Mathematical Programming
https://doi.org/10.1007/s10107-020-01612-z

FULL LENGTH PAPER

Series B

Subregular recourse in nonlinear multistage stochastic optimization

Darinka Dentcheva[1] · Andrzej Ruszczyński[2]

Received: 19 January 2020 / Accepted: 25 December 2020

Abstract
We consider nonlinear multistage stochastic optimization problems in the spaces of integrable functions. We allow for nonlinear dynamics and general objective functionals, including dynamic risk measures. We study causal operators describing the dynamics of the system and derive the Clarke subdifferential for a penalty function involving such operators. Then we introduce the concept of subregular recourse in nonlinear multistage stochastic optimization and establish subregularity of the resulting systems in two formulations: with built-in nonanticipativity and with explicit nonanticipativity constraints. Finally, we derive optimality conditions for both formulations and study their relations.

Keywords Nonlinear Causal Operators · Subregularity · Nonanticipativity

Mathematics Subject Classification 49K27 · 90C15

1 Introduction

The concepts of metric regularity and subregularity of multifunctions are at the core of modern variational analysis, with applications to stability theory of systems of inclusions and derivation of optimality conditions. We refer the readers to the mono-

Funding was provided by the National Science Foundation, Division of Mathematical Sciences (Award No. 1907522).

✉ Darinka Dentcheva
darinka.dentcheva@stevens.edu

Andrzej Ruszczyński
rusz@rutgers.edu

[1] Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, NJ 07030, USA

[2] Department of Management Science and Information Systems, Rutgers University, Piscataway, NJ 08854, USA

Published online: 13 January 2021

graph [9] and to [14,19] for a comprehensive exposition of this vast field and its applications. However, most research on these properties and their implications in infinite-dimensional spaces focuses on fairly abstract settings in general metric or Banach spaces.

Our objective is to concentrate on specific questions arising in the analysis of nonlinear optimization problems in the spaces of p-integrable functions defined on a probability space $(\Omega, \mathscr{F}, P)$, with $p \in [1, \infty)$. The decision vector x is an element of the space $\mathscr{X} = \mathscr{L}_p(\Omega, \mathscr{F}, P; \mathbb{R}^n)$. The constraints of the problem are defined by a nonlinear operator $F : \mathscr{X} \to \mathscr{Y}$, where $\mathscr{Y} = \mathscr{L}_p(\Omega, \mathscr{F}, P; \mathbb{R}^m)$, multifunctions $X : \Omega \rightrightarrows \mathbb{R}^n$ and $Y : \Omega \rightrightarrows \mathbb{R}^m$, and a subspace $\mathscr{N} \subset \mathscr{X}$. The problem has the form

$$
\begin{aligned}
&\min \varphi(x) \\
&\ \text{s.t. } F(x) \in Y \quad \text{a.s.,} \\
&\qquad x \in X \quad \text{a.s.,} \\
&\qquad x \in \mathscr{N}.
\end{aligned}
$$

The objective function $\varphi : \mathscr{X} \to \mathbb{R}$ is a Lipschitz continuous functional.

In stochastic optimization and control, the constraint operator is causal, as defined in Sect. 3; it is given pointwise:

$$
F(x)(\omega) = f(x(\omega), \omega), \quad \omega \in \Omega,
$$

where $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^m$ describes the dynamics of the system.

The existing theory of stochastic multistage optimization provides optimality conditions for convex problems with linear dynamics and expected-value objective functionals involving convex integrands. Our goal is to analyze problems with nonlinear dynamics, which are prevalent in applications. Unfortunately, in the nonlinear case, the assumption of the Fréchet differentiability of $F(\cdot)$, common in the optimization theory in abstract Banach spaces, is unrealistic, except for very special cases (see Sect. 3). This poses a challenge in developing optimality conditions. Additionally, standard constraint qualification conditions are not satisfied.

Furthermore, we allow for general objective functionals $\varphi(\cdot)$ which include complex risk functionals that cannot be expressed as expected values of stage-wise costs. The classical approaches, exploiting the properties of convex integral functionals and conjugate duality, are inapplicable to such models.

Our contributions can be summarized as follows.

- Exact Clarke subdifferentials of penalty functions involving causal operators.
- A new concept of subregular recourse for multistage stochastic optimization problems with nonlinear dynamics in two settings: with built-in nonanticipativity and with explicit nonanticipativity constraints. It allows to deduce subregularity of the infinite-dimensional constraint system from the subregularity of the finite-dimensional systems associated with each stage and each elementary event.
- Optimality conditions for nonlinear multistage stochastic optimization problems with general objective functions in both settings.

The paper is organized as follows. In Sect. 2, we review several concepts and results on sets, tangent cones, and subregularity in spaces of integrable functions, which are essential for our analysis. In Sect. 3, we derive useful properties of causal operators describing the dynamics of the system. Finally, Sect. 4 is devoted to the analysis of multistage stochastic optimization problems with nonlinear causal operators and general objective functionals.

2 Preliminaries

For a given probability space $(\Omega, \mathscr{F}, P)$, the notation $\mathscr{X} = \mathscr{L}_p(\Omega, \mathscr{F}, P; \mathbb{R}^n)$ stands for the vector space of measurable functions $x : \Omega \to \mathbb{R}^n$, such that $\int \|x(\omega)\|^p \; P(d\omega) < \infty$, where $p \in [1, \infty)$. We denote the norm in $\mathscr{X}$ by $\|\cdot\|$; it will be clear from the context in which space the norm is taken. The distance function to a set A in a functional space will be denoted by $\operatorname{dist}(\cdot, A)$, while the distance to B in a finite dimensional space will be denoted by $d(\cdot, B)$.

We pair the space $\mathscr{X}$ with the space $\mathscr{X}^* = \mathscr{L}_q(\Omega, \mathscr{F}, P; \mathbb{R}^n)$, $1/p + 1/q = 1$, and with the bilinear form

$$\langle y, x\rangle = \int_\Omega y(\omega)^\top x(\omega) \; P(d\omega), \quad y \in \mathscr{X}^*, \quad x \in \mathscr{X}.$$

Here, $y(\omega)^\top$ refers to the transposed vector $y(\omega) \in \mathbb{R}^n$.

Definition 1 Suppose A is a closed subset of $\mathscr{X}$ and $x \in A$. The *contingent cone* to A at x is the set

$$\mathscr{T}_A(x) = \Big\{v \in \mathscr{X} : \liminf_{\tau \downarrow 0} \frac{1}{\tau} \operatorname{dist}(x + \tau v, A) = 0\Big\}.$$

Recall that for a cone $\mathscr{K} \subset \mathscr{X}$ its *polar cone* is defined as follows:

$$\mathscr{K}^\circ = \Big\{y \in \mathscr{X}^* : \langle y, x\rangle \le 0 \text{ for all } x \in \mathscr{K}\Big\}.$$

Definition 2 A set $A \subset \mathscr{X}$ is *derivable* at $x \in A$ if for every $v \in \mathscr{T}_A(x)$

$$\lim_{\tau \downarrow 0} \frac{1}{\tau} \operatorname{dist}_{\mathscr{X}}(x + \tau v, A) = 0.$$

We recall the notion of a decomposable set in $\mathscr{X}$ (cf. [2]).

Definition 3 A set $\mathscr{K} \subset \mathscr{X}$ is *decomposable* if a measurable multifunction $K : \Omega \rightrightarrows \mathbb{R}^n$ exists, such that $\mathscr{K} = \big\{x \in \mathscr{X} : x(\omega) \in K(\omega) \text{ a.s.}\big\}$.

The following fact is well-known in set-valued analysis (see, e.g., [2, Cor. 8.5.2]).

Lemma 1 *Suppose $A \subset \mathscr{X}$ is decomposable and $A(\omega)$ are closed and derivable sets for P-almost all $\omega \in \Omega$. Then*

$$\mathscr{T}_A(x) = \big\{ v \in \mathscr{X} : \textit{for } P\textit{-almost all } \omega,\ v(\omega) \in \mathscr{T}_{A(\omega)}\big(x(\omega)\big) \big\}.$$

Polar cones of convex decomposable cones are also decomposable. We provide a simple proof for convenience of the readers.

Lemma 2 *The polar cone $\mathscr{K}^\circ$ of a convex decomposable cone $\mathscr{K} \subset \mathscr{X}$ is a convex decomposable cone, and $K^\circ(\omega) = \big(K(\omega)\big)^\circ$ a.s.*

Proof Consider the convex decomposable cone $D : \Omega \rightrightarrows \mathbb{R}^n$ defined pointwise as follows: $D(\omega) = \big(K(\omega)\big)^\circ$. Evidently, if $y \in D$ then for all $x \in \mathscr{K}$ we have

$$\langle y, x \rangle = \int_\Omega y(\omega)^\top x(\omega)\ P(d\omega) \le 0.$$

Hence, $y \in \mathscr{K}^\circ$ and $D \subset \mathscr{K}^\circ$. We show that $\mathscr{K}^\circ = D$ by contradiction. Suppose an element $y \in \mathscr{K}^\circ$ exists, such that the event

$$S = \big\{ y(\omega) \notin \big(K(\omega)\big)^\circ \big\}$$

has positive probability. Then, for every $C > 0$ we can find a function $x \in \mathscr{X}$ such that $x(\omega) \in K(\omega)$ and $\langle y(\omega), x(\omega) \rangle > C$ for all $\omega \in S$. For $\omega \in \Omega \backslash S$ we select $x(\omega) \in K(\omega) \cap B_\delta$, where B_δ is a ball in $\mathbb{R}^n$ of radius $\delta > 0$. Then

$$\langle y, x \rangle = \int_S y(\omega)^\top x(\omega)\ P(d\omega) + \int_{\Omega \backslash S} y(\omega)^\top x(\omega)\ P(d\omega) \ge C P(S) - \delta \|y\|_{\mathscr{X}^*}.$$

The number C may be arbitrarily large, and δ may be arbitrarily small, which leads to a contradiction. This concludes the proof. □

Remark 1 When deriving optimality conditions, we shall use normal cones to convex sets defined as follows:

$$N_A(x) = \big[\mathscr{T}_A(x)\big]^\circ.$$

It follows from Lemmas 1 and 2 that if A is decomposable, then the normal cone is decomposable as well and consists of all measurable selectors $n(\omega) \in N_{A(\omega)}\big(x(\omega)\big)$.

