Subregular Recourse in Nonlinear Multistage Stochastic Optimization

Darinka Dentcheva*
Andrzej Ruszczynski†

January 22, 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

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 monographs [9, 13] 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 $p$-integrable functions on a probability space, with $p \in [1, \infty)$. Such problems arise in stochastic optimization, and despite much effort devoted to their analysis in the past, they still pose formidable theoretical challenges. These challenges are compounded by the fact that stochastic optimization models may involve complex risk functionals which cannot be expressed as expected values of stage-wise costs. The classical approaches, exploiting the properties of convex integral functionals, are inapplicable to such models. Furthermore, we consider problems with nonlinear dynamics, where techniques of conjugate duality cannot be used.

Our contributions can be summarized as follows. We introduce a new concept of subregular recourse, and we establish subregularity of a system of constraints in a multistage stochastic optimization problem

*Stevens Institute of Technology, Department of Mathematical Sciences, Hoboken, NJ 07030, USA; email: darinka.dentcheva@stevens.edu
†Rutgers University, Department of Management Science and Information Systems, Piscataway, NJ 08854, USA; email: rusz@rutgers.edu
with nonlinear dynamics in two settings: with built-in nonanticipativity and with explicit nonanticipativity constraints. We derive exact Clarke subdifferentials of penalty functions involving causal operators. Our main results are optimality conditions for nonlinear multistage stochastic optimization problems with general objective functions in both settings.

The paper is organized as follows. In §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 §3, we derive useful properties of causal operators describing the dynamics of the system. Finally, §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 $\left(\Omega, \mathcal{F}, P\right)$, the notation $\mathcal{X} = \mathcal{L}_p\left(\Omega, \mathcal{F}, P; \mathbb{R}^n\right)$ stands for the vector space of measurable functions $x : \Omega \to \mathbb{R}^n$, such that $\int \left\| x(\omega) \right\|^p P(d\omega) < \infty$, where $p \in [1, \infty)$. We denote the norm in $\mathcal{X}$ 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 $\text{dist}\left(\cdot, A\right)$, while the distance to $B$ in a finite dimensional space will be denoted by $d(\cdot, B)$.

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

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

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

**Definition 2.1.** Suppose $A$ is a closed subset of $\mathcal{X}$ and $x \in A$. The contingent cone to $A$ at $x$ is the set

$$\mathcal{T}_A(x) = \left\{ v \in \mathcal{X} : \liminf_{\tau \downarrow 0} \frac{1}{\tau} \text{dist}_{\mathcal{X}}(x + \tau v, A) = 0 \right\}.$$

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

$$\mathcal{K}^\circ = \left\{ y \in \mathcal{X}^* : \langle y, x \rangle \leq 0 \text{ for all } x \in \mathcal{K} \right\}.$$

**Definition 2.2.** A set $A \subset \mathcal{X}$ is derivable at $x \in A$ if for every $v \in \mathcal{T}_A(x)$

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

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

**Definition 2.3.** A set $\mathcal{K} \subset \mathcal{X}$ is decomposable if a measurable multifunction $K : \Omega \rightrightarrows \mathbb{R}^n$ exists, such that $\mathcal{K} = \left\{ x \in \mathcal{X} : x(\omega) \in K(\omega) \text{ a.s.} \right\}$.

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

**Lemma 2.4.** Suppose $A \subset \mathcal{X}$ is decomposable and $A(\omega)$ are closed and derivable sets for $P$-almost all $\omega \in \Omega$. Then

$$\mathcal{T}_A(x) = \left\{ v \in \mathcal{X} : \text{ for } P\text{-almost all } \omega, v(\omega) \in \mathcal{T}_{A(\omega)}(x(\omega)) \right\}.$$
Polar cones of convex decomposable cones are also decomposable.

**Lemma 2.5.** The polar cone $K^\circ$ of a decomposable cone $K \subset X$ is a convex decomposable cone, and $K^\circ(\omega) = (K(\omega))^\circ$ a.s.

