

On Optimality Conditions for Some Nonsmooth Optimization Problems over L^p Spaces¹

J. V. OUTFATA² AND W. RÖMISCH³

Communicated by P. Tseng

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.

Key Words. Normal integrands, integral functionals, normal cones, subdifferentials, fuzzy calculus, coderivatives, stochastic programming, two-stage programs, multistage programs.

1. Introduction

In Ref. 1, 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 Ref. 1 can well be applied, under some assumptions, to the minimization of certain integral functionals subject to the constraints

$$x(s) \in \Gamma(s), \quad \text{for a.e. } s \in \Omega, \quad (1a)$$

$$x \in L^p(\Omega; \mathbb{R}^n), \quad (1b)$$

¹The authors express their gratitude to Boris Mordukhovich (Detroit) for his extensive support during this research and to Marian Fabian (Prague) and Alexander Kruger (Ballarat) for valuable discussions. They are indebted also to two anonymous referees for helpful suggestions.

²Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic. The research of this author was partly supported by Grant 1075005 of the Czech Academy of Sciences.

³Humboldt-University Berlin, Institute of Mathematics, Berlin, Germany. The research of this author was supported by the Deutsche Forschungsgemeinschaft.

where $\Gamma[\Omega \rightrightarrows \mathbb{R}^n]$ 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, parametrized by $s \in \Omega$.

Concerning the 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 (Refs. 2–6) 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 by comparison with Refs. 7–8, where the conditions have been derived on the basis of the generalized differential calculus of Clarke (Ref. 9).

Unfortunately, the situation changes substantially whenever our optimization problem contains, besides the constraints (1), also a nonpointwise 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 Ref. 10 in terms of the Clarke generalized differential calculus under some additional assumptions imposed on the nonpointwise constraint. Trying to weaken these additional assumptions as much as possible, we have confined ourselves to the Hilbert space $L^2(\Omega; \mathbb{R}^n)$ 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 nonpointwise constraints. The satisfaction of such constraint qualifications happens to be the most serious hurdle on the way to the classical KKT conditions, at least in 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 (Section 5) 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: $\text{cl } A$ is the closure of a set A and $\mathbb{B}(a; \rho)$ is the closed ball with the center at a and radius ρ . If $a=0$ and $\rho=1$, we write simply \mathbb{B} . $\bar{\mathbb{R}}$ is the extended real line. For a function $f[X \rightarrow \bar{\mathbb{R}}]$, $\text{epi } f$ denotes its epigraph and $\bar{\partial}f(x)$ denotes the Clarke

subdifferential of f at x . If F maps \mathbb{R}^n into \mathbb{R}^m , then $\bar{\partial}F(x)$ is the Clarke generalized Jacobian of F at x . Analogously, $\bar{N}_A(x)$ denotes the Clarke normal cone to A at x . For a multifunction $\Phi[X \rightrightarrows Y]$,

$$\text{gph } \Phi = \{(x, y) | y \in \Phi(x)\}.$$

$\delta_C(\cdot)$ denotes the indicator functional of a set C and $\text{dist}(x|C)$ is the distance from x to C . $|\cdot|$ is a norm in \mathbb{R}^n , whereas $\|\cdot\|$ is the norm in a considered function space. $o(t)$ indicates a term with the property that

$$o(t)/t \rightarrow 0, \quad \text{as } t \rightarrow 0, \quad t \neq 0.$$

2. Problem Formulation and Preliminaries

Consider a function f mapping $\Omega \times \mathbb{R}^n$ into $\bar{\mathbb{R}}$. Following Ref. 1, f is called a normal integrand provided its epigraphical mapping,

$$\mathcal{E}_f(s) := \text{epi} f(s, \cdot) = \{(y, \alpha) \in \mathbb{R}^n \times \mathbb{R} | f(s, y) \leq \alpha\},$$

is closed-valued and measurable. We recall that a multifunction $\Theta[\Omega \rightrightarrows \mathbb{R}^n]$ is measurable if, for every closed set $\mathcal{O} \subset \mathbb{R}^n$, the set $\Theta^{-1}(\mathcal{O})$ is measurable, i.e., $\Theta^{-1}(\mathcal{O}) \in \mathcal{S}$ (see Ref. 11). In particular, the set $\text{dom } \Theta = \Theta^{-1}(\mathbb{R}^n)$ must be measurable. For a comprehensive treatment of measurable multifunctions, we refer the reader e.g. to Ref. 1 or Ref. 12. In what follows, we adopt the notation of Ref. 1, where $\int_{\Omega} f(s, x(s))\mu(ds)$ is denoted by $I_f[x]$. The next section concerns essentially optimization problems of the form

$$\min I_f[x], \tag{2a}$$

$$\text{s.t. } x(s) \in \Gamma(s), \quad \text{for a.e. } s \in \Omega, \tag{2b}$$

$$x \in L^p(\Omega; \mathbb{R}^n), \tag{2c}$$

where f is a normal integrand, Γ is a closed-valued and measurable multifunction, and where $1 \leq p \leq \infty$. In Section 4, we add to the constraints of (2) another one, namely

$$x \in C,$$

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

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

Consider an arbitrary set $\Pi \subset \mathbb{R}^p$.

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

$$\hat{N}_\Pi(a) := \left\{ x^* \in \mathbb{R}^p \mid \limsup_{x \xrightarrow{\Pi} a} \langle x^*, x - a \rangle / |x - a| \leq 0 \right\}$$

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

$$N_\Pi(a) = \limsup_{a' \xrightarrow{\text{cl } \Pi} a} \hat{N}_\Pi(a'). \tag{3}$$

The \limsup in (3) 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 $\text{cl } \Pi$ (with respect to $\text{cl } \Pi$). This is essential in the calculus of Mordukhovich subdifferentials and coderivatives introduced below.

Definition 2.2. Let $\varphi[\mathbb{R}^p \rightarrow \bar{\mathbb{R}}]$ be an arbitrary extended real-valued function and let $a \in \text{dom } \varphi$. The sets

$$\begin{aligned} \partial\varphi(a) &:= \{a^* \in \mathbb{R}^p \mid (a^*, -1) \in N_{\text{epi}\varphi}(a, \varphi(a))\}, \\ \partial^\infty\varphi(a) &:= \{a^* \in \mathbb{R}^p \mid (a^*, 0) \in N_{\text{epi}\varphi}(a, \varphi(a))\} \end{aligned}$$

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

It was shown in Ref. 2 that

$$\bar{\partial}\varphi(a) = \text{cl conv } (\partial\varphi(a) + \partial^\infty\varphi(a));$$

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

Definition 2.3. Let $\Phi[\mathbb{R}^p \rightrightarrows \mathbb{R}^q]$ be a multifunction and $b \in \Phi(a)$. The multifunction $D^*\Phi(a, b)[\mathbb{R}^q \rightrightarrows \mathbb{R}^p]$, defined by

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

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. See Refs. 13–14. Let $a \in \text{cl } \Pi$. 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 $N_\Pi^P(a)$.

It is easy to see that, for $H = \mathbb{R}^n$, one has $N_\Pi^P(a) \subset \hat{N}_\Pi(a)$. However, as explained e.g. in Ref. 1, the limiting normal cone has been originally defined via the proximal normal cone and (3) holds true with $\hat{N}_\Pi(a')$ replaced by $N_\Pi^P(a')$.