We recall the subregularity concept regarding set-constrained systems [9, sec. 3H]; see also [8,17] and the references therein. For a multifunction $\mathfrak{H} : \mathscr{X} \rightrightarrows \mathscr{Y}$, where $\mathscr{Y}$ is a Banach space, we consider the relation

$$0 \in \mathfrak{H}(x). \tag{1}$$

Definition 4 The multifunction $\mathfrak{H}$ is *subregular* at $\hat{x} \in \mathscr{X}$ with $0 \in \mathfrak{H}(\hat{x})$, if $\delta > 0$ and $C > 0$ exist such that for all $x \in \mathscr{X}$ with $\|x - \hat{x}\|_{\mathscr{X}} \le \delta$ a point $\widetilde{x}$ satisfying (1) exists such that

$$\|\widetilde{x} - x\|_{\mathscr{X}} \le C \operatorname{dist}_{\mathscr{Y}}(0, \mathfrak{H}(x)).$$

In our analysis of multistage stochastic optimization problems, we shall use systems of the form

$$F(x) \in Y, \tag{2}$$

where $\mathscr{Y}$ is an $\mathscr{L}_p$-space, $F : \mathscr{X} \to \mathscr{Y}$ is Lipschitz continuous, and $Y \subset \mathscr{Y}$ is a closed convex set. With the multifunction $\mathfrak{H} = F(x) - Y$, the property of subregularity of (2) means that a constant C exists, such that for all x in a neighborhood of $\hat{x}$,

$$\operatorname{dist}\big(x, F^{-1}(Y)\big) \le C \operatorname{dist}\big(F(x), Y\big).$$

The subregularity of (2) at $\hat{x}$ is equivalent to the calmness of the multifunction

$$M(z) = \big\{x : F(x) \in Y - z\big\} = F^{-1}(Y - z)$$

at the point $(0, \hat{x})$, with the concept of calmness of a multifunction defined in [24] (under the name of the "upper Lipschiz property"); see also [15,16] and [9, Thm. 3H3].

3 Causal operators

We are interested in nonlinear operators acting between two spaces of sequences of integrable functions. For a probability space $(\Omega, \mathscr{F}, P)$ with filtration $\{\emptyset, \Omega\} = \mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}_T = \mathscr{F}$, we define the spaces $\mathscr{X}_t = \mathscr{L}_p(\Omega, \mathscr{F}_t, P; \mathbb{R}^n)$ and $\mathscr{Y}_t = \mathscr{L}_p(\Omega, \mathscr{F}_t, P; \mathbb{R}^m)$ with $p \in [1, \infty)$, $t = 1, \dots, T$. Let $\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_T$ and $\mathscr{Y} = \mathscr{Y}_1 \times \cdots \times \mathscr{Y}_T$. We use $x_{1:t}$ as the shorthand notation for $(x_1, \dots, x_t)$, and $\mathscr{X}_{1:t}$ for $\mathscr{X}_1 \times \cdots \times \mathscr{X}_t$.

We adapt the following concept from the dynamical system theory (see [7] and the references therein).

Definition 5 An operator $F : \mathscr{X} \to \mathscr{Y}$ is *causal*, if functions $f_t : \mathbb{R}^{nt} \times \Omega \to \mathbb{R}^m$ exist, such that for all $t = 1, \dots, T$

$$F_t(x)(\omega) = f_t(x_{1:t}(\omega), \omega), \quad \omega \in \Omega, \tag{3}$$

and each $f_t(\cdot, \cdot)$ is superpositionally measurable.

Superpositional measurability is discussed in detail in [1]; this property is guaranteed for Carathéodory functions, in particular, for functions that satisfy the assumption below (*op. cit.*, Thm. 1.1).

We use the notation

$$f(x(\omega),\omega) = \big\{ f_t(x_{1:t}(\omega),\omega)\big\}_{t=1,\dots,T}.$$

Then $F(x)(\omega) = f(x(\omega),\omega)$.

Assumption 1 For all $t = 1,\dots,T$:

(i) $f_t(\xi,\cdot)$ is an element of $\mathscr{Y}_t$ for all $\xi \in \mathbb{R}^{nt}$;
(ii) For almost all $\omega \in \Omega$, $f_t(\cdot,\omega)$ is continuously differentiable, with the Jacobian $f_t'(\cdot,\omega)$;
(iii) A constant C_f exists, such that $\| f_t'(\cdot,\omega)\| \le C_f$, almost surely.

Under Assumption 1, each F_t given by (3) indeed maps the product space $\mathscr{X}_{1:t}$ into a subset of $\mathscr{Y}_t$. In fact, condition (iii) is related to the necessary condition for the Lipschitz continuity of $F(\cdot)$; see [1, Thm. 3.10].

Notice that each Jacobian $f_t'(x_{1:t}(\omega),\omega)$ acts on the realization of the subvector $h_{1:t}(\omega)$ of an element $h \in \mathscr{X}$. For simplicity, we use the same notation as if it were acting on the entire $h(\omega)$. Then we can write

$$f'(x(\omega),\omega) = \big\{ f_t'(x_{1:t}(\omega),\omega)\big\}_{t=1,\dots,T}$$

to represent the Jacobian of $[F(x)](\omega)$ with respect to $x(\omega)$.

Lemma 3 *A causal operator $F(\cdot)$ satisfying Assumption 1 is Gâteaux differentiable with the derivative $F'(x)$ defined by*

$$[F'(x)\,h](\omega) = f'(x(\omega),\omega)\,h(\omega), \quad \omega \in \Omega. \tag{4}$$

Proof We define $J(x) : \mathscr{X} \to \mathscr{Y}$ by using the right hand side of formula (4):

$$[J(x)\,h](\omega) = f'(x(\omega),\omega)\,h(\omega), \quad \omega \in \Omega.$$

Notice that $J(\cdot)$ is a continuous linear operator.

We calculate the directional derivative of the function F at x in the direction h. First, we observe that for any $h \in \mathscr{X}$ and $\tau > 0$

$$\frac{1}{\tau}\big\| f(x(\omega)+\tau h(\omega),\omega) - f(x(\omega),\omega) - \tau f'(x(\omega),\omega)h(\omega)\| \le 2C_f \|h(\omega)\| \quad \text{a.s.}$$

and the function at the right-hand side is p-integrable. This yields the following estimate:

$$\begin{aligned}
&\frac{1}{\tau}\big\| F(x+\tau h) - F(x) - \tau J(x)\,h\big\|_{\mathscr{Y}} \\
&\quad = \bigg(\int \Big\| \frac{1}{\tau}\big(f(x(\omega)+\tau h(\omega),\omega) - f(x(\omega),\omega) - \tau f'(x(\omega),\omega)h(\omega)\big)\Big\|^p \; P(d\omega)\bigg)^{1/p} \\
&\quad \le 2C_f \bigg(\int \|h(\omega)\|^p \; P(d\omega)\bigg)^{1/p} = 2C_f \|h\|_{\mathscr{Y}}.
\end{aligned}$$

Using Lebesgue's dominated convergence theorem, we obtain

$$\begin{aligned}
&\lim_{\tau\downarrow 0}\frac{1}{\tau}\big\| F(x+\tau h)-F(x)-\tau J(x)\,h\big\|_{\mathscr{Y}}\\
&=\bigg(\int \lim_{\tau\to 0}\Big\|\frac{1}{\tau}\big(f(x(\omega)+\tau h(\omega),\omega)-f(x(\omega),\omega)-\tau f'(x(\omega),\omega)h(\omega)\big)\Big\|^{p}\ P(d\omega)\bigg)^{1/p}\\
&=0.
\end{aligned}$$

Therefore, $J(x)$ is the Gâteaux derivative of $F(\cdot)$ at x. □

It is worth mentioning that our assumptions do not guarantee the Fréchet differentiability of $F(\cdot)$. Unfortunately, in a nonlinear stochastic setting, the Fréchet differentiability of $F(\cdot)$ is very difficult to guarantee, except for very special cases [1, Sec. 2.7 and Thm. 3.12]. As an illustration, we provide the following example.