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

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

Hence, $y \in K^\circ$ and $D \subset K^\circ$. We show that $K^\circ = D$ by contradiction. Suppose an element $y \in K^\circ$ exists, such that the event $S = \{y(\omega) \notin (K(\omega))^\circ\}$ has positive probability. Then, for every $C > 0$ we can find a function $x \in X^*$ such that $x(\omega) \in K(\omega)$ and $\langle y(\omega), x(\omega) \rangle > C$ for all $\omega \in S$. For $\omega \in \Omega \setminus S$ we select $x(\omega) \in K(\omega) \cap B_\delta$, where $B_\delta$ is a ball in $\mathbb{R}^n$ of radius $\delta > 0$. Then

$$\langle y, x \rangle = \int_S y(\omega)^T x(\omega) \, P(d\omega) + \int_{\Omega \setminus S} y(\omega)^T x(\omega) \, P(d\omega) \geq CP(S) - \delta \|y\|_{X^*}.$$

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

We recall the subregularity concept regarding set-constrained systems. For thorough treatment of regularity conditions, we refer the reader to [16, 8] and the references therein. For a multifunction $H: X \Rightarrow Y$, where $Y$ is a Banach space, we consider the relation

$$0 \in H(x). \quad (1)$$

**Definition 2.6.** The multifunction $H$ is subregular at $\hat{x} \in X$ with $0 \in H(\hat{x})$, if $\delta > 0$ and $C > 0$ exist such that for all $x \in X$ with $\|x - \hat{x}\|_X \leq \delta$ a point $\tilde{x}$ satisfying (1) exists such that

$$\|\tilde{x} - x\|_X \leq C \text{dist}_Y(0, H(x)).$$

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

$$F(x) \in Y, \quad (2)$$

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

$$\text{dist}(x, F^{-1}(Y)) \leq C \text{dist}(F(x), Y).$$
3 Causal Operators

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

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

**Definition 3.1.** An operator \(F : \mathcal{X} \to \mathcal{Y}\) is causal, if functions \(f_t : \mathbb{R}^n \times \Omega \to \mathbb{R}^m\) exist, such that for all \(t = 1, \ldots, T\)

\[
F_t(x)(\omega) = f_t(x_{1:t}(\omega), \omega), \quad \omega \in \Omega,
\]

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).

**Assumption 1.** For all \(t = 1, \ldots, T\):

(i) \(f_t(\xi, \cdot)\) is an element of \(\mathcal{Y}_t\) for all \(\xi \in \mathbb{R}^m\);

(ii) For almost all \(\omega \in \Omega\), \(f_t(\cdot, \omega)\) is continuously differentiable with respect to its first argument, with the Jacobian \(f'_t(\cdot, \omega)\);

(iii) A constant \(C_f\) exists, such that \(\|f'_t(\cdot, \omega)\| \leq C_f\), almost surely.

Under Assumption 1 each \(F_t\) given by (3) indeed maps the product space \(\mathcal{X}_{1:t}\) into a subset of \(\mathcal{Y}_t\).

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 \(h \in \mathcal{X}\). For simplicity, we use the same notation as if it were acting on the entire \(h(\omega)\). Then we can write

\[
f'_t(x(\omega), \omega) = \{f'_t(x_{1:t}(\omega), \omega)\}_{t=1,\ldots,T}
\]
to represent the Jacobian of \(F(x)(\omega)\) with respect to \(x(\omega)\).

**Lemma 3.2.** If Assumption 1 is satisfied, then \(F(\cdot)\) is Gâteaux differentiable with the derivative \(F'(x)\) defined by

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

**Proof.** We define \(J(x) : \mathcal{X} \to \mathcal{Y}\) 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 \(h \in \mathcal{X}\) and \(\tau > 0\)

\[
\frac{1}{\tau} \left\| f(x(\omega) + \tau h(\omega), \omega) - f(x(\omega), \omega) - \tau f'(x(\omega), \omega) h(\omega) \right\| \leq 2C_f \|h(\omega)\| \quad \text{a.s.}
\]