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

Definition 2.5. See Ref. 13. Let $\varphi[H \rightarrow \bar{\mathbb{R}}]$ be lower semicontinuous (lsc) at $a \in \text{dom } \varphi$. The set

$$\partial_P \varphi(a) := \left\{ a^* \in H \mid (a^*, -1) \in N_{\text{epi } \varphi}^P(a, \varphi(a)) \right\}$$

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 consist above all in handling integral functionals. For example, in $L^2(\Omega; \mathbb{R}^n)$, one has the implication

$$v \in \partial_P I_f[x] \Rightarrow v(s) \in \partial_P f(s, x(s)), \quad \text{for a.e. } s \in \Omega, \tag{4}$$

where the subdifferential of f is computed in the second argument only (cf. Refs. 14–15).

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 Ref. 1, Theorem 14.60. So, we start with the optimization problem (2)

under the assumptions posed in Section 2 and introduce the essential integrand $\tilde{f}[\Omega \times \mathbb{R}^n \rightarrow \bar{\mathbb{R}}]$ by

$$\tilde{f}(s, x) := f(s, x) + \delta_{\Gamma(s)}(x). \tag{5}$$

In this way, (2) amounts to the minimization of $I_{\tilde{f}}$ over $L^p(\Omega; \mathbb{R}^n)$.

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

- (i) $\hat{x} \in \operatorname{argmin}_{x \in L^p(\Omega; \mathbb{R}^n)} I_{\tilde{f}}[x]$;
- (ii) $\hat{x}(s) \in \operatorname{argmin}_{y \in \mathbb{R}^n} \tilde{f}(s, y)$, for a.e. $s \in \Omega$.

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

$$\begin{aligned} \mathcal{E}_{\tilde{f}}(s) &= \operatorname{epi} \tilde{f}(s, \cdot) \\ &= \operatorname{epi} f(s, \cdot) \cap (\Gamma(s) \times \mathbb{R}_+). \end{aligned}$$

Since intersections and products preserve measurability (Ref. 1, Proposition 14.11), the normality of \tilde{f} is implied by the normality of f and the closed-valuedness and measurability of Γ . The L^p spaces are decomposable (Ref. 1, Definition 14.59) and so the mentioned Theorem 14.60 from Ref. 1 can be specialized to the above form. □

Remark 3.1. In most cases, f is 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 . However, the considered structure enables 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 5.1 and formulate immediately the optimality conditions for (2).

Theorem 3.2. Let \hat{x} be a (local) solution of (2) and let the assumptions of Theorem 3.1 be fulfilled. Then, for a.e. $s \in \Omega$, there exist a vector \hat{v}_s^* and a real $\lambda_s \geq 0$, not both simultaneously equal to zero, such that

$$(\hat{v}_s^* - \lambda_s) \in N_{\operatorname{epi} f(s, \cdot)}(\hat{x}, f(s, \hat{x}(s))), \quad -\hat{v}_s^* \in N_{\Gamma(s)}(\hat{x}(s)). \tag{6}$$

If the constraint qualification

$$\partial^\infty f(s, \hat{x}(s)) \cap (-N_{\Gamma(s)}(\hat{x}(s))) = \{0\}, \quad \text{for a.e. } s \in \Omega, \tag{7}$$

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

$$0 \in \partial f(s, \hat{x}(s)) + N_{\Gamma(s)}(\hat{x}(s)). \tag{8}$$

The subdifferentials in (7), (8) 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 having the shape similar to the “Mercedes star”,

$$\min \int_0^1 (|x_1(s)| + |x_2(s)|) ds, \tag{9a}$$

$$\text{s.t. } x(s) \in \left\{ y \in \mathbb{R}^2 \mid k(s) - |y_1| - y_2 = 0 \right\} \cup \left\{ y \in \mathbb{R}^2 \mid y_1 = 0, y_2 \geq k(s) \right\}, \tag{9b}$$

a.e. in $[0, 1]$,

$$x \in L^2(0, 1; \mathbb{R}^2), \tag{9c}$$

where $k \in L^2(0, 1; \mathbb{R}^2)$ is a given function. We note that

$$\hat{x}(\cdot) = (0, \max\{0, k(\cdot)\})$$

is a solution of (9). It is not difficult to check that the all the assumptions of Theorem 3.1 are fulfilled.

Further, in this simple situation, the constraint qualification (7) is fulfilled and the normal cones $N_{\Gamma(s)}(\hat{x}(s))$ as well as the subdifferentials $\partial f(s, \hat{x}(s))$ can be computed easily:

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

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

Furthermore,

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

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

The optimality condition (8) is thus evidently fulfilled, because e.g. the function $\hat{v}^*[[0, 1] \rightarrow \mathbb{R}^2]$ defined by

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

fulfills the relation

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

In what follows, we examine three particular situations in which the statement of Theorem 3.2 can be strengthened or specialized.

Corollary 3.1. Let \hat{x} be a solution of (2) and let the assumptions of Theorem 3.1 be fulfilled. Further, assume that $p < \infty$ and that there exists a function $k \in L^q(\Omega)$, $(1/p + 1/q = 1)$, such that, for all $s \in \Omega$,

$$|f(s, y_1) - f(s, y_2)| \leq k(s)|y_1 - y_2|, \quad \text{for all } y_1, y_2 \text{ in } \mathbb{R}^n. \quad (10)$$

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

$$\hat{v}^*(s) \in \partial f(s, \hat{x}(s)) \cap (-N_{\Gamma(s)}(\hat{x}(s))), \quad \text{for a.e. } s \in \Omega. \quad (11)$$

Proof. We start with the observation that, due to (10), $f(s, \cdot)$ is Lipschitz on \mathbb{R}^n , for a.e. $s \in \Omega$, and consequently

$$\partial^\infty f(s, \hat{x}(s)) = \{0\}, \quad \text{for a.e. } s \in \Omega.$$

This implies that the constraint qualification (7) is fulfilled. Further, by virtue of (10), for all $s \in \Omega$ one has

$$\partial f(s, \hat{x}(s)) \subset \bar{\partial} f(s, \hat{x}(s)) \subset k(s)\mathbb{B},$$

so that each measurable selection of $\partial f(s, \hat{x}(s))$, a.e. on Ω , belongs to $L^q(\Omega; \mathbb{R}^n)$. Thus, it remains to show that there is a measurable function $\hat{v}^*[\Omega \rightarrow \mathbb{R}^n]$ satisfying condition (11). However, this follows immediately

from Ref. 1, Proposition 14.11 (measurability of intersections) and Ref. 1, 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. \square

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

$$\Gamma(s) = \{y \in \Theta(s) \mid G(s, y) \in \Lambda(s)\}, \tag{12}$$

$G[\Omega \times \mathbb{R}^n \rightarrow \mathbb{R}^m]$ being a Carathéodory map and the multifunctions $\Theta[\Omega \rightrightarrows \mathbb{R}^n]$, $\Lambda[\Omega \rightrightarrows \mathbb{R}^m]$ being closed-valued and measurable. As shown in Ref. 12, Theorem 8.2.9, under these conditions, Γ is also closed-valued and measurable.