Example 1 Let $\Omega=[0,1]$ and P be the Lebesgue measure on $[0,1]$. We define the spaces $\mathscr{X}=\mathscr{Y}=\mathscr{L}_1(\Omega,\mathscr{F},P)$, and the operator $F:\mathscr{X}\to\mathscr{Y}$ given by

$$F(x)(\omega)=f(x(\omega),\omega)=\begin{cases}\big(x(\omega)\big)^2 & \text{if } -1\le x(\omega)\le 1,\\ 2\big|x(\omega)\big|-1 & \text{otherwise.}\end{cases}$$

Note that $\|F(x)\|\le 2\|x\|$, and thus indeed $F:\mathscr{X}\to\mathscr{Y}$. Unfortunately, $F(\cdot)$ is not Fréchet differentiable at 0. By Lemma 3, the Gâteaux derivative of F at 0 is 0, and thus the Fréchet derivative, if it existed, would be $F'(0)=0$ as well. Consider the sequence of functions

$$x_n(\omega)=\begin{cases}1 & \text{if } 0\le\omega\le\frac{1}{n},\\ 0 & \text{otherwise,}\end{cases}\qquad n=1,2,\ldots$$

We have $\|x_n\|=\frac{1}{n}$ and thus $x_n\to 0$. By construction, $F(x_n)=x_n$, $F(0)=0$, and then, by the definition of the Fréchet derivative, we would have

$$0=\lim_{n\to\infty}\frac{F(x_n)-F(0)-F'(0)x_n}{\|x_n\|}=\lim_{n\to\infty}\frac{x_n}{\|x_n\|}.$$

This is a contradiction, because all elements on the right hand side have norm 1. □

In the next result, we calculate the Clarke subdifferential of the function

$$\Phi(\cdot)=\operatorname{dist}\big(F(\cdot),Y\big)\tag{5}$$

with $F(\cdot)$ being only Gâteaux differentiable.

Theorem 1 *Suppose $F(\cdot)$ is a causal operator satisfying Assumption 1, $Y\subset\mathscr{Y}$ is convex and closed, and $F(x)\in Y$. Then*

$$\partial\Phi(x)=\big[F'(x)\big]^*\big(N_Y(F(x))\cap\mathbb{B}_{\mathscr{Y}^*}\big),$$

where $\big[F'(x)\big]^*$ *is the adjoint operator to the Gâteaux derivative* $F'(x)$*, and* $\mathbb{B}_{\mathscr{Y}^*}$ *is the closed unit ball in* $\mathscr{Y}^*$.

Proof Since Y is convex, the function $\operatorname{dist}(\cdot, Y)$ is convex as well, and we can use the subgradient inequality:

$$\operatorname{dist}\big(F(z+\tau h), Y\big) - \operatorname{dist}\big(F(z), Y\big) \le \big\langle g, F(z+\tau h) - F(z)\big\rangle,$$

for any $g \in \partial \operatorname{dist}(y, Y)$ at $y = F(z+\tau h)$. The Clarke directional derivative of $\Phi(\cdot)$ at x in the direction h can thus be bounded from above as follows:

$$\begin{aligned}\Phi^0(x; h) &= \limsup_{\substack{z\to x\\ \tau\downarrow 0}} \frac{1}{\tau}\Big(\operatorname{dist}\big(F(z+\tau h), Y\big) - \operatorname{dist}\big(F(z), Y\big)\Big)\\ &\le \limsup_{\substack{z\to x\\ \tau\downarrow 0}} \Big\langle g, \frac{1}{\tau}\big(F(z+\tau h) - F(z)\big)\Big\rangle, \qquad (6)\end{aligned}$$

for any $g \in \partial\operatorname{dist}(F(z+\tau h), Y)$. Consider arbitrary sequences $\{z_k\} \to x$ and $\{\tau_k\} \downarrow 0$, as $k \to \infty$. By the mean value theorem, for each $\omega \in \Omega$, each component of the quotient on the right hand side of (6) can be expressed as follows:

$$\begin{aligned}&\frac{1}{\tau_k}\big[f_j(z_k(\omega) + \tau_k h(\omega), \omega) - f_j(z_k(\omega), \omega))\big]\\ &\quad = f_j'(\bar{z}_{k,j}(\omega), \omega)\, h(\omega), \quad j = 1, \dots, mT,\end{aligned}$$

where $\bar{z}_{k,j}(\omega) = z_k(\omega) + \tau_k\theta_{k,j}(\omega)h(\omega)$ with $\theta_{k,j}(\omega) \in [0, 1]$. Then

$$\frac{1}{\tau_k}\big[f(z_k(\omega) + \tau_k h(\omega), \omega) - f(z_k(\omega), \omega))\big] = [F'(x)\,h](\omega) + \Delta_k(\omega), \qquad (7)$$

with the error $\Delta_k(\omega)$ having coordinates

$$\Delta_{k,j}(\omega) = \big[f_j'(\bar{z}_{k,j}(\omega), \omega) - f_j'(x(\omega), \omega)\big]h(\omega), \quad j = 1, \dots, mT.$$

We shall verify that $\{\Delta_k\} \to 0$ in $\mathscr{Y}$, as $k \to \infty$. For an arbitrary $\varepsilon > 0$, we define the events

$$\Omega_{k,\epsilon} = \Big\{\omega \in \Omega : \max_{1\le j\le mT} \|\bar{z}_{k,j}(\omega) - x(\omega)\| > \varepsilon\Big\}.$$

Since $\{\bar{z}_{k,j}\} \to x$ in $\mathscr{X}$, as $k \to \infty$, the convergence in probability follows:

$$\lim_{k\to\infty} P\big[\Omega_{k,\epsilon}\big] = 0. \qquad (8)$$

Let

$$\delta(\varepsilon, \omega) = \sup_{\|w - x(\omega)\|\le\varepsilon} \; \max_{1\le j\le mT} \big\| f_j'(w, \omega) - f_j'(x(\omega), \omega)\big\|.$$

By the boundedness and continuity of the derivatives, $\delta(\varepsilon,\omega) \le 2C_f$, and $\delta(\varepsilon,\omega) \to 0$ a.s., when $\varepsilon \downarrow 0$. The error from our desired representation of the differential quotient can be bounded as follows:

$$\|\Delta_k(\omega)\| \le 2C_f \mathbb{1}_{\Omega_{k,\varepsilon}}(\omega)\|h(\omega)\| + \delta(\varepsilon,\omega)\mathbb{1}_{\Omega^c_{k,\varepsilon}}(\omega)\|h(\omega)\|. \tag{9}$$

Consider the first term on the right hand side of (9). Suppose that with some $\alpha > 0$,

$$\int \mathbb{1}_{\Omega_{k,\varepsilon}}\|h(\omega)\|^p \; P(d\omega) > \alpha, \quad \text{for} \quad k \in \mathscr{K} \subset \mathscr{N}, \tag{10}$$

where the set of indices $\mathscr{K}$ is infinite. By the Banach–Alaoglu theorem [3, Ch.VII,§7], the sequence $\left\{\mathbb{1}_{\Omega_{k,\varepsilon}}\right\}_{k\in\mathscr{K}}$ of elements in the unit ball of $\mathscr{L}_\infty(\Omega,\mathscr{F},P)$ must have a weakly* convergent subsequence, indexed by $k \in \mathscr{K}_1 \subset \mathscr{K}$. By (8), its weak* limit is zero. Consequently,

$$\lim_{\substack{k\to\infty\\ k\in\mathscr{K}_1}} \int \mathbb{1}_{\Omega_{k,\varepsilon}}\|h(\omega)\|^p \; P(d\omega) = 0,$$

which contradicts (10). Therefore, for any $\alpha > 0$, the inequality (10) may be satisfied only finitely many times, and thus $\mathbb{1}_{\Omega_{k,\varepsilon}} h \to 0$ in $\mathscr{Y}$.

Combining this with (9), we obtain (in the space $\mathscr{Y}$)

$$\limsup_{k\to\infty} \left\|\Delta_k\right\| \le \left(\int \big(\delta(\varepsilon,\omega)\|h(\omega)\|\big)^p \; P(d\omega)\right)^{1/p}.$$

Letting $\varepsilon \downarrow 0$ and using the Lebesgue dominated convergence theorem, we conclude that $\{\Delta_k\} \to 0$ in $\mathscr{Y}$.

For arbitrary $g_k \in \partial \mathrm{dist}(F(z_k + \tau_k h), Y)$, in view of (7),

$$\Phi^0(x;h) \le \limsup_{\substack{z_k\to x\\ \tau_k\downarrow 0}} \left\langle g_k, \frac{1}{\tau_k}\big(F(z_k+\tau_k h) - F(z_k)\big)\right\rangle \le \limsup_{\substack{z_k\to x\\ \tau_k\downarrow 0}} \left\langle g_k, F'(x)h + \Delta_k\right\rangle.$$

All subgradients g_k are bounded by the Lipschitz constant 1 of the distance function. Therefore, $\langle g_k, \Delta_k\rangle \to 0$. Consider an arbitrary accumulation point α of the sequence $\left\langle g_k, F'(x)\,h\right\rangle$. By the Banach–Alaoglu theorem, we can choose a sub-subsequence $\{g_k\}_{k\in\mathscr{K}}$ which is weakly* convergent to some g in $\mathscr{Y}^*$. Then $\alpha = \left\langle g, F'(x)\,h\right\rangle$. By the norm-to-weak* upper semicontinuity of the subdifferential [23, Prop. 2.5], $g \in \partial \mathrm{dist}(F(x), Y)$. Therefore,

$$\Phi^0(x;h) \le \max_{g\in\partial \mathrm{dist}(F(x),Y)} \left\langle g, F'(x)\,h\right\rangle. \tag{11}$$