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

\[
\frac{1}{\tau} \left\| F(x + \tau h) - F(x) - \tau J(x) h \right\|_\mathcal{Y}
= \left( \int \left\| \frac{1}{\tau} \left( f(x(\omega) + \tau h(\omega), \omega) - f(x(\omega), \omega) - \tau f'(x(\omega), \omega) h(\omega) \right) \right\|^p P(d\omega) \right)^{1/p}
\leq 2C_f \left( \int \|h(\omega)\|^p P(d\omega) \right)^{1/p} = 2C_f \|h\|_\mathcal{Y}.
\]
Using Lebesgue’s dominated convergence theorem, we obtain

\[
\lim_{\tau \to 0} \frac{1}{\tau} \|F(x + \tau h) - F(x) - \tau J(x) h\|_{\mathcal{Y}} = \\
\left( \int \lim_{\tau \to 0} \frac{1}{\tau} \left( f(x(\omega) + \tau h(\omega), \omega) - f(x(\omega), \omega) - \tau f'(x(\omega), \omega) h(\omega) \right) \|P(d\omega) \right)^{1/p} = 0.
\]

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

\[\square\]

It is worth mentioning that our assumptions do not guarantee the Fréchet differentiability of \( F(\cdot) \). Nonetheless, in the next result, we are able to calculate the Clarke subdifferential of the function

\[
\Phi(\cdot) = \text{dist}(F(\cdot), Y).
\]

**Theorem 3.3.** Suppose \( Y \subseteq \mathcal{Y} \) is convex and closed, \( F(x) \in Y \), and Assumption \( \square \) is satisfied. Then

\[
\partial \Phi(x) = \left[ F'(x) \right]^* (N_Y(F(x)) \cap \mathcal{B}_{\mathcal{Y}^*}),
\]

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

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

\[
\text{dist}_{\mathcal{Y}}(F(z + \tau h), Y) - \text{dist}_{\mathcal{Y}}(F(z), Y) \leq \langle g, F(z + \tau h) - F(z) \rangle,
\]

for any \( g \in \partial \text{dist}(y, Y) \) 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:

\[
\Phi^0(x; h) = \sup_{\tau \to 0} \frac{1}{\tau} \left( \text{dist}_{\mathcal{Y}}(F(z + \tau h), Y) - \text{dist}_{\mathcal{Y}}(F(z), Y) \right) \leq \sup_{\tau \to 0} \left( g, \frac{1}{\tau} (F(z + \tau h) - F(z)) \right),
\]

for any \( g \in \partial \text{dist}(F(z + \tau h), Y) \). Consider arbitrary sequences \( \{z_k\} \to x \) and \( \{\tau_k\} \downarrow 0 \). 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:

\[
\frac{1}{\tau_k} \left[ f_j(z_k(\omega) + \tau_k h(\omega), \omega) - f_j(z_k(\omega), \omega) \right] = f'_j(\bar{z}_{k,j}(\omega), \omega) h(\omega), \quad j = 1, \ldots, mT,
\]

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

\[
\frac{1}{\tau_k} \left[ f(z_k(\omega) + \tau_k h(\omega), \omega) - f(z_k(\omega), \omega) \right] = [F'(x)] h(\omega) + \Delta_k(\omega),
\]

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

\[
\Delta_{k,j}(\omega) = \left[ f'_j(\bar{z}_{k,j}(\omega), \omega) - f'_j(x(\omega), \omega) \right] h(\omega), \quad j = 1, \ldots, mT.
\]

We shall verify that \( \{\Delta_k\} \to 0 \) in \( \mathcal{Y} \). For an arbitrary \( \varepsilon > 0 \) we define the events

\[
\Omega_{k,\varepsilon} = \left\{ \omega \in \Omega : \max_{1 \leq j \leq mT} \|\bar{z}_{k,j}(\omega) - x(\omega)\| > \varepsilon \right\}.
\]

5
Since \( \{z_{k,j}\} \to x \) in \( \mathcal{Y} \), as \( k \to \infty \), the convergence in probability follows:

\[
\lim_{k \to \infty} P\left[ \Omega_{k,\varepsilon} \right] = 0. \tag{8}
\]

Let

\[
\delta(\varepsilon, \omega) = \sup_{\|y-x(\omega)\| \leq \varepsilon} \max_{1 \leq i \leq m} \left\| f'_i(y, \omega) - f'_i(x(\omega), \omega) \right\|. \tag{9}
\]