Corollary 3.2. Let \hat{x} be a solution of (2), where Γ is given in the form (12). Further, suppose that the assumptions of Theorem 3.1 are fulfilled, $f(s, \cdot)$ for a.e. $s \in \Omega$ is Lipschitz around $\hat{x}(s)$, and the following qualification conditions hold true:

$$D^*G(s, \hat{x}(s)) \circ N_{\Lambda(s)}(G(s, \hat{x}(s))) \cap (-N_{\Theta(s)}(\hat{x}(s))) = \{0\}, \tag{13a}$$

$$N_{\Lambda(s)}(G(s, \hat{x}(s))) \cap \text{Ker } D^*G(s, \hat{x}(s)) = \{0\}. \tag{13b}$$

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

$$0 \in \partial f(s, \hat{x}(s)) + D^*G(s, \hat{x}(s))(\hat{y}_s^*) + N_{\Theta(s)}(\hat{x}(s)). \tag{14}$$

Remark 3.2. Analogously to Theorem 3.2, the coderivatives in (13), (14) concern the map $G(s, \cdot)$.

Proof of Corollary 3.2. Under the qualification conditions (13), Theorem 5.2 provides us with an upper approximation of $N_{\Gamma(s)}(\hat{x}(s))$ in the form

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

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

The mapping which assigns the vector \hat{y}_s^* to a.e. $s \in \Omega$ can be viewed as the Karush–Kuhn–Tucker (KKT) function of the program (2) with Γ given by (12). In this connection, a natural question arises: Under what assumptions there exists a measurable KKT function? This question is answered in the next statement.

Corollary 3.3. In addition to the assumptions of Corollary 3.2, suppose that $p < \infty$ and that condition (10) holds true. Then, there exist an element $\hat{v}^* \in L^q(\Omega; \mathbb{R}^n)$ and a measurable KKT function $s \mapsto \hat{y}^*(s)$ such that

$$\hat{y}^*(s) \in N_{\Lambda(s)}(G(s, \hat{x}(s))), \tag{15a}$$

$$\hat{v}^*(s) \in \partial f(s, \hat{x}(s)) \cap [-D^*G(s, \hat{x}(s))(\hat{y}^*(s)) - N_{\Theta(s)}(\hat{x}(s))], \tag{15b}$$

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

Proof. The statement of Corollary 3.3 can be reformulated in the following form: There exist mappings $\hat{v}^*[\Omega \rightarrow \mathbb{R}^n]$, $\hat{w}^*[\Omega \rightarrow \mathbb{R}^n]$, $\hat{y}^*[\Omega \rightarrow \mathbb{R}^m]$ such that

$$\hat{v}^*(s) \in \partial f(s, \hat{x}(s)), \tag{16a}$$

$$-\hat{v}^*(s) - \hat{w}^*(s) \in N_{\Theta(s)}(\hat{x}(s)), \tag{16b}$$

$$\hat{y}^*(s) \in N_{\Lambda(s)}(G(s, \hat{x}(s))), \tag{16c}$$

$$(\hat{w}^*(s), -\hat{y}^*(s)) \in N_{\text{gph}G(s, \cdot)}(\hat{x}(s), G(s, \hat{x}(s))), \tag{16d}$$

for a.e. $s \in \Omega$. In the above relations, $\text{gph } G(s, \cdot)$ denotes the set $\{(y, z) \in \mathbb{R}^n \times \mathbb{R}^m \mid z = G(s, y)\}$. Using the argumentation of Corollary 3.1, we infer easily that $\hat{v}^* \in L^q(\Omega; \mathbb{R}^n)$ provided it is measurable. Thus, it suffices to prove the existence of measurable functions \hat{v}^* , \hat{w}^* , \hat{y}^* satisfying (16). To this purpose, we introduce the function $q[\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m]$ defined by

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

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

$$\begin{aligned} Q(s) := & \partial f(s, \hat{x}(s)) \times N_{\Theta(s)} \times N_{\Lambda(s)}(G(s, \hat{x}(s))) \\ & \times N_{\text{gph}G(s, \cdot)}(\hat{x}(s), G(s, \hat{x}(s))). \end{aligned} \tag{17}$$

The functions \hat{v}^* , \hat{w}^* , \hat{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 Carathéodory function and \hat{x} is in $L^p(\Omega; \mathbb{R}^n)$, we observe that the function $G(\cdot, \hat{x}(\cdot))$

and $(\hat{x}(\cdot), G(\cdot, \hat{x}(\cdot)))$ are measurable. The first three multifunctions in the Cartesian product (17) are measurable due to the results from Ref. 1 mentioned in the proof of Corollary 3.1. To see the measurability of the 4th multifunction, note first that

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

Hence, Theorem 8.2.9 from Ref. 12 applies and yields the measurability of $\text{gph } G(s, \cdot)$. The measurability of $N_{\text{gph}G(s, \cdot)}(\hat{x}(s), G(s, \hat{x}(s)))$ follows then from Ref. 1, Theorem 14.26. The product multifunction Q is thus also measurable [Ref. 1, Proposition 14.11(d)] and we can apply Ref. 12, 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. \square

Remark 3.3. If $G(s, \cdot)$ is Lipschitz near $\hat{x}(s)$ for some $s \in \Omega$, then one has

$$\begin{aligned} D^*G(s, \hat{x}(s))(\hat{y}^*(s)) &= \partial(\hat{y}^*(s), G(s, \cdot))(\hat{x}(s)) \subset \{C(s)^T \hat{y}^*(s) \mid C(s) \\ &\in \bar{\partial}G(s, \hat{x}(s))\}; \end{aligned} \tag{18}$$

see Ref. 2. The 2nd or the 3rd set in (18) can sometimes be computed more easily than the coderivative $D^*G(s, \hat{x}(s))(\hat{y}^*(s))$.

Let us comment on the optimality conditions (11), (15). The limiting subdifferential and the limiting normal cone $N_{\Gamma(s)}(\hat{x}(s))$ are in nonconvex situations usually much smaller than the corresponding objects of Clarke for which an analogous condition has been proved in Ref. 10. This is illustrated strikingly in Example 3.1, where the Clarke normal cone to $\Gamma(s)$ at $\hat{x}(s)$ is the whole space \mathbb{R}^2 provided $k(s) \geq 0$. Hence, our conditions are sharper. In Corollary 3.2.3, 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 $\Phi[\mathbb{R}^p \rightrightarrows \mathbb{R}^q]$ is pseudo-Lipschitz around $(a, b) \in \text{gph } \Phi$, provided there exists neighborhoods \mathcal{U} of a and \mathcal{V} of b and a modulus $\ell \geq 0$ such that

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

Let $G(s, \cdot)$ be Lipschitz near $\hat{x}(s)$ for a.e. $s \in \Omega$. From Theorem 5.3, the following constraint qualification:

$$0 \in D^*G(s, \hat{x}(s))(\xi) + N_{\Theta(s)}(\hat{x}(s)), \xi \in N_{\Lambda(s)}(G(s, \hat{x}(s))), \text{ imply } \xi = 0 \tag{19}$$