The converse inequality follows from (6) by setting $z = x$ and using Lemma 3:

$$\begin{aligned}\Phi^0(x;h) &\geq \limsup_{\tau\downarrow 0} \frac{1}{\tau}\Big(\operatorname{dist}\big(F(x+\tau h),Y\big) - \operatorname{dist}\big(F(x),Y\big)\Big)\\ &\geq \limsup_{\tau\downarrow 0} \frac{1}{\tau}\langle g, F(x+\tau h) - F(x)\rangle = \big\langle g, F'(x)\,h\big\rangle,\end{aligned}$$

for any $g \in \partial\operatorname{dist}(F(x),Y)$. Therefore,

$$\Phi^0(x;h) \geq \max_{g\in\partial\operatorname{dist}(F(x),Y)} \big\langle g, F'(x)\,h\big\rangle.$$

Combining this with (11), we infer that

$$\Phi^0(x;h) = \max_{g\in\partial\operatorname{dist}(F(x),Y)} \big\langle [F'(x)]^* g, h\big\rangle.$$

Since $\Phi^0(x;h)$ is the support function of $\partial\Phi(x)$ (*cf.* [6, Proposition 2.1.2]) and the support function provides a unique description of a weakly* closed and convex set, we conclude that

$$\partial\Phi(x) = \big\{[F'(x)]^* g : g \in \partial\operatorname{dist}(F(x),Y)\big\}.$$

Having in mind that $\partial\operatorname{dist}(y,Y) = N_Y(y) \cap \mathbb{B}_{\mathscr{Y}^*}$ whenever $y \in Y$, we obtain the stated result. $\square$

Corollary 1 *The same argument provides the Clarke subdifferential of a composition of a convex subdifferentiable functional $\rho : \mathscr{Y} \to \mathbb{R}$ and a causal $F(\cdot)$ satisfying Assumption 1: $\partial(\rho\circ F)(x) = [F'(x)]^*\,\partial\rho(F(x))$.*

4 Multistage stochastic optimization and nonanticipativity

We study nonlinear multistage stochastic optimization with general objective functionals which include dynamic measures if risk. The multistage problems can be formulated in two different ways regarding the way implementability of the solution is reflected in the model. One possibility is to formulate the model in such a way that the definition of the decision spaces includes the $\mathscr{F}_t$-measurability of the decisions at time t, $t = 1,\dots,T$. In another formulation, we consider decision spaces of $\mathscr{F}$-measurable decisions at each stage, but we add additional linear constraints enforcing $\mathscr{F}_t$-measurability.

4.1 The model with built-in nonanticipaticity

A probability space $(\Omega, \mathscr{F}, P)$ with filtration $\{\emptyset, \Omega\} = \mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}_T = \mathscr{F}$ is given. At each stage $t = 1,\dots,T$, a decision x_t with values in $\mathbb{R}^n$ is made. We require that x_t be an element of the space $\mathscr{X}_t = \mathscr{L}_p(\Omega, \mathscr{F}_t, P; \mathbb{R}^n)$ with $p \in [1, \infty)$. We define the space $\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_T$. We denote the spaces in which our dynamics operators will take values by $\mathscr{Y}_t = \mathscr{L}_p(\Omega, \mathscr{F}_t, P; \mathbb{R}^m)$, $t = 1, \ldots, T$.

The dynamics of the system is represented by the relation

$$F(x) \in Y, \tag{12}$$

where $F : \mathscr{X} \to \mathscr{Y}$ is a causal operator, and $Y = Y_1 \times \cdots \times Y_T$, with each $Y_t : \Omega \rightrightarrows \mathbb{R}^m$, $t = 1, \ldots, T$, being an $\mathscr{F}_t$-measurable multifunction with convex and closed values. In a more explicit way, the relation (12) has the form:

$$F_t(x_{1:t}) \in Y_t, \quad t = 1, \ldots, T, \tag{13}$$

and, due to the causality of $F(\cdot)$ and the decomposability of Y,

$$f_t(x_{1:t}(\omega), \omega) \in Y_t(\omega), \quad t = 1, \ldots, T, \quad \omega \in \Omega.$$

Additionally, $\mathscr{F}_t$-measurable mulitifunctions with closed convex images $X_t : \Omega \rightrightarrows \mathbb{R}^n$, $t = 1, \ldots, T$, are defined.

The objective function is a Lipschitz continuous functional $\varphi : \mathscr{X} \to \mathbb{R}$. The multistage stochastic optimization problem is formulated as follows:

$$\min \varphi(x_{1:T}) \tag{14}$$
$$\text{s.t.} F_t(x_{1:t}) \in Y_t \quad \text{a.s.}, \quad t = 1, \ldots, T, \tag{15}$$
$$x_t \in X_t \quad \text{a.s.}, \quad t = 1, \ldots, T. \tag{16}$$

Evidently, we could have aggregated the relations (15) and (16) into one inclusion, but it is convenient to distinguish between the causal relations describing the dynamics of the system, and the stage-wise constraints.

The existing theory of stochastic optimization provides optimality conditions for convex versions of problem (14)–(16), with linear operators $F_t(\cdot)$ and expected value functionals

$$\varphi(x_1, \ldots, x_T) = \mathbb{E}\left[\sum_{t=1}^{T} c_t(x_t(\omega), \omega)\right], \tag{17}$$

involving convex integrands $c_t(\cdot, \cdot)$, see [10,11,22,27–29,31]. Here, $\mathbb{E}[\cdot]$ is the expected value operator. A formulation with nonlinear constraints is analyzed in [22], with the use of Mordukhovich calculus and fuzzy proximal subgradients. It is unclear, though, how these fairly abstract objects can be calculated and applied.

We expand the theory by allowing non-linear dynamics and more general functionals in the model description. Consider stage-wise random cost operators $C_t : \mathscr{X}_t \to \mathscr{Z}_t$, given by $C_t(x_t)(\omega) = c_t(x_t(\omega), \omega)$, where $\mathscr{Z}_t = \mathscr{L}_p(\Omega, \mathscr{F}_t, P)$. A fairy general class of objective functionals results from replacing the expected value operator with a dynamic measure of risk $\varrho : \mathscr{Z}_1 \times \cdots \times \mathscr{Z}_T \to \mathbb{R}$:

$$\varphi(x_{1:T}) = \varrho\big(C_1(x_1), \ldots, C_T(x_T)\big). \tag{18}$$

We refer the readers to [32, Ch. 6] and the references therein for the theory of dynamic risk measures and their use in multistage stochastic optimization. An important form of $\varrho(\cdot)$, resulting from time-consistency and other technical assumptions, is the following

$$\begin{aligned}&\varrho\big(C_1(x_1), \ldots, C_T(x_T)\big)\\ &\quad = C_1(x_1) + \rho_1\Big(C_2(x_2) + \rho_2\big(C_3(x_3) + \cdots + \rho_{T-1}(C_T(x_T))\cdots\big)\Big).\end{aligned} \tag{19}$$

In this formula, each $\rho_t : \mathscr{Z}_{t+1} \to \mathscr{Z}_t, t = 1, \ldots, T-1$, is a one-step conditional risk measure. It generalizes the conditional expected value operator $\mathbb{E}_t[\,\cdot\,] = \mathbb{E}[\,\cdot\,|\mathscr{F}_t]$. With the special selection of $\rho_t(\cdot) = \mathbb{E}_t[\,\cdot\,]$ the formula (18) reduces to (17). Other popular choices of $\rho_t(\cdot)$ lead to objective functionals $\varphi(\cdot)$ which cannot be expressed as expected values and are highly nonlinear in decisions and the underlying probability measure. One such example is the mean–semideviation conditional mapping:

$$\rho_t(Z_{t+1}) = \mathbb{E}_t\big[Z_{t+1}\big] + \varkappa\bigg(\mathbb{E}_t\Big[\Big(\max\big(0, Z_{t+1} - \mathbb{E}_t\big[Z_{t+1}\big]\big)\Big)^p\Big]\bigg)^{1/p}, \quad \varkappa \in [0, 1].$$

With this, and many other choices of the conditional risk mappings, the operator $\varrho(Z_1, \ldots, Z_T)$ is convex and Lipschitz continuous. Therefore, the composition (18) is Lipschitz continuous, as long as the cost operators $C_t(\cdot)$ are Lipschitz continuous. In our further considerations, we will only use the Lipschiz continuity of $\varphi(\cdot)$, without specificity resulting from the structure mentioned above.

Our key idea is to use uniform parametric subregularity of deterministic finite-dimensional set-constrained systems associated with each stage $t = 1, \ldots, T$ and each elementary event $\omega \in \Omega$:

$$f_t(\zeta_{1:t-1}, \xi, \omega) \in Y_t(\omega), \tag{20}$$
$$\xi \in X_t(\omega). \tag{21}$$

Here, $\zeta_{1:t-1} \in \mathbb{R}^{n(t-1)}$ represents the history of decisions at the particular elementary event, and the elementary event $\omega \in \Omega$ itself are parameters of the system. For uniformity of notation, for $t = 1$ the parameter $\zeta_{1:t-1}$ is non-existent.

We introduce the following concept.

Definition 6 The system (20)–(21) *admits complete subregular recourse*, if a constant C exist, such that for almost all $\omega \in \Omega$, every $\zeta_{1:t-1} \in X_{1:t-1}(\omega)$ and every $\eta \in \mathbb{R}^n$, a solution ξ of (20)–(21) exists, satisfying the inequality

$$\|\xi - \eta\| \le C\big(d(f_t(\zeta_{1:t-1}, \eta, \omega), Y_t(\omega)) + d(\eta, X_t(\omega))\big).$$

In two-stage stochastic linear programming the concept of *relatively complete recourse* is well-established (see [32, sec. 2.1.3] and the references therein). It means the solvability of the system (20)–(21) for almost all $\omega \in \Omega$ and every $\zeta_1 \in X_1(\omega)$. But even in this case, the uniform subregularity of this system is not guaranteed.