By the boundedness and continuity of the derivatives, \( \delta(\varepsilon, \omega) \leq 2C_f \), 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:

\[
\|\Delta_k(\omega)\| \leq 2C_f \mathbb{1}_{\Omega_{k,\varepsilon}}(\omega)\|h(\omega)\| + \delta(\varepsilon, \omega) \mathbb{1}_{\Omega_{k,\varepsilon}^c}(\omega)\|h(\omega)\|. \tag{10}
\]

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

\[
\int \mathbb{1}_{\Omega_{k,\varepsilon}}\|h(\omega)\|^p \, P(d\omega) > \alpha, \quad \text{for} \quad k \in \mathcal{K} \subset \mathcal{N},
\]

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

\[
\lim_{k \to \infty} \int_{k \in \mathcal{K}_1} \mathbb{1}_{\Omega_{k,\varepsilon}}\|h(\omega)\|^p \, P(d\omega) = 0,
\]

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 \( \mathcal{Y} \).

Combining this with (7), we obtain (in the space \( \mathcal{Y} \))

\[
\limsup_{k \to \infty} \|\Delta_k\| \leq \left( \int (\delta(\varepsilon, \omega)\|h(\omega)\|)^p \, P(d\omega) \right)^{1/p}.
\]

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

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

\[
\Phi^0(x; h) \leq \sup_{z \in \mathcal{K}} \frac{1}{\tau_k} \left( g_k, \frac{1}{\tau_k} \left( F(z_k + \tau_k h) - F(z_k) \right) \right) \leq \sup_{z \in \mathcal{K}} \left( g_k, F'(x)h + \Delta_k \right).
\]

All subgradients \( g_k \) are bounded by the Lipschitz constant 1 of the distance function. Therefore, \( \left( g_k, \Delta_k \right) \to 0 \). Consider an arbitrary accumulation point \( \alpha \) of the sequence \( \left( g_k, F'(x)h \right) \). By the Banach–Alaoglu theorem, we can choose a sub-subsequence \( \{g_k\}_{k \in \mathcal{K}} \) which is weakly* convergent to some \( g \) in \( \mathcal{Y}^* \). Then \( \alpha = \left( g, F'(x)h \right) \). By the norm-to-weak* upper semicontinuity of the subdifferential [19, Prop. 2.5], \( g \in \partial \text{dist}(F(x), Y) \). Therefore,

\[
\Phi^0(x; h) \leq \max_{g \in \partial \text{dist}(F(x), Y)} \left( g, F'(x)h \right). \tag{11}
\]

The converse inequality follows from (6) by setting \( z = x \) and using Lemma 3.2.

\[
\Phi^0(x; h) \geq \limsup_{\tau \to 0} \frac{1}{\tau} \left( \text{dist}_\mathcal{Y}(F(x + \tau h), Y) - \text{dist}_\mathcal{Y}(F(x), Y) \right)
\geq \limsup_{\tau \to 0} \frac{1}{\tau} \left( g, F(x + \tau h) - F(x) \right) = \left( g, F'(x)h \right).
\]
for any \( g \in \partial \text{dist}(F(x), Y) \). Therefore,
\[
\Phi^0(x; h) \geq \max_{g \in \partial \text{dist}(F(x), Y)} \langle g, F'(x) h \rangle.
\]
Combining this with (11), we infer that
\[
\Phi^0(x; h) = \max_{g \in \partial \text{dist}(F(x), Y)} \langle [F'(x)]^* g, h \rangle.
\]
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
\[
\partial \Phi(x) = \{ [F'(x)]^* g : g \in \partial \text{dist}(F(x), Y) \}.
\]
Having in mind that \( \partial \text{dist}(y, Y) = N_Y(y) \cap \mathbb{R} \) whenever \( y \in Y \), we obtain the stated result.

\[ \square \]

**Remark 3.4.** The causality of the operator \( F \) does not play a role in the proof of Theorem 3.3. The result is true for any superposition operator \( \tilde{F} : \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m) \rightarrow \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m) \), defined by \( \tilde{F}(x)(\omega) = f(x(\omega), \omega) \), whenever \( f \) satisfies conditions (ii)–(iii) of Assumption 1.