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

Theorem 3.3. Let \hat{x} be a solution of (2), where Γ is given in the form (12) with $G(s, \cdot)$ Lipschitz near $\hat{x}(s)$ for a.e. $s \in \Omega$. Let $p < \infty$ and let condition (10) be fulfilled. Further assume that, for a.e. $s \in \Omega$, the constraint qualification (19) holds true and that the function $\rho(\cdot)$, assigning $s \in \Omega$ the modulus of pseudo-Lipschitz continuity of Σ around $(0, \hat{x}(s))$, belongs to $L^p(\Omega)$. Then, there exists a KKT function $\hat{u}^* \in L^1(\Omega; \mathbb{R}^m)$ such that

$$\hat{u}^*(s) \in N_{\Lambda(s)}(G(s, \hat{x}(s))), \tag{20a}$$

$$0 \in \partial f(s, \hat{x}(s)) + D^*G(s, \hat{x}(s))(\hat{u}^*(s)) + N_{\Theta(s)}(\hat{x}(s)), \tag{20b}$$

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

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

$$\min \quad f(s, v) + R(s)|u|, \tag{21a}$$

$$\text{s.t.} \quad (u, v) \in \text{gph } \Sigma(s, \cdot), \tag{21b}$$

see Ref. 16, Lemma 3.1, provided the penalty parameter $R(s)$ is sufficiently large. Writing down the optimality conditions for (21), we obtain the existence of a pair $(\hat{u}^*(s), -\hat{v}^*(s)) \in N_{\text{gph}\Sigma(s, \cdot)}(0, \hat{x}(s))$ such that

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

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

$$\hat{u}^*(s) \in D^*\Sigma(s, 0, \hat{x}(s))(\hat{v}^*(s))$$

and, by invoking (40), we infer that

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

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

It follows that \hat{u}^* is a KKT function and $\hat{v}^* \in L^q(\Omega; \mathbb{R}^n)$. Moreover, by virtue of Ref. 5, Theorem 3.2, for a.e. $s \in \Omega$ one has

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

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

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

and we are done. □

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

3.2. Reduced Optimization. In Ref. 1, 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 an integrand. Such a situation arises typically 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 strengthen the optimality conditions of Ref. 7. Correspondingly, in this subsection we will be dealing with the optimization problem

$$\min \quad h(z) + I_g[z, x], \tag{22a}$$

$$\text{s.t.} \quad z \in D \subset \mathbb{R}^m, \tag{22b}$$

$$x(s) \in \Gamma(s, z), \quad \text{for a.e. } s \in \Omega, \tag{22c}$$

$$x \in L^p(\Omega; \mathbb{R}^n), \tag{22d}$$

where $h[\mathbb{R}^m \rightarrow \mathbb{R}]$ 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 assume that g maps $\Omega \times \mathbb{R}^m \times \mathbb{R}^n$ into \mathbb{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 Ref. 1, Example 14.62 attains the following form.

Theorem 3.4. Consider problem (22) and suppose, in addition to the posed assumptions, that for a.e. $s \in \Omega$ the essential integrand

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

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

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

let $h + I_f$ be proper. Then, the following two statements are equivalent:

- (i) $(\hat{z}, \hat{x}) \in D \times L^p(\Omega; \mathbb{R}^n)$ is a (local) solution of (22);
- (ii) \hat{z} is a (local) solution of the optimization problem

$$\min h(z) + \int_{\Omega} f(s, z)\mu(ds), \tag{23a}$$

$$\text{s.t. } z \in D, \tag{23b}$$

and

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

Since $z \in \mathbb{R}^m$, it is clear that, in Theorem 3.4, it suffices to assume the level-boundedness of \tilde{g} in y uniformly only with respect to z from a neighborhood of \hat{z} . On the basis of Theorem 3.4, we can now derive the optimality conditions for problem (22), i.e., the counterpart of Ref. 7, Theorem 5. In the first step, we invoke Ref. 3, Theorem 4.1 and observe that, under the posed assumptions, for all $s \in \Omega$ one has

$$\partial^\infty f(s, \hat{z}) \subset \bigcup_{y_0 \in \operatorname{argmin}_{y \in \mathbb{R}^n} \tilde{g}(s, \hat{z}, y)} D^*\Gamma(s, \hat{z}, y_0)(0). \tag{24}$$

Therefore, whenever the set on the right-hand side of the inclusion (24) contains only the zero vector, the value function $f(s, \cdot)$ is Lipschitz near \hat{z} with a Lipschitz modulus $k(s)$, $s \in \Omega$.

Theorem 3.5. Let $\hat{z} \in D$ be a (locally) optimal value of the variable z in (22) and let all assumptions of Theorem 3.4 be fulfilled. Further, assume that $D^*\Gamma(s, \hat{z}, y_0)(0) = \{0\}$, for all $y_0 \in \operatorname{argmin}_{y \in \mathbb{R}^n} \tilde{g}(s, \hat{z}, y)$, $s \in \Omega$, and that the Lipschitz modulus k is integrable. Then, there exists an integrable mapping $z^*[\Omega \rightarrow \mathbb{R}^m]$ such that

$$0 \in \partial h(\hat{z}) + \int_{\Omega} z^*(s)\mu(ds) + N_D(\hat{z}) \tag{25}$$

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

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

with

$$\Delta(s) = \left\{ v^* \in \mathbb{R}^m \mid (v^*, x^*) \in \partial g(s, \hat{z}, x_s^0) \right\}, \quad x_s^0 \in \operatorname{argmin}_{y \in \mathbb{R}^n} \tilde{g}(s, \hat{z}, y).$$

Proof. By Theorem 3.4, it suffices to derive optimality conditions for the finite-dimensional problem (23), where the only difficult part represents the special integral functional

$$J(z) := \int_{\Omega} f(s, z) \mu(ds).$$

For all z_1, z_2 sufficiently close to \hat{z} , one has

$$\begin{aligned} |J(z_1) - J(z_2)| &\leq \int_{\Omega} |f(s, z_1) - f(s, z_2)| \mu(ds) \\ &\leq |z_1 - z_2| \int_{\Omega} k(s) \mu(ds); \end{aligned}$$

i.e., J is Lipschitz near \hat{z} by the assumed integrability of Lipschitz modulus k . By virtue of Ref. 9, Theorem 2.7.2, to each $\xi \in \bar{\partial}J(z)$, there exists an integrable mapping $\tilde{z}^*[\Omega \rightarrow \mathbb{R}^m]$ such that

$$\xi = \int_{\Omega} \tilde{z}^*(s) \mu(ds)$$

and

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

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

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

Now, the Lyapunov–Aumann theorem implies the existence of an integrable selection z^* such that