We shall prove subregularity of the entire infinite-dimensional system of constraints (15)–(16) when complete subregular recourse is admitted.

Theorem 2 *If the system* (20)–(21) *admits complete subregular recourse, then the system* (15)–(16) *is subregular at any feasible point* $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_T)$.

Proof Let $u = (u_1, \ldots, u_T) \in \mathscr{X}$ be chosen from a sufficiently small neighborhood of $\hat{x}$. We shall construct a solution $\bar{x}$ of (15)–(16) which is close to u, with an appropriate error bound.

For $t = 1, \ldots, T$ we consider the system in the space $\mathscr{X}_t$:

$$F_t(\bar{x}_{1:t-1}, x_t) \in Y_t,$$
$$x_t \in X_t.$$

Our intention is to find a solution $\bar{x}_t$ to this system, which is sufficiently close to u_t. By Lipschitz continuity of $F_t(\cdot, \cdot)$,

$$\big\| F_t\big(\bar{x}_{1:t-1}, u_t\big)\big\| \leq \big\| F_t\big(u_{1:t}\big)\big\| + L\big\|\bar{x}_{1:t-1} - u_{1:t-1}\big\|. \tag{22}$$

We define a multifunction $\mathfrak{G} : \Omega \rightrightarrows \mathbb{R}^n$ by the relations

$$\mathfrak{G}(\omega) = \Big\{\xi \in \mathbb{R}^n : \ f_t(\bar{x}_{1:t-1}(\omega), \xi, \omega) \in Y_t(\omega), \ \xi \in X_t(\omega),$$
$$\big\|\xi - u_t(\omega)\big\| \leq C\Big(d\big(f_t(\bar{x}_{1:t-1}(\omega), u_t(\omega), \omega\big), Y_t(\omega)\big) + d\big(u_t(\omega), X_t(\omega)\big)\Big)\Big\}.$$

We observe that both distance functions in the definition of $\mathfrak{G}(\cdot)$ are $\mathscr{F}_t$-measurable by [2, Corollary 8.2.13]. Therefore, the multifunction $\mathfrak{G}$ is $\mathscr{F}_t$-measurable. It has nonempty images due to Definition 6 applied with $\eta = u_t(\omega)$ and $\zeta_{1:t-1} = \bar{x}_{1:t-1}(\omega)$. Hence, a measurable selection $\bar{x}_t$ of $\mathfrak{G}$ exists (cf. [18]). From the construction of the multifunction $\mathfrak{G}$,

$$\big\|\bar{x}_t(\omega) - u_t(\omega)\big\| \leq C\Big(d\big(f_t(\bar{x}_{1:t-1}(\omega), u_t(\omega), \omega\big), Y_t(\omega)\big) + d\big(u_t(\omega), X_t(\omega)\big)\Big).$$

Therefore, with the norms and distances in the spaces $\mathscr{X}_t$ and $\mathscr{Y}_t$,

$$\big\|\bar{x}_t - u_t\big\| \leq C\Big(\operatorname{dist}\big(F_t(\bar{x}_{1:t-1}, u_t\big), Y_t\big) + \operatorname{dist}\big(u_t, X_t\big)\Big). \tag{23}$$

Combining inequalities (23) and (22), we infer that

$$\big\|\bar{x}_t - u_t\big\| \leq \ C\Big(\operatorname{dist}\big(F_t(u_{1:t}), Y_t\big) + L\big(\big\|\bar{x}_{1:t-1} - u_{1:t-1}\big\|\big) \ + \operatorname{dist}\big(u_t, X_t\big)\Big). \tag{24}$$

We can now prove by induction that constants $\bar{C}_t$ exist such that

$$\|\bar{x}_t - u_t\| \leq \bar{C}_t \sum_{\ell=1}^{t} \Big(\operatorname{dist}\big(F_\ell\big(u_{1:\ell}\big), Y_\ell\big) + \operatorname{dist}\big(u_\ell, X_\ell\big)\Big).$$

For $t = 1$, the result is provided by (24), because the term $\left\|\bar{x}_{1:t-1} - u_{1:t-1}\right\|$ is not present there. Supposing it is true for $t-1$, we verify it for t using (24). The last relation for $t = T$ establishes the subregularity of the system (15)–(16). □

Under Assumption 1, we denote:

$$F_t'(\hat{x}_{1:t}) = A_t = \big(A_{t,1}, \dots, A_{t,t}\big), \quad t = 1, \dots, T,$$

with partial Jacobians $A_{t,\ell} : \mathscr{X}_\ell \to \mathscr{Y}_t$,

$$A_{t,\ell} = \frac{\partial F_t(\hat{x}_{1:t})}{\partial x_\ell}, \quad \ell = 1, \dots, t, \quad t = 1, \dots, T. \tag{25}$$

These linear operators are defined pointwise:

$$A_{t,\ell}(\omega) = \frac{\partial f_t(\hat{x}_{1:t}(\omega), \omega)}{\partial x_\ell(\omega)}, \quad \ell = 1, \dots, t, \quad t = 1, \dots, T, \quad \omega \in \Omega. \tag{26}$$

Due to Assumption 1, all operators $A_{t,\ell}$ are continuous linear operators.

Now, we establish necessary conditions of optimality for problem (14)–(16).

Theorem 3 *Suppose the system (20)–(21) admits complete subregular recourse and the policy $\hat{x}$ is a local minimum of problem (14)–(16). Then a subgradient $\hat{g} \in \partial\varphi(\hat{x})$, multipliers $\hat{\psi}_t \in N_{Y_t}(F_t(\hat{x}_{1:t}))$, $t = 1, \dots, T$, and normal elements $\hat{n}_t \in N_{X_t}(\hat{x}_t)$, $t = 1, \dots, T$, exist, such that for P-almost all $\omega \in \Omega$ we have:*

$$\hat{g}_t + A_{t,t}^\top \hat{\psi}_t + \mathbb{E}_t\Bigg[\sum_{\ell=t+1}^{T} A_{\ell,t}^\top \hat{\psi}_\ell\Bigg] + \hat{n}_t = 0, \quad t = 1, \dots, T. \tag{27}$$

Proof Since $\varphi(\cdot)$ is Lipschitz continuous about $\hat{x}$ with some constant L_φ, then for every $K > L_\varphi$ the point $\hat{x}$ is a local minimum of the function

$$\varphi(x) + K \operatorname{dist}(x, X \cap F^{-1}(Y));$$

see [6, Prop. 2.4.3]. The system (15)–(16) is subregular with some constant $\bar{C}$ by virtue of Theorem 2. Consequently, $\hat{x}$ is a local minimum of the function

$$\varphi(x) + K\bar{C}\big(\operatorname{dist}(F(x), Y) + \operatorname{dist}(x, X)\big).$$

This type of argument is discussed in detail in [4,13,15]. We use Clarke's necessary conditions of optimality for Lipschitz continuous functions:

$$0 \in \partial\varphi(\hat{x}) + K\bar{C}\,\partial\big[\operatorname{dist}(F(\cdot), Y)\big](\hat{x}) + K\bar{C}\,\partial\big[\operatorname{dist}(\cdot, X)\big](\hat{x}).$$

The Clarke subdifferential of the function $\operatorname{dist}(F(\cdot), Y)$ is calculated in Theorem 1:

$$\partial\Phi(\hat{x}) = \big[F'(\hat{x})\big]^* \left(N_Y(F(\hat{x})) \cap \mathbb{B}_{\mathscr{Y}^*}\right).$$

The subdifferential of $\operatorname{dist}(\hat{x}, X)$ is $N_X(\hat{x}) \cap \mathbb{B}_{\mathscr{X}^*}$. We infer that a subgradient $\hat{g} \in \partial\varphi(\hat{x})$, an element $\hat{\psi} \in N_Y(F(\hat{x}))$, and a normal vector $\hat{n} \in N_X(\hat{x})$ exist, such that

$$\hat{g} + \big[F'(\hat{x})\big]^* \hat{\psi} + \hat{n} = 0.$$

We can derive a more explicit form of the vector $\big[F'(\hat{x})\big]^* \hat{\psi}$. Due to the decomposability of X_t the normal cone $N_{X_t}(x)$ is composed of elements which are selectors of $N_{X_t(\cdot)}\big(x(\cdot)\big)$ (*cf* Remark 1); we have $N_{X_t}(\hat{x}_t)(\omega) = N_{X_t(\omega)}\big(\hat{x}_t(\omega)\big)$ a.s.. Using the same argument and the causality of F_t, we obtain

$$\hat{\psi}_t(\omega) \in N_{Y_t(\omega)}\big(f_t(\hat{x}_{1:t}(\omega), \omega)\big) \quad t = 1, \ldots, T, \quad \text{for almost all } \omega \in \Omega.$$

Now, using the block-triangular form of $A = F'(\hat{x})$, for any $h \in \mathscr{X}$ we can write

$$\begin{aligned} \langle A^* \hat{\psi}, h \rangle = \langle \hat{\psi}, Ah \rangle &= \sum_{t=1}^{T} \langle \hat{\psi}_t, A_t h \rangle \\ &= \sum_{t=1}^{T} \sum_{\ell=1}^{t} \langle \hat{\psi}_t, A_{t,\ell} h_\ell \rangle = \sum_{\ell=1}^{T} \sum_{t=\ell}^{T} \langle A_{t,\ell}^* \hat{\psi}_t, h_\ell \rangle. \end{aligned} \tag{28}$$

