F(\overline{x})) \end{array}\right\} \text{ implies } \xi = 0 \tag{A.5}$$

is fulfilled.

Then conditions (A.2), (A.4) hold true and Q is pseudo-Lipschitz around $(0,\overline{x})$.

The above statement shows a well-known connection between the pseudo-Lipschitz continuity of Q and the possibility to express (an upper approximation of) $N_A(\overline{x})$ in terms of the constraint data.

Acknowledgement. The authors would like to express their gratitude to Boris Mordukhovich (Detroit) for his extensive support during this research and to Alexander Kruger (Minsk) and Marian Fabian (Prague) for valuable discussions.

References

[1] J.–P. Aubin, H. Frankowska: *Set–Valued Analysis*. Birkhäuser, Boston 1990.

[2] J. M. Borwein, Q. J. Zhu: A survey of subdifferential calculus with applications. *Nonl. Analysis, Theory, Methods & Applications* 38 (1999), 687–773.

[3] F. H. Clarke: *Optimization and Nonsmooth Analysis*. Wiley, New York 1983.

[4] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, P. R. Wolenski: *Nonsmooth Analysis and Control Theory*. Springer, New York 1998.

[5] J.–B. Hiriart-Urruty: Gradients generalises de fonctions marginales. *SIAM J. Contr. Optim.* 16 (1978), 301–316.

[6] J.–B. Hiriart-Urruty: Conditions necessaires d'optimalite pour un programme stochastique avec recours. *SIAM J. Contr. Optim.* 16 (1978), 317–329.

[7] J.–B. Hiriart-Urruty: Extensions of Lipschitz integrands and minimization of nonconvex integral functionals: Application to the optimal recourse problem in discrete time. *Probability and Math. Statistics* 3 (1982), 19–36.

[8] A. D. Ioffe, R. T. Rockafellar: The Euler and Weierstrass conditions for nonsmooth variational problems. *Calculus of Variations* 4 (1996), 59–87.

[9] B. S. Mordukhovich: Maximum principle in the problem of time optimal response with nonsmooth constraints. *J. Appl. Math. Mech.* 40 (1976), 960–969.

[10] B. S. Mordukhovich: *Approximation Methods in Problems of Optimization and Control* (in Russian). Nauka, Moscow 1988 (English edition to appear in Wiley–Interscience).

[11] B. S. Mordukhovich: Sensitivity analysis in nonsmooth optimization. In: *Theoretical Aspects in Industrial Design* (D. A. Field, V. Komkov, eds.), SIAM Publications, Proc. in Applied Mathematics, Vol. 58, 1992, pp. 32–46.

[12] B. S. Mordukhovich: Generalized differential calculus for nonsmooth and set–valued mappings. *J. Math. Anal. Appl.* 183 (1994), 250–288.

[13] B. S. Mordukhovich: Lipschitzian stability of constraint systems and generalized equations. *Nonl. Analysis, Theory, Methods & Applications* 22 (1994), 173–206.

[14] B. S. Mordukhovich, Y. Shao: Nonconvex differential calculus for infinite-dimensional multifunctions. *Set-Valued Analysis* 4 (1996), 205–236.

[15] B. S. Mordukhovich, B. Wang: Necessary suboptimality and optimality conditions via variational principles. Preprint # 4, Wayne State University, Detroit, 2000.

[16] B. S. Mordukhovich, B. Wang: Sequential normal compactness in variational analysis. *Nonl. Analysis, Theory, Methods & Applications* 47 (2001), 717–728.

[17] P. Olsen: Multistage stochastic programming with recourse as mathematical programming in an L_p-space, *SIAM J. Contr. Optim.* 14 (1976), 528–537.

[18] R. T. Rockafellar: Proximal subgradients, marginal values and augmented Lagrangians in nonconvex optimization, *Math. Oper. Res.* 6 (1981), 424–436.

[19] R. T. Rockafellar, R. J-B Wets: The optimal recourse problem in discrete time: L^1-multipliers for inequality costraints. *SIAM J. Contr. Optim.* 16 (1978), 16–36.

[20] R. T. Rockafellar, R. J-B Wets: *Variational Analysis*. Springer Verlag, Berlin 1998.

[21] S. Vogel: Necessary optimality conditions for two-stage stochastic programming problems. *Optimization* 16 (1985), 607–616.

[22] R. J-B Wets: Stochastic programs with fixed recourse: The equivalent deterministic program. *SIAM Review* 16 (1974), 309–339.

[23] J. J. Ye, X. Y. Ye: Necessary optimality conditions for optimization problems with variational inequality constraints. *Math. Oper. Res.* 22 (1997), 977–997.