## 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 \( \mathcal{F}_t \)-measurability of the decisions at time \( t \), \( t = 1, \ldots, T \). In another formulation, we consider decision spaces of \( \mathcal{F}_t \)-measurable decisions at each stage, but add additional linear constraints enforcing \( \mathcal{F}_t \)-measurability.

### 4.1 The Model with Build-In Nonanticipativity

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

The dynamics of the system is represented by the relation
\[
F(x) \in Y,
\]
where \( F : \mathcal{X} \rightarrow \mathcal{Y} \) is a causal operator, and \( Y = Y_1 \times \cdots \times Y_T \), with each \( Y_t : \Omega \Rightarrow \mathbb{R}^m, t = 1, \ldots, T \), being an \( \mathcal{F}_t \)-measurable multifunction with convex and closed values. In a more explicit way, the relation (12) has the form:
\[
F_t(x_{1:t}) \in Y_t, \quad t = 1, \ldots, T,
\]
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, $\mathcal{F}_t$-measurable multifunctions with closed convex images $X_t : \Omega \to \mathbb{R}^n$, $t = 1, \ldots, T$, are defined.

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

$$\begin{align*}
\min & \quad \varphi(x_{1:T}) \\
\text{s.t.} & \quad F_t(x_{1:t}) \in Y_t \quad \text{a.s.,} \quad t = 1, \ldots, T, \\
& \quad x_t \in X_t \quad \text{a.s.,} \quad t = 1, \ldots, T.
\end{align*}$$

(14)–(15), with linear operators

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.

Current theory of stochastic optimization provides optimality conditions for convex versions of problem (14)–(15), with linear operators $F_t(\cdot)$ and expected value functionals $\varphi(x_1, \ldots, x_T) = \mathbb{E} \left[ \sum_{t=1}^{T} c_t(x_t(\omega), \omega) \right]$, involving convex integrands $c_t(\cdot, \cdot)$, see [10, 22, 24, 23, 11, 25, 18].

We expand the theory by allowing non-linear dynamics and more general functionals in the model description.

We use uniform parametric subregularity of deterministic set-constrained systems associated with each stage $t = 1, \ldots, T$ and each elementary event $\omega \in \Omega$:

$$f_t(\xi_{t-1}, \xi_t, \omega) \in Y_t(\omega), \quad \xi \in X_t(\omega).$$

(17)–(18)

Here, $\xi_{t-1} \in \mathbb{R}^{n(t-1)}$ representing 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 $\xi_{t-1}$ is non-existent.

We introduce the following concept.

**Definition 4.1.** The system (17)–(18) admits complete subregular recourse, if a constant $C$ exist, such that for almost all $\omega \in \Omega$, every $\xi_{t-1} \in X_{t-1}(\omega)$ and every $\eta \in \mathbb{R}^n$, a solution $\xi$ of (17)–(18) exists, satisfying the inequality

$$\| \xi - \eta \| \leq C \left( d(f_t(\xi_{t-1}, \eta, \omega), Y_t(\omega)) + d(\eta, X_t(\omega)) \right).$$

(19)

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

**Theorem 4.2.** If the system (17)–(18) 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 $u = (u_1, \ldots, u_T) \in \mathcal{X}$ be chosen from a sufficiently small neighborhood of $\hat{x}$. We shall construct a solution $\tilde{x}$ 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 $\mathcal{X}_t$:

$$F_t(\xi_{t-1}, x_t) \in Y_t,$$

$$x_t \in X_t.$$

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

$$\| F_t(\tilde{x}_{t-1}, u_t) \| \leq \| F_t(u_{1:t}) \| + L \| \tilde{x}_{t-1} - u_{1:t-1} \|,$$