It follows that $A_{t,\ell}^* \hat{\psi}_t = \mathbb{E}\big[A_{t,\ell}^\top \hat{\psi}_t \,\big|\, \mathscr{F}_\ell\big]$. This yields the equations (27). □

4.2 Nonanticipativity constraints

A different situation arises with the use of *nonanticipativity constraints*. The fundamental idea reflected in this formulation, due to [33], is to consider extended spaces $\widetilde{\mathscr{X}_t} = \mathscr{L}_p(\Omega, \mathscr{F}, P; \mathbb{R}^n)$, $t = 1, \ldots, T$ and a relaxed policy

$$x = (x_1, \ldots, x_T) \in \widetilde{\mathscr{X}_1} \times \cdots \times \widetilde{\mathscr{X}_T} = \widetilde{\mathscr{X}}.$$

In order to enforce that the relaxed policy can be identified with an element of the space $\mathscr{X}$, we impose the following requirement known as the *nonaticipativity constraint*:

$$x_t = \mathbb{E}[x_t | \mathscr{F}_t], \quad t = 1, \ldots, T. \tag{29}$$

The equations (29) define a closed subspace $\mathscr{N}$ in $\widetilde{\mathscr{X}}$. This subspace can be identified with the space $\mathscr{X}$ in the original problem.

Several important theoretical and practical advantages are associated with this reformulation (see [32, Ch. 3] and the references therein). It allows to study individual scenario models, for each $\omega \in \Omega$, and to analyze the effect of the information constraint. It may also serve as the theoretical foundation for a variety of decomposition methods, similar to the case of linear dynamics and integral functionals (see [30] and the references therein).

In order to formally define the nonlinear problem in the space $\widetilde{\mathscr{X}}$ we need to extend the domains of the functional $\varphi(\cdot)$ and the domain and range of the operator $F(\cdot)$. We denote by $\widetilde{\varphi} : \widetilde{\mathscr{X}} \to \mathbb{R}$ a Lipschitz continuous extension of φ, that is, $\widetilde{\varphi}(x) = \varphi(x)$ for all $x \in \mathscr{N}$ (here we identify $\mathscr{N}$ with $\mathscr{X}$). Such an extension may be defined in various ways, for example, as

$$\widetilde{\varphi}(x_1, x_2, \dots, x_T) = \varphi\big(\mathbb{E}_1[x_1], \mathbb{E}_2[x_2], \dots, \mathbb{E}_T[x_T]\big).$$

An extension of a causal operator $F(\cdot)$ is natural from its definition; it is still given by (3). Its value space is $\widetilde{\mathscr{Y}} = \widetilde{\mathscr{Y}}_1 \times \cdots \times \widetilde{\mathscr{Y}}_T$ with $\widetilde{\mathscr{Y}}_t = \mathscr{L}_p(\Omega, \mathscr{F}, P; \mathbb{R}^m)$, $t = 1, \dots, T$. The decomposable sets X_t and Y_t can still be viewed as subsets $\widetilde{X}_t$ of $\widetilde{\mathscr{X}_t}$ and $\widetilde{Y}_t$ of $\widetilde{\mathscr{Y}_t}$:

$$\begin{aligned}
\widetilde{X}_t &= \{x_t \in \widetilde{\mathscr{X}_t} : x_t(\omega) \in X_t(\omega) \text{ a.s.}\},\\
\widetilde{Y}_t &= \{y_t \in \widetilde{\mathscr{Y}_t} : y_t(\omega) \in Y_t(\omega) \text{ a.s.}\}, \quad t = 1, \dots, T.
\end{aligned}$$

Notice that the sets $\widetilde{X}_t$ and $\widetilde{Y}_t$ contain more elements than their counterparts in the previous formulation, because they allow for a broader class of measurable selections from $X_t(\cdot)$ and $Y_t(\cdot)$, respectively.

The problem is re-formulated as follows:

$$\min\ \widetilde{\varphi}(x_1, \dots, x_T) \tag{30}$$
$$\text{s.t. } x_t - \mathbb{E}_t x_t = 0 \quad \text{a.s.}, \quad t = 1, \dots, T, \tag{31}$$
$$F_t(x_{1:t}) \in \widetilde{Y}_t \quad \text{a.s.}, \quad t = 1, \dots, T, \tag{32}$$
$$x_t \in \widetilde{X}_t \quad \text{a.s.}, \quad t = 1, \dots, T. \tag{33}$$

Simplified versions of this problem are considered in [12,26], under the assumption that $\varphi(x_1, \dots, x_T) = \mathbb{E}\big[\sum_{t=1}^T c_t(x_t(\omega), \omega)\big]$, with $c_t(\cdot, \cdot)$ being convex normal integrands. The authors use the space $\mathscr{L}_\infty(\Omega, \mathscr{F}, P; \mathbb{R}^n)$ to allow for the interior point conditions for the sets $\widetilde{X}_t$, but the price for this setting was that the dual elements live in the spaces of bounded finitely additive measures and can contain singular components. Specific properties of subdifferentials of expected value functionals in $\mathscr{L}_\infty$ spaces (see, [25] and [5, Ch. VII]) allow for the restriction of the dual elements to $\mathscr{L}_1(\Omega, \mathscr{F}_t, P; \mathbb{R}^n)$.

Our approach is different. We work in the space $\mathscr{L}_p(\Omega, \mathscr{F}, P; \mathbb{R}^n)$, with $p \in [1, \infty)$. We consider general Lipschitz continuous functionals $\varphi(\cdot)$, and a nonlinear causal operator $F(\cdot)$. Our idea is to require the existence of subregular recourse and to exploit its properties, as well as specific properties of causal operators to derive the optimality conditions. In this way, we relate assumptions on finite-dimensional systems associated with elementary events $\omega \in \Omega$ and stages $1, \dots, T$ with the optimality conditions for the entire system.

First, we prove subregularity of the constraints present in the problem formulation with explicit nonaticipativity constraints.

Theorem 4 *If the system (20)–(21) admits complete subregular recourse, then the system (31)–(33) is subregular at any feasible point $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_T)$.*

Proof Let $u = (u_1, \ldots, u_T) \in \widetilde{\mathscr{X}}$ be fixed. We shall construct a solution $\bar{x}$ of (31)–(33) which is close to u, with an appropriate error bound.

For $t = 1, \ldots, T$, we consider the following system in the space $\widetilde{\mathscr{X}_t}$:

$$\begin{aligned} F_t(\bar{x}_{1:t-1}, x_t) &\in \widetilde{Y}_t, \\ x_t - \mathbb{E}_t[x_t] &= 0, \\ x_t &\in \widetilde{X}_t. \end{aligned}$$

Our intention is to find a solution $\bar{x}_t$ to this system, which is sufficiently close to $\mathbb{E}_t[u_t]$. Using the Lipschitz continuity of $F_t(\cdot)$, we obtain

$$\left\| F_t\big(\bar{x}_{1:t-1}, \mathbb{E}_t[u_t]\big) \right\| \le \left\| F_t\big(u_{1:t}\big) \right\| + L\Big(\left\|\bar{x}_{1:t-1} - u_{1:t-1}\right\| + \left\|u_t - \mathbb{E}_t[u_t]\right\|\Big). \tag{34}$$

We define a multifunction $\mathfrak{G} : \Omega \rightrightarrows \mathbb{R}^n$ by the relations

$$\begin{aligned} \mathfrak{G}(\omega) = \Big\{ \xi :\ & f_t(\bar{x}_{1:t-1}(\omega), \xi, \omega) \in Y_t(\omega), \quad \xi \in X_t(\omega), \\ & \left\| \xi - \mathbb{E}_t[u_t](\omega) \right\| \\ & \le C\Big(d\big(f_t\big(\bar{x}_{1:t-1}(\omega), \mathbb{E}_t[u_t](\omega), \omega\big), Y_t(\omega)\big) + d\big(\mathbb{E}_t[u_t](\omega), X_t(\omega)\big)\Big)\Big\}. \end{aligned}$$

We observe that both distance terms on the right hand side are $\mathscr{F}_t$-measurable by [2, Corollary 8.2.13]. Therefore, the multifunction $\mathfrak{G}$ is $\mathscr{F}_t$-measurable. It has non-empty images due to Definition 6 applied with $\eta = \mathbb{E}_t[u_t](\omega)$ and $\zeta_{1:t-1} = \bar{x}_{1:t-1}(\omega)$. Hence, an $\mathscr{F}_t$-measurable selection $\bar{x}_t$ of $\mathfrak{G}$ exists (*cf.* [18]). From the construction of $\mathfrak{G}$,

$$\begin{aligned} & \left\| \bar{x}_t(\omega) - \mathbb{E}_t[u_t](\omega) \right\| \\ & \quad \le C\Big(d\big(f_t\big(\bar{x}_{1:t-1}(\omega), \mathbb{E}_t[u_t](\omega), \omega\big), Y_t(\omega)\big) + d\big(\mathbb{E}_t[u_t](\omega), X_t(\omega)\big)\Big). \end{aligned} \tag{35}$$

We view both sides of this inequality as nonnegative elements of the space $\mathscr{L}_p(\Omega, \mathscr{F}_t, P)$. Since it is a Banach lattice, the functional norm of the element on the left hand side does not exceed the functional norm of the element on right hand side. The triangle inequality yields:

$$\left\| \bar{x}_t - \mathbb{E}_t[u_t] \right\| \le C\Big(\operatorname{dist}\big(F_t(\bar{x}_{1:t-1}, \mathbb{E}_t[u_t]), Y_t\big) + \operatorname{dist}\big(\mathbb{E}_t[u_t], X_t\big)\Big). \tag{36}$$

For every $\tilde{x}_t \in \widetilde{\mathscr{X}_t}$, Jensen inequality implies that

$$\left\| \mathbb{E}_t[u_t] - \mathbb{E}_t[\tilde{x}_t] \right\| \le \left\| \mathbb{E}_t[u_t] - \tilde{x}_t \right\|$$

and $\mathbb{E}_t[\tilde{x}_t] \in X_t$ by convexity. Therefore, $\operatorname{dist}\big(\mathbb{E}_t[u_t], X_t\big) = \operatorname{dist}\big(\mathbb{E}_t[u_t], \widetilde{X}_t\big)$. Using a similar argument, we have $\operatorname{dist}\big(F_t(\bar{x}_{1:t-1}, \mathbb{E}_t[u_t]), Y_t\big) = \operatorname{dist}\big(F_t(\bar{x}_{1:t-1}, \mathbb{E}_t[u_t]), \widetilde{Y}_t\big)$.