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

and relation (25) is fulfilled. It remains to recall from Ref. 3, Theorem 4.1, that, under our assumptions,

$$\begin{aligned} \partial f(s, \hat{z}) \subset \bigcup_{x_s^0 \in \text{argmin}_{y \in \mathbb{R}^n} \tilde{g}(s, \hat{z}, y)} &\left\{ z_1^* + z_2^* \mid z_1^* \in D^* \Gamma(s, \hat{z}, x_s^0)(x^*), (z_2^*, x^*) \right. \\ &\left. \in \partial g(s, \hat{z}, x_s^0) \right\} \end{aligned}$$

and we are done. □

Under our assumptions, the multifunction

$$P : (s, z) \rightrightarrows \operatorname{argmin}_{y \in \mathbb{R}^n} \tilde{g}(s, z, y)$$

is closed-valued and measurable (Ref. 1, Theorem 14.37). Of course, this does not imply the measurability of the map $s \mapsto x_s^0$ in Theorem 3.5. Nevertheless, if P possesses a measurable selection \bar{y} such that $\bar{y}(s, \cdot)$ is continuous at \hat{z} for a.e. $s \in \Omega$, then one can set

$$x_s^0 = \bar{y}(s, \hat{z}), \quad \text{for a.e. } s \in \Omega.$$

This idea comes from Ref. 17 and follows in our context from the following observations. Let $s \in \Omega$ be fixed and $\xi \in \partial f(s, \hat{z})$. Then, as explained e.g. in Refs. 1–2, there are sequences $z_i \rightarrow \hat{z}$ and $\xi_i \rightarrow \xi$ such that, for all i , the element ξ_i is a regular subgradient of f at (s, z_i) ; i.e., for all $z \in \mathbb{R}^m$,

$$f(s, z) - f(s, z_i) \geq \langle \xi_i, z - z_i \rangle - o(\|z - z_i\|).$$

Definitely, for all $z \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$,

$$f(s, z) - f(s, z_i) \leq \tilde{g}(s, z, y) - \tilde{g}(s, z_i, \bar{y}(s, z_i)).$$

This implies that $(\xi_i, 0)$ is a regular subgradient of \tilde{g} at $(s, z_i, \bar{y}(s, z_i))$. Moreover, since $z_i \rightarrow \hat{z}$, $\bar{y}(s, z_i) \rightarrow \bar{y}(s, \hat{z})$ and $\xi_i \rightarrow \xi$, it follows that $(\xi, 0) \in \partial \tilde{g}(s, \hat{z}, \bar{y}(s, \hat{z}))$. It remains to apply the sum rule for limiting subdifferentials which yields the decomposition

$$\xi = z_1^* + z_2^*,$$

with

$$z_1^* \in D^* \Gamma(s, \hat{z}, \bar{y}(s, \hat{z}))(x^*) \quad \text{and} \quad (z_2^*, x^*) \in \partial g(s, \hat{z}, \bar{y}(s, \hat{z})).$$

Theorem 3.5 together with the above discussion represents a sharper variant of Ref. 7, Theorem 5. Indeed, 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 Ref. 7 that $\Gamma(s, \cdot)$ is given by parameter-dependent inequalities and derive readily the counterpart of Ref. 7, Theorem 7. Unfortunately this approach, leading to an improvement in the case of two-stage stochastic programs, could not be applied in the case of multi-stage stochastic programs as we will see in Section 4.

4. Nonpointwise Constraints

4.1. Fuzzy Optimality Conditions. This section deals with the optimization problem

$$\min I_f[x], \tag{26a}$$

$$\text{s.t. } x(s) \in \Gamma(s), \text{ for a.e. } s \in \Omega, \tag{26b}$$

$$x \in C, \tag{26c}$$

where f and Γ fulfill the assumptions posed in connection with problem (2) and C is a nonempty and closed subset of $L^2(\Omega; \mathbb{R}^n)$. As pointed out in the introduction, C cannot be expressed in pointwise form; therefore, the approach of Section 3 is not applicable. Clearly, (26) amounts to the problem

$$\min I_{\tilde{f}}[x], \tag{27a}$$

$$\text{s.t. } x \in C, \tag{27b}$$

where \tilde{f} is the essential integrand introduced in (5). Since the sum $I_{\tilde{f}} + \delta_C$ is lsc, from the definition of the proximal subdifferential it follows that

$$0 \in \partial_P(I_{\tilde{f}} + \delta_C)(\hat{x}), \tag{28}$$

whenever \hat{x} is a (local) minimum in (26) and $I_f[\hat{x}] \in \mathbb{R}$. To express the relation (28) in terms of the problem data, we invoke the proximal variant of the weak fuzzy sum rule from Ref. 18. For the sake of simplicity, we formulate this result 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 let $f_1, f_2[X \rightarrow \bar{\mathbb{R}}]$ 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 $x_1, x_2 \in \mathbb{B}(x; \varepsilon)$, $x_1^* \in \partial_P f(x_1)$, $x_2^* \in \partial_P f(x_2)$ such that

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

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

and

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

In the proof one needs just to combine the ideas from the proof of Ref. 18, Theorem 2.7, with the strong proximal fuzzy sum rule in Ref. 15. On the basis of Theorem 4.1, we obtain the following weak fuzzy optimality conditions for problem (26).

Theorem 4.2. Let \hat{x} be a (local) solution of problem (26). In addition to the posed assumptions, suppose that either (A) or (A)' below is satisfied:

(A) there exist a function $k \in L^2(\Omega; \mathbb{R})$ such that, for all $s \in \Omega$,

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

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

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

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

$$|I_f[x_1] - I_f[\hat{x}]| < \varepsilon, \quad x_2(s) \in \Gamma(s), \quad \text{for a.e. } s \in \Omega, \quad x_3 \in C, \tag{29a}$$

$$\|x_1 - \hat{x}\| \leq \varepsilon, \quad \|x_2 - \hat{x}\| \leq \varepsilon, \quad \|x_3 - \hat{x}\| \leq \varepsilon, \tag{29b}$$

$$x_1^*(s) \in \partial_P f(s, x_1(s)), \quad \text{for a.e. } s \in \Omega, \tag{29c}$$

$$x_2^*(s) \in N_{\Gamma(s)}^P(x_2(s)), \quad \text{for a.e. } s \in \Omega, \tag{29d}$$

$$x_3^* \in N_C^P(x_3), \tag{29e}$$

$$0 \in x_1^* + x_2^* + x_3^* + V. \tag{29f}$$

Before, we prove this statement, we note that, under (A), I_f is defined and (globally) Lipschitz on $L^2(\Omega; \mathbb{R}^n)$; under (A)', I_f is Lipschitz only on bounded subsets of $L^2(\Omega; \mathbb{R}^n)$; see Ref. 9, Theorem 2.7.5.

Proof of Theorem 4.2. Since both functions $I_{\tilde{f}}$ and δ_C are lsc, we are entitled to apply Theorem 4.1 to (28). This yields the existence of functions $\tilde{x}, x_3 \in \mathbb{B}(\hat{x}; \varepsilon/2)$, $\tilde{x}^* \in \partial_P I_{\tilde{f}}[\tilde{x}]$, $x_3^* \in N_C^P(x_3)$ such that

$$|I_f[\tilde{x}] - I_f[\hat{x}]| < \varepsilon/2, \tag{30a}$$

$$\tilde{x}(s) \in \Gamma(s), \quad \text{for a.e. } s \in \Omega, \tag{30b}$$

$$x_3 \in C, \tag{30c}$$

$$0 \in \tilde{x}^* + x_3^* + U, \tag{30d}$$

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

$$U + \mathbb{B}(0; \varepsilon/2) \subset V.$$

The first two relations in (30) follow from the inequality

$$|I_{\tilde{f}}[\tilde{x}] - I_{\tilde{f}}[\hat{x}]| < \varepsilon/2.$$

Now, we apply the strong proximal sum rule (Ref. 15, Theorem 2) to the relation $\tilde{x}^* \in \partial_P(I_f + I_{\delta_\Gamma})(\tilde{x})$. This is possible, because I_f is Lipschitz on a neighborhood of \hat{x} so that the uniform lower semicontinuity property is satisfied. It follows that there exist functions $x_1, x_2 \in \mathbb{B}(\tilde{x}; \varepsilon/2)$, $x_1^* \in \partial_P I_f[x_1]$, $x_2^* \in \partial_P I_{\delta_\Gamma}[x_2]$ such that

$$|I_f[x_1] - I_f[\tilde{x}]| < \varepsilon/2 \tag{31a}$$

$$x_2(s) \in \Gamma(s), \quad \text{for a.e. } s \in \Omega, \tag{31b}$$

$$\tilde{x}^* \in x_1^* + x_2^* + \mathbb{B}(0, \varepsilon/2). \tag{31c}$$

It remains to put relations (30), (31) together and take into account the implication (4). □

Since V is a weak neighborhood, it is not possible to get a limiting version of (29) by letting $\varepsilon \rightarrow 0$. On the other hand, in Theorem 4.2, we do not have to consider any constraint qualification arising usually in optimality conditions of the KKT type, and the subdifferentials and normal cone in (29) 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 \mathbb{R}^P (Definition 2.1). The limiting normal cone to Ξ at \bar{z} is then the set

$$\begin{aligned} N_\Xi(\bar{z}) &:= \limsup_{z \rightarrow \bar{z}} \hat{N}_\Xi(z) \\ &= \left\{ z^* \in Z^* \mid \exists \text{ sequences } z_k \rightarrow \bar{z} \text{ and } z_k^* \xrightarrow{w^*} z^*, \text{ with} \right. \\ &\quad \left. z_k^* \in \hat{N}_\Xi(z_k), \forall k \in \mathbb{N} \right\}. \end{aligned}$$

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

$$z_k^* \in \hat{N}_\Xi(z_k), \quad z_k \rightarrow \bar{z}, \quad z_k^* \xrightarrow{w^*} 0,$$

contains a subsequence $\{(z_{k'}, z_{k'}^*)\}$ with $\|z_{k'}^*\| \rightarrow 0$. To derive the optimality conditions for problem (26) in a standard KKT form, we need to ensure the inclusion

$$N_{C \cap D}(\hat{x}) \subset N_C(\hat{x}) + N_D(\hat{x}), \tag{32}$$

where

$$D := \{x \in L^2(\Omega; \mathbb{R}^n) \mid x(s) \in \Gamma(s), \text{ for a.e. } s \in \Omega\}.$$

Following Ref. 19, this can be done by requiring that either C or D is sequentially normally compact at \hat{x} and

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

Unfortunately, we are not able to prove the sequential normal compactness of any from the sets C, D in the applications that we have in mind, in particular in multistage stochastic programs. Additionally, a pointwise description of $N_D(\hat{x})$ in terms of Γ is generally not available to our knowledge. To summarize, for problem (26), we dispose with the fuzzy optimality conditions stated in Theorem 4.2, which are valid under very weak conditions imposed on the problem data. However, if these data fulfill some more restrictive assumptions, optimality conditions in the classical KKT form can be derived. In the next statement, we require among others the regularity of $\Gamma(s)$ at $\hat{x}(s)$ for a.e. $s \in \Omega$ [i.e., $N_{\Gamma(s)}(\hat{x}(s))$ equals the negative polar of the Bouligand (contingent) cone to $\Gamma(s)$ at $\hat{x}(s)$]. Other regularity notions and their relations are studied in Ref. 20.

Theorem 4.3. Let \hat{x} be a (local) solution of (26), where C is sequentially normally compact at \hat{x} and $\Gamma(s)$ is regular at $\hat{x}(s)$ for a.e. $s \in \Omega$. Further, suppose that either assumption (A) or assumption (A') from Theorem 4.2 are fulfilled and that the constraint qualification

$$N_C(\hat{x}) \cap \{x^* \in L^2(\Omega; \mathbb{R}^n) \mid -x^*(s) \in N_{\Gamma(s)}(\hat{x}(s)), \text{ a.e. in } \Omega\} = \{0\} \tag{33}$$

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

$$\begin{aligned} x_1^*(s) &\in \bar{\partial} f(s, \hat{x}(s)), & \text{for a.e. } s \in \Omega, \\ x_2^*(s) &\in N_{\Gamma(s)}(\hat{x}(s)), & \text{for a.e. } s \in \Omega, \\ x_3^*(s) &\in N_C(\hat{x}), \\ 0 &= x_1^* + x_2^* + x_3^*. \end{aligned}$$

Proof. Due to the regularity of $\Gamma(s)$ at $\hat{x}(s)$, for a.e. $s \in \Omega$, one can invoke Ref. 12, Corollary 8.5.2, according to which

$$N_D(\hat{x}) = \{x^* \in L^2(\Omega; \mathbb{R}^n) \mid x^*(s) \in N_{\Gamma(s)}(\hat{x}(s)), \text{ a.e. in } \Omega\}.$$

Since C is sequentially normally compact at \hat{x} , the constraint qualification (33) ensures by virtue of Ref. 19, Proposition 2.2, the inclusion (32). Under assumption (A) or (A') the objective is Lipschitz around \hat{x} . Therefore,

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

It remain to observe that

$$\partial I_f[\hat{x}] \subset \bar{\partial} I_f[\hat{x}] \subset \{\xi \in L^2(\Omega; \mathbb{R}^n) \mid \xi(s) \in \bar{\partial} f(s, \hat{x}(s)), \text{ a.e. in } \Omega\},$$

and we are done. □

In Ref. 21, one can find useful conditions ensuring the sequential normal compactness of C for sets with various frequently appearing structures.

A favorable situation for the construction of the classical KKT optimality conditions for (26) arises if C is a decomposable subspace of $L^2(\Omega; \mathbb{R}^n)$ or even $L^p(\Omega; \mathbb{R}^n)$, $1 \leq p \leq \infty$. It is shown in Ref. 22, Theorem 3.1, that a nonempty closed subset C of $L^p(\Omega; \mathbb{R}^n)$, $1 \leq p < \infty$, is decomposable iff there exists a closed-valued measurable multifunction Θ from Ω to \mathbb{R}^n such that C coincides with the set of all functions in $L^p(\Omega; \mathbb{R}^n)$ that are measurable selections of Θ , a.e. in Ω . In such a case, the approach via extended Lipschitz integrands from Ref. 10 can be used and it is not necessary to assume either the sequential normal compactness of C or any constraint qualification of the type (33). In the rest of this section, we examine the form of conditions (29) in the 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 is based on only 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 \mathbb{R}^{n_k} is measurable with respect to \mathcal{S}_k . The latter property is called nonanticipativity. Furthermore, we assume that

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

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 Refs. 23–24 and formulate the sequential decision model as a mathematical program in a space of integrable functions [here, the space $L^2(\Omega; \mathbb{R}^n)$, 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 $\Omega \times \mathbb{R}^n$ to $\bar{\mathbb{R}}$ 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 (nonpointwise) constraints

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

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

$$\min \quad I_f[x] = \int_{\Omega} f(s, x(s)) \mu(ds), \tag{34a}$$

$$\text{s.t.} \quad \varphi(s, x(s)) \leq 0, \quad \text{for a.e. } s \in \Omega, \tag{34b}$$

$$x \in C := \{x \in L^2(\Omega; \mathbb{R}^n) | x_k = \mathbb{E}[x_k | \mathcal{S}_k], k = 1, \dots, K\}. \tag{34c}$$

In general, the set C forms a closed linear subspace of $L^2(\Omega; \mathbb{R}^n)$ and has the specific structure $\mathbb{R}^{n_1} \times L^2(\Omega; \mathbb{R}^{n_2})$ in the two-stage situation (i.e., $K = 2$). While the latter structure allows one to reduce the model (34) to the model (22), the situation for $K > 2$ becomes quite different, since C is not decomposable in general.