We observe that both distances above are finite because $\big\| f_t(\bar{x}_{1:t-1}(\cdot), \mathbb{E}_t[u_t](\cdot), \cdot)\big\|$ has a finite $\mathscr{L}_p$-norm by virtue of (34) and the term $d\big(\mathbb{E}_t[u_t](\cdot), X_t^0(\cdot)\big)$ is bounded from above by $\|\mathbb{E}_t[u_t](\cdot) - \hat{x}_t(\cdot)\|$, which has a finite $\mathscr{L}_p$-norm by assumption.

Combining these observations with inequalities (36) and (34), we infer that

$$\begin{aligned}\big\|\bar{x}_t - \mathbb{E}_t[u_t]\big\| \leq\ & C\Big(\operatorname{dist}\big(F_t\big(u_{1:t}\big), \widetilde{Y}_t\big) + L\big(\big\|\bar{x}_{1:t-1} - u_{1:t-1}\big\| \\ & + \big\|u_t - \mathbb{E}_t[u_t]\big\|\big)\ + \operatorname{dist}\big(\mathbb{E}_t[u_t], \widetilde{X}_t\big)\Big).\end{aligned}$$

Since $\operatorname{dist}\big(\mathbb{E}_t[u_t], \widetilde{X}_t\big) \leq \operatorname{dist}\big(u_t, \widetilde{X}_t\big) + \big\|u_t - \mathbb{E}_t[u_t]\big\|$, we conclude that

$$\begin{aligned}\|\bar{x}_t - u_t\| \leq\ & (1 + C + CL)\big\|u_t - \mathbb{E}_t[u_t]\big\| \\ & + C\Big(\operatorname{dist}\big(F_t\big(u_{1:t}\big), \widetilde{Y}_t\big) + L\big\|\bar{x}_{1:t-1} - u_{1:t-1}\big\|\ + \operatorname{dist}\big(u_t, \widetilde{X}_t\big)\Big).\end{aligned} \tag{37}$$

We can now prove by induction that constants $\bar{C}_t$ exist such that

$$\|\bar{x}_t - u_t\| \leq \bar{C}_t \sum_{\ell=1}^{t} \Big(\big\|u_\ell - \mathbb{E}_\ell[u_\ell]\big\| + \operatorname{dist}\big(F_\ell(u_{1:\ell}), \widetilde{Y}_\ell\big) + \operatorname{dist}\big(u_\ell, \widetilde{X}_\ell\big)\Big).$$

For $t = 1$, the result follows from (37), because the term $\big\|\bar{x}_{1:t-1} - u_{1:t-1}\big\|$ is not present. Supposing it is true for $t-1$, we verify it for t using (37). The last relation for $t = T$ establishes the subregularity of the system (31)–(33). □

Abusing notation, we shall use the same notation for the operators

$$F_t'(\hat{x}_{1:t}) = A_t = \big(A_{t,1}, \ldots, A_{t,t}\big), \quad t = 1, \ldots, T,$$

referring to the partial Jacobians $A_{t,\ell} : \widetilde{\mathscr{X}_\ell} \to \widetilde{\mathscr{Y}_t}$, which are defined by (25)-(26), but are acting as linear operators between larger spaces.

Now, we can formulate the main result of this section.

Theorem 5 *Suppose the system (20)–(21) admits complete subregular recourse. If a policy $\hat{x}$ is a local minimum of problem (30)–(33) then a subgradient $\tilde{g} \in \partial\widetilde{\varphi}(\hat{x})$, multipliers $\lambda_t \in \widetilde{\mathscr{X}_t^*}$, $\widetilde{\psi}_t \in N_{\widetilde{Y}_t}(F_t(\hat{x}_{1:t}))$, $t = 1, \ldots, T$, and normal elements $\tilde{n}_t \in N_{\widetilde{X}_t}(\hat{x}_t)$, $t = 1, \ldots, T$, exist, such that for P-almost all $\omega \in \Omega$ we have:*

$$\tilde{g}_t + \lambda_t + \sum_{\ell=t}^{T} A_{\ell,t}^\top \widetilde{\psi}_\ell + \tilde{n}_t = 0, \quad t = 1, \ldots, T, \tag{38}$$

$$\mathbb{E}_t[\lambda_t] = 0, \quad t = 1, \ldots, T. \tag{39}$$

Proof We follow a similar line of argument as in Theorem 3. Using the Lipschitz continuity of $\widetilde{\varphi}(\cdot)$ about $\hat{x}$ with some Lipschitz constant L_φ, we infer that, for every $K > L_\varphi$, the point $\hat{x}$ is a local minimum of the function

$$\widetilde{\varphi}(x) + K\,\mathrm{dist}(x, \widetilde{X} \cap F^{-1}(\widetilde{Y}) \cap \mathscr{N}).$$

We define the linear operator $\Pi : \widetilde{\mathscr{X}} \to \widetilde{\mathscr{X}}$, by

$$\Pi(x_1, \ldots, x_T) = (\mathbb{E}_1[x_1], \ldots, \mathbb{E}_T[x_T]). \tag{40}$$

Theorem 4 implies that the system (31)–(33) is metrically subregular with some constant $\bar{C}$. Consequently, $\hat{x}$ is a local minimum of the function

$$\widetilde{\varphi}(x) + K\bar{C}\big(\mathrm{dist}(F(x), \widetilde{Y}) + \mathrm{dist}(x, \widetilde{X}) + \|x - \Pi x\|\big).$$

We use necessary conditions of optimality for Lipschitz continuous functions:

$$0 \in \partial\varphi(\hat{x}) + K\bar{C}\,\partial_x\big[\mathrm{dist}(F(\hat{x}), \widetilde{Y})\big] + K\bar{C}\,\partial\big[\mathrm{dist}(\hat{x}, \widetilde{X})\big] + K\bar{C}\,\partial\|\hat{x} - \Pi\hat{x}\|.$$

By virtue of Theorem 1, the subdifferential of the function $\mathrm{dist}(F(\cdot), \widetilde{Y})$ it is equal to $\big[F'(\hat{x})\big]^* \big(N_{\widetilde{Y}}(F(\hat{x})) \cap \mathbb{B}_{\widetilde{\mathscr{Y}}^*}\big)$. The subdifferential of $\mathrm{dist}(\hat{x}, \widetilde{X})$ is $N_{\widetilde{X}}(\hat{x}) \cap \mathbb{B}_{\widetilde{\mathscr{X}}^*}$. The subdifferential of the last term is $(I - \Pi^*)\mathbb{B}_{\widetilde{\mathscr{X}}^*}$. Using the tower property of conditional expectations, we see that

$$\Pi^*(v_1, \ldots, v_T) = (\mathbb{E}_1[v_1], \ldots, \mathbb{E}_T[v_T]).$$

Therefore,

$$\partial\|\hat{x} - \Pi\hat{x}\| = (I - \Pi^*)\mathbb{B}_{\widetilde{\mathscr{X}}^*} = [\ker(\Pi^*)] \cap \mathbb{B}_{\widetilde{\mathscr{X}}^*}.$$

Summing up, it follows that a subgradient $\tilde{g} \in \partial\widetilde{\varphi}(\hat{x})$, an element $\widetilde{\psi} \in N_{\widetilde{Y}}(F(\hat{x}))$, a normal vector $\tilde{n} \in N_{\widetilde{X}}(\hat{x})$, and a multiplier $\lambda \in \ker(\Pi^*)$ exist, such that

$$\tilde{g} + \lambda + \big[F'(\hat{x})\big]^*\widetilde{\psi} + \tilde{n} = 0.$$

The condition $\lambda \in \ker(\Pi^*)$ is equivalent to (39). Equations (38) can now be derived as in the proof of Theorem 3, using the block-triangular form of $A = F'(\hat{x})$, and equation (28) for any $h \in \widetilde{\mathscr{X}}$. Since both spaces, $\widetilde{\mathscr{Y}}_t^*$ and $\widetilde{\mathscr{X}}_\ell^*$, are defined with the use of the full σ-algebra $\mathscr{F}$, we simply have $A^*_{t,\ell} = A^\top_{t,\ell}$. That is why no conditional expectation appears in (38). $\square$

It may be of interest to explore the relations of two sets of optimality conditions of Theorems 3 and 5.