Theorem 4.4. Let $\hat{x} \in L^2(\Omega; \mathbb{R}^n)$ be a (local) solution of problem (34). Suppose that $\varphi(s, \cdot)$ is locally Lipschitz on \mathbb{R}^n for each $s \in \Omega$ and that either assumption (A) or assumption (A)' of Theorem 4.2 is fulfilled. Further, assume that, for all $x \in L^2(\Omega; \mathbb{R}^n)$ such that $x(s) \in \Gamma(s)$, for a.e. $s \in \Omega$, one has

$$0 \notin \left\{ \sum_{i \in I(s, x(s))} \lambda_i \xi_i | \xi_i \in \partial \varphi^i(s, x(s)), \lambda_i \geq 0, \sum_{i \in I(s, x(s))} \lambda_i = 1 \right\},$$

for a.e. $s \in \Omega$, (35)

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 $L^2(\Omega; \mathbb{R}^n)$, there exist functions $x_1, x_2, x_3, x_1^*, x_2^*, x_3^* \in L^2(\Omega; \mathbb{R}^n)$ such that:

- (i) $|I_f[x_1] - I_f[\hat{x}]| < \varepsilon, \varphi(s, x_2(s)) \leq 0$, 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) $\|x_1 - \hat{x}\| \leq \varepsilon, \|x_2 - \hat{x}\| \leq \varepsilon, \|x_3 - \hat{x}\| \leq \varepsilon$;
- (iii) for a.e. $s \in \Omega$, one has $x_1^*(s) \in \partial_P f(s, x_1(s))$ and there exist multipliers $\lambda_s \in \mathbb{R}_+^m$, depending measurably on s , such that

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

- (iv) for $x_3^* = (x_{31}^*, x_{32}^*, \dots, x_{3K}^*)$, it holds that $\mathbb{E}[x_{3k}^* | \mathcal{S}_k] = 0$, a.e. for $k = 1, \dots, K$;
- (v) $0 \in x_1^* + x_2^* + x_3^* + V$.

Proof. It suffices to express the cones $N_{\Gamma(s)}^P(x_2(s))$, for a.e. $s \in \Omega$, and $N_C^P(x_3)$ from Theorem 4.2 in terms of the data of the problem (34). To this purpose, we invoke Ref. 2, Corollary 4.4.2, according to which the inclusion

$$\begin{aligned} N_{\Gamma(s)}^P(x_2(s)) &\subseteq N_{\Gamma(s)}(x_2(s)) \\ &\subseteq \bigcup \left\{ \sum_{i=1}^m \lambda^i \partial \varphi^i(s, x_2(s)) \mid \lambda^i \geq 0, \lambda^i \varphi^i(s, x_2(s)) = 0 \right\} \end{aligned} \quad (36)$$

holds, whenever the constraint qualification (35) is fulfilled. Since the sub-gradient mappings $s \mapsto \partial \varphi^i(s, x_2(s))$ are closed-valued and measurable by Ref. 1, Theorem 14.56, for $i = 1, \dots, m$, the set-valued mapping, which assigns s the set $\Theta(s)$ on the right-hand side of (36), is closed-valued and measurable, too; see Ref. 1, Exercise 14.12. By appealing to the implicit measurable functions theorem (see Ref. 1, Theorem 14.16), there exist measurable functions $s \mapsto \lambda_s^i$ from Ω to \mathbb{R}_+^m such that $\lambda_s^i \varphi^i(s, x_2(s)) = 0, i = 1, \dots, m$, and the measurable selection x_2^* of Θ is a.e. contained in the set

$$\sum_{i=1}^m \lambda_s^i \partial \varphi^i(s, x_2(s)).$$

The cone $N_C^P(x_3)$ coincides with the normal cone of convex analysis to the closed linear subspace C of $L^2(\Omega; \mathbb{R}^n)$ 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 that

$$C^\perp = \{x^* \in L^2(\Omega; \mathbb{R}^n) \mid \mathbb{E}[x_k^* | \mathcal{S}_k] = 0, \text{ a.e.}, k = 1, \dots, K\},$$

completing the proof. \square

The usage of limiting subdifferentials in the upper approximation of $N_{\Gamma(s)}^P(x_2(s))$ makes possible to evaluate all needed subdifferentials $\partial\varphi^i, i \in \{1, 2, \dots, m\}$, at the points $x_2(s), s \in \Omega$. However, one could 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 (35) which has to be replaced by

$$\liminf_{y \rightarrow x(s)} \text{dist}(0 | \partial_P \varphi^i(s, y)) > 0, \quad \text{for a.e. } s \in \Omega, \quad i = 1, 2, \dots, m. \quad (37)$$

Then, to each $\varepsilon > 0$ and to each weak neighborhood V of 0 in $L^2(\Omega; \mathbb{R}^n)$, there exist functions $x_1, x_2, x_3, x_1^*, x_2^*, x_3^* \in L^2(\Omega; \mathbb{R}^n)$ such that the assertions (i), (ii), (iv), (v) of Theorem 4.4 hold true and assertion (iii) is replaced by (iii)' 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 $y_{is} \in \mathbb{B}(x_2(s); \varepsilon), i = 1, 2, \dots, m$, a multiplier $\lambda_s \in \text{int } \mathbb{R}_+^m$, and proximal subgradients $\xi_{is} \in \partial_P \varphi^i(s, y_{is}), i = 1, 2, \dots, m$, such that

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

In the proof, it suffices to express the proximal normal cones to the corresponding sets $\{y \in \mathbb{R}^n \mid \varphi^i(s, y) \leq 0\}$ at $y = x_2(s), s \in \Omega$, on the basis of Ref. 18, Theorem 3.6, which is possible by virtue of the qualification condition (37). Then, one applies the proximal variant of the weak fuzzy sum rule (Ref. 18, Theorem 2.7) and arrives at the above result.

In the description of $N_{\Gamma(s)}^P(x_2(s))$, 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 (37) concerns the single functions $\varphi^i, i = 1, 2, \dots, m$, separately and not jointly as condition (35).

5. Appendix

Consider first an abstract mathematical program of the form

$$\min \varphi(x), \tag{38a}$$

$$\text{s.t. } x \in \Pi, \tag{38b}$$

where φ maps \mathbb{R}^p into $\bar{\mathbb{R}}$ and Π is a closed subset of \mathbb{R}^p . In Ref. 2 the 1st-order necessary optimality conditions for problem (38) have been proved in the following form.