Corollary 2 *The subgradient $\hat{g} \in \partial\varphi(\hat{x})$ given by $\hat{g}_t = \mathbb{E}_t[\tilde{g}_t]$, $t = 1, \ldots, T$, together with the multipliers $\hat{\psi}_t = \mathbb{E}_t[\widetilde{\psi}_t]$, $t = 2, \ldots, T$, and normal vectors $\hat{n}_t = \mathbb{E}_t[\tilde{n}_t]$ satisfy the optimality conditions* (27).

Proof We take the conditional expectation of both sides of a typical relation in (38), first with respect to with respect to $\mathscr{F}_t$. Since $\mathbb{E}_t[\pi_t] = 0$, using the tower property and $\mathscr{F}_\ell$-measurability of $A_{\ell,t}$, we obtain

$$0 = \mathbb{E}_t\big[\tilde{g}_t\big] + \mathbb{E}_t\Bigg[\sum_{\ell=t}^{T} A_{\ell,t}^\top \widetilde{\psi}_\ell\Bigg] + \mathbb{E}_t\big[\tilde{n}_t\big] = \mathbb{E}_t\big[\tilde{g}_t\big]$$
$$+\mathbb{E}_t\Bigg[\sum_{\ell=t}^{T} A_{\ell,t}^\top \mathbb{E}_\ell[\widetilde{\psi}_\ell]\Bigg] + \mathbb{E}_t\big[\tilde{n}_t\big].$$

We shall verify that $\hat{g}$ is a subgradient of $\varphi(\cdot)$ at $\hat{x}$. Having in mind that $\tilde{g} \in \partial\widetilde{\varphi}(\hat{x})$, for any $x \in \mathscr{X}$, we have

$$\varphi(x) - \varphi(\hat{x}) \geq \sum_{t=1}^{T}\langle \tilde{g}_t, x_t - \hat{x}_t\rangle = \sum_{t=1}^{T}\big\langle \tilde{g}_t, \mathbb{E}_t[x_t - \hat{x}_t]\big\rangle = \sum_{t=1}^{T}\big\langle \mathbb{E}_t[\tilde{g}_t], x_t - \hat{x}_t\big\rangle,$$

and, thus, $\hat{g} \in \partial\varphi(\hat{x})$.

In a similar way, if $\tilde{n}_t \in N_{\widetilde{X}_t}(\hat{x}_t)$, then, for every $x_t \in X_t$, we have

$$0 \geq \langle \tilde{n}_t, x_t - \hat{x}_t\rangle = \big\langle \tilde{n}_t, \mathbb{E}_t[x_t - \hat{x}_t]\big\rangle = \big\langle \mathbb{E}_t[\tilde{n}_t], x_t - \hat{x}_t\big\rangle.$$

This proves that $\hat{n}_t \in N_{X_t}(\hat{x}_t)$, $t = 1, \ldots, T$. Likewise, we obtain $\hat{\psi}_t \in N_{Y_t}(F_t(\hat{x}_{1:t}))$ for $t = 1, \ldots, T$. □

5 Conclusions and future research

The concepts of metric subregularity and calmness and the associated penalty approach are very fruitful in the derivation of optimality conditions for nonlinear multistage stochastic optimization problems with general cost functionals. These new optimality conditions rest on two main contributions.

First, the concept of subregular recourse of Definition 6 allows for the verification of the subregularity of the constraint system in abstract spaces by establishing subsegularity of finite-dimensional systems associated with each stage and each elementary event.

Second, the calculation of the Clarke subdifferential of a composition of the distance function and a causal operator (Thm. 1) allows for exact subdifferentiation of the penalty function associated with system's dynamics.

In the course of these derivations we have introduced novel analysis techniques. We believe that this approach has much potential in addressing nonlinear stochastic dynamic optimization problems.

One direction of research would be to focus on specific dynamic risk measures and exploit their specific structure to further refine the conditions.

Another important avenue of research would be to adopt the setting of [20, Prop. 5.3] and [21, Thm. 5]. This requires restricting the spaces $\mathscr{X}$ and $\mathscr{Y}$ to those with $p \in (1, \infty)$. However, it would allow for the treatment of non-convex sets X_t and Y_t in the problem formulation, and the use of more accurate subdifferentials of non-convex objective functionals. The main challenge in this setting is to derive the explicit form of the coderivative of the operator $F(\cdot)$ describing the dynamics of the system.

Finally, the necessary conditions of optimality are a prerequisite for the development of numerical methods for solving nonlinear stochastic programming problems. We hope that this formidable challenge will be undertaken soon.

References

1. Appell, J., Zabrejko, P.P.: Nonlinear Superposition Operators. Cambridge University Press, Cambridge (1990)
2. Aubin, J.-P., Frankowska, H.: Set-Valued Analysis. Springer, Berlin (2009)
3. Banach, S.: Théorie des Opérations Linéaires. Monografje Matematyczne, Warszawa (1932)
4. Burke, J.V.: Calmness and exact penalization. SIAM J. Control Optim. **29**(2), 493–497 (1991)
5. Castaing, C., Valadier, M.: Convex Analysis and Measurable Multifunctions. Springer, Berlin (1977)
6. Clarke, F.H.: Optimization and Nonsmooth Analysis. SIAM, New Delhi (1990)
7. Corduneanu, C.: Functional Equations with Causal Operators. CRC Press, Boca Raton (2005)
8. Cuong, N. D., Kruger, A. Y.: Transversality properties: primal sufficient conditions. *Set-Valued and Variational Analysis*, pp. 1–36 (2020) (online first)
9. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings. Springer, Berlin (2009)
10. Eisner, M.J., Olsen, P.: Duality for stochastic programming interpreted as lp in l_p-space. SIAM J. Appl. Math. **28**(4), 779–792 (1975)
11. Evstigneev, I. V. : Lagrange multipliers for the problems of stochastic programming. In: M. Łoś, J. Łoś, and A. Wieczorek (eds.) Warsaw Fall Seminars in Mathematical Economics 1975, pp. 34–48. Springer (1976)
12. Flåm, S.D.: Nonanticipativity in stochastic programming. J. Optim. Theory Appl. **46**(1), 23–30 (1985)
13. Ioffe, A.D.: Necessary and sufficient conditions for a local minimum. 1: a reduction theorem and first order conditions. SIAM J. Control Optim. **17**(2), 245–250 (1979)
14. Ioffe, A.D.: Variational Analysis of Regular Mappings. Springer, Berlin (2017)
15. Klatte, D., Kummer, B.: Constrained minima and Lipschitzian penalties in metric spaces. SIAM J. Optim. **13**(2), 619–633 (2002)
16. Klatte, D., Kummer, B.: Nonsmooth equations in optimization: regularity, calculus, methods and applications, vol. 60. Springer, Berlin (2006)
17. Kruger, A.Y., Luke, D.R., Thao, N.H.: Set regularities and feasibility problems. Math. Program. **168**(1–2), 279–311 (2018)
18. Kuratowski, K., Ryll-Nardzewski, C.: A general theorem on selectors. Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **13**(6), 397–403 (1965)
19. Mordukhovich, B.S.: Complete characterization of openness, metric regularity, and lipschitzian properties of multifunctions. Trans. Am. Math. Soc. **340**(1), 1–35 (1993)
20. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation II Applications. Springer, Berlin (2006)
21. Movahedian, N.: Calmness and inverse image characterizations for asplund spaces. Optim. Lett. **7**(2), 361–373 (2013)
22. Outrata, J.V., Römisch, W.: On optimality conditions for some nonsmooth optimization problems over L_p spaces. J. Optim. Theory Appl. **126**(2), 411–438 (2005)
23. Phelps, R.R.: Convex Functions, Monotone Operators and Differentiability, vol. 1364. Springer, Berlin (2009)
24. Robinson, S. M.: Generalized equations and their solutions, Part I: Basic theory. In: Huard, P., Auslender, A (eds.) *Point-to-Set Maps and Mathematical Programming*, pp. 128–141. Springer (1979)
25. Rockafellar, R.T.: Integrals which are convex functionals. II. Pacific J. Math. **39**(2), 439–469 (1971)

26. Rockafellar, R.T., Wets, R.J.-B.: Nonanticipativity and l_1-martingales in stochastic optimization problems. Math. Program. Study Stochast. Syst. Model. Ident. Optim. **II**(6), 170–187 (1976)
27. Rockafellar, R.T., Wets, R.J.-B.: Stochastic convex programming: basic duality. Pacific J. Math. **62**(1), 173–195 (1976)
28. Rockafellar, R.T., Wets, R.J.-B.: Stochastic convex programming: relatively complete recourse and induced feasibility. SIAM J. Control Optim. **14**(3), 574–589 (1976)
29. Rockafellar, R.T., Wets, R.J.-B.: Stochastic convex programming: singular multipliers and extended duality singular multipliers and duality. Pacific J. Math. **62**(2), 507–522 (1976)
30. Ruszczyński, A.: Decomposition methods. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, pp. 141–211. Elsevier, Amsterdam (2003)
31. Ruszczyński, A., Shapiro, A.: Stochastic programming models. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, pp. 1–64. Elsevier, Amsterdam (2003)
32. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming. SIAM, Philadelphia, Pennsylvania (2009)
33. Wets, R. J.-B.: On the relation between stochastic and deterministic optimization. In: Bensoussan, A., Lions, J. L. (eds.) Control Theory, Numerical Methods and Computer Systems Modelling, pp. 350–361. Springer (1975)