Theorem 5.1. Let \hat{x} be a local solution of (38) and let φ be lower semicontinuous in a neighborhood of \hat{x} . Then, there exist an element $\hat{x}^* \in \mathbb{R}^p$ and a real $\lambda \geq 0$, not both equal to zero, such that

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

Under the additional condition

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

one has $\lambda \neq 0$ and

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

The generalized differential calculus of Mordukhovich is rather rich and enables one to compute generalized normal cones or their upper approximations to a large number of sets with different structure. The next statement can be proved easily on the basis of Ref. 2, Theorem 1.2, and Ref. 4, Corollary 5.5 and Theorem 6.10.

Theorem 5.2. 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 $F[\mathbb{R}^n \rightarrow \mathbb{R}^m]$ is continuous and C, D are closed subset of $\mathbb{R}^n, \mathbb{R}^m$, respectively. Let $\bar{x} \in A$ and, for $y^* \in N_D(F(\bar{x}))$, let

$$D^*F(\bar{x})(y^*) \cap (-N_C(\bar{x})) = \{0\}. \quad (39)$$

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

$$D^*Q(0, \bar{x})(x^*) \subset \{y^* \in N_D(F(\bar{x})) \mid 0 \in x^* + D^*F(\bar{x})(y^*) + N_C(\bar{x})\}. \quad (40)$$

Furthermore, under the condition

$$N_D(F(\bar{x})) \cap \text{Ker } D^*F(\bar{x}) = \{0\}, \quad (41)$$

the inclusion

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

holds true.

By combining (40) and Ref. 5, Theorem 3.2, we obtain the following criterion of pseudo-Lipschitz continuity of Q around $(0, \bar{x})$.

Theorem 5.3. Consider the map Q from Theorem 5.2 with F Lipschitz near \bar{x} and assume that the following constraint qualification is fulfilled:

$$0 \in D^*F(\bar{x})(\xi) + N_C(\bar{x}), \xi \in N_D(F(\bar{x})) \text{ imply } \xi = 0. \quad (42)$$

Then, conditions (39), (41) hold true and Q is pseudo-Lipschitz around $(0, \bar{x})$.

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

References

1. ROCKAFELLAR, R. T., and WETS, R., J. B., *Variational Analysis*, Springer Verlag, Berlin, Germany, 1998.
2. MORDUKHOVICH, B. S., *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, Russia, 1988 (in Russian); English Edition to appear in Wiley-Interscience.
3. MORDUKHOVICH, B. S., *Sensitivity Analysis in Nonsmooth Optimization*, Theoretical Aspects in Industrial Design, Edited by D. A. Field and V. Komkov, SIAM Publications, Proceedings in Applied Mathematics, Vol. 58, pp. 32–46, 1992.

4. MORDUKHOVICH, B. S., *Generalized Differential Calculus for Nonsmooth and Set-Valued Mappings*, Journal of Mathematical Analysis and Applications, Vol. 183, pp. 250–288, 1994.
5. MORDUKHOVICH, B. S., *Lipschitzian Stability of Constraint Systems and Generalized Equations*, Nonlinear Analysis; Theory, Methods and Applications, Vol. 22, pp. 173–206, 1994.
6. MORDUKHOVICH, B. S., and SHAO, Y., *Nonconvex Differential Calculus for Infinite-Dimensional Multifunctions*, Set-Valued Analysis, Vol. 4, pp. 205–236, 1996.
7. HIRIART-URRUTY, J. B., *Conditions Necessaires d'Optimalité pour un Programme Stochastique avec Recours*, SIAM Journal on Control and Optimization, Vol. 16, pp. 317–329, 1978.
8. VOGEL, S., *Necessary Optimality Conditions for Two-Stage Stochastic Programming Problems*, Optimization, Vol. 16, pp. 607–616, 1985.
9. CLARKE, F. H., *Optimization and Nonsmooth Analysis*, Wiley, New York, NY, 1983.
10. HIRIART-URRUTY, J. B., *Extensions of Lipschitz Integrands and Minimization of Nonconvex Integral Functionals: Applications to the Optimal Recourse Problem in Discrete Time*, Probability and Mathematical Statistics, Vol. 3, pp. 19–36, 1982.
11. CASTAING, C., and VALADIER, M., *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Mathematics, Springer Verlag, Berlin, Germany, Vol. 580, 1977.
12. AUBIN, J. P., and FRANKOWSKA, H., *Set-Valued Analysis*, Birkhäuser, Boston, Massachusetts, 1990.
13. ROCKAFELLAR, R. T., *Proximal Subgradients, Marginal Values, and Augmented Lagrangians in Nonconvex Optimization*, Mathematics of Operations Research, Vol. 6, pp. 424–436, 1981.
14. CLARKE, F. H., LEDYAEV, YU. S., STERN, R. J., and WOLENSKI, P. R., *Nonsmooth Analysis and Control Theory*, Graduate Texts in Mathematics, Springer Verlag, New York, NY, Vol. 178, 1998.
15. IOFFE, A. D., and ROCKAFELLAR, R. T., *The Euler and Weierstrass Conditions for Nonsmooth Variational Problems*, Calculus of Variations, Vol. 4, pp. 59–87, 1996.
16. YE, J. J., and YE, X. Y., *Necessary Optimality Conditions for Optimization Problems with Variational Inequality Constraints*, Mathematics of Operations Research, Vol. 22, pp. 977–997, 1997.
17. HIRIART-URRUTY, J. B., *Gradients Généralisés des Fonctions Marginales*, SIAM Journal on Control and Optimization, Vol. 16, pp. 301–316, 1978.
18. BORWEIN, J. M., and ZHU, Q. J., *A Survey of Subdifferential Calculus with Applications*, Nonlinear Analysis: Theory, Methods, and Applications, Vol. 38, pp. 687–773, 1999.
19. MORDUKHOVICH, B. S., and WANG, B., *Necessary Suboptimality and Optimality Conditions via Variational Principles*, SIAM Journal on Control and Optimization, Vol. 41, pp. 623–640, 2002.

20. BOUNKHEL, M., and THIBAUT, L., *On Various Notions of Regularity of Sets in Nonsmooth Analysis*, *Nonlinear Analysis: Theory, Methods, and Applications*, Vol. 48, pp. 223–246, 2002.
21. MORDUKHOVICH, B. S., and WANG, B., *Sequential Normal Compactness in Variational Analysis*, *Nonlinear Analysis: Theory, Methods, and Applications*, Vol. 47, pp. 717–728, 2001.
22. HIAI, F., and UMEGAKI, H., *Integrals, Conditional Expectations, and Martingales of Multivalued Functions*, *Journal of Multivariate Analysis*, Vol. 7, pp. 149–182, 1977.
23. OLSEN, P., *Multistage Stochastic Programming with Recourse as Mathematical Programming in an L_p Space*, *SIAM Journal on Control and Optimization*, Vol. 14, pp. 528–537, 1976.
24. ROCKAFELLAR, R. T., and WETS, R., J. B., *The Optimal Recourse Problem in Discrete Time: L^1 Multipliers for Inequality Constraints*, *SIAM Journal on Control and Optimization*, Vol. 16, pp. 16–36, 1978.