(19)
We define a multifunction \( \mathcal{G} : \Omega \rightarrow \mathbb{R}^n \) by the relations
\[
\mathcal{G}(\omega) = \left\{ \tilde{x} \in \mathbb{R}^n : f_t(\tilde{x}_{1:t-1}, \tilde{x}_t, \omega) \in Y_t(\omega), \tilde{x}_t \in X_t(\omega), \right\}
\]
\[
\|\tilde{x}_t(\omega) - u_t(\omega)\| \leq C \left( d(\tilde{x}_{1:t-1}(\omega), u_t(\omega), Y_t(\omega)) + d(u_t(\omega), X_t(\omega)) \right) .
\]
We observe that both distance functions in the definition of \( \mathcal{G}(\cdot ) \) are \( \mathcal{F}_t \)-measurable by [2, Corollary 8.2.13]. Therefore, the multifunction \( \mathcal{G} \) is \( \mathcal{F}_t \)-measurable. It has non-empty images due to Definition 4.1 applied with with \( \eta = u_t(\omega) \) and \( \xi_{1:t-1} = \tilde{x}_{1:t-1}(\omega) \). Hence, a measurable selection \( \bar{x}_t \) of \( \mathcal{G} \) exists (cf. [17]). From the construction of the multifunction \( \mathcal{G} \),
\[
\|\bar{x}_t(\omega) - u_t(\omega)\| \leq C \left( d(\tilde{x}_{1:t-1}(\omega), u_t(\omega), Y_t(\omega)) + d(u_t(\omega), X_t(\omega)) \right) .
\]
Therefore, with the norms and distances in the spaces \( \mathcal{X}_t \) and \( \mathcal{Y}_t \),
\[
\|\bar{x}_t - u_t\| \leq C \left( \text{dist}(F_t(\tilde{x}_{1:t-1}, u_t), Y_t) + \text{dist}(u_t, X_t) \right) .
\](20)
Combining inequalities (20) and (19), we infer that
\[
\|\bar{x}_t - u_t\| \leq C \left( \text{dist}(F_t(u_{1:t}), Y_t) + L(\|\tilde{x}_{1:t-1} - u_{1:t-1}\|) + \text{dist}(u_{t}, X_t) \right) .
\](21)
We can now prove by induction that constants \( \bar{C}_t \) exist such that
\[
\|\bar{x}_t - u_t\| \leq \bar{C}_t \sum_{\ell=1}^{t} \left( \text{dist}(F_{\ell}(u_{1:\ell}), Y_{\ell}) + \text{dist}(u_{\ell}, X_{\ell}) \right) .
\]
For \( t = 1 \), the result is provided by (21), because the term \( \|\tilde{x}_{1:t-1} - u_{1:t-1}\| \) is not present there. Supposing it is true for \( t - 1 \), we verify it for \( t \) using (21). The last relation for \( t = T \) establishes the subregularity of the system (15)–(16).

\[ \square \]

Under Assumption 1, we denote:
\[
F_t'(\bar{x}_{1:t}) = A_t = (A_{t,1}, \ldots, A_{t,T}), \quad t = 1, \ldots, T,
\]
with partial Jacobians \( A_{t,\ell} : \mathcal{X}_\ell \rightarrow \mathcal{Y}_\ell \),
\[
A_{t,\ell} = \frac{\partial F_t(\bar{x}_{1:t})}{\partial x_{\ell}}, \quad \ell = 1, \ldots, T, \quad t = 1, \ldots, T.
\](22)
These linear operators are defined pointwise:
\[
A_{t,\ell}(\omega) = \frac{\partial f_t(\tilde{x}_{1:t}(\omega), \omega)}{\partial x_{\ell}(\omega)}, \quad \ell = 1, \ldots, T, \quad t = 1, \ldots, T, \quad \omega \in \Omega.
\](23)
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 4.3. Suppose the system (17)–(18) 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 \( \hat{\psi}_t \in N_{Y_t}(F_t(\hat{x}_t)) \), \( t = 1, \ldots, T \), and normal elements \( \hat{n}_t \in N_X(\hat{x}_t) \), \( t = 1, \ldots, T \), exist, such that for \( P \)-almost all \( \omega \in \Omega \) we have:

\[
\hat{g}_t + A_{t,1}^* \hat{\psi}_t + \mathbb{E}_t \left[ \sum_{t'=t+1}^T A_{t',1}^* \hat{\psi}_{t'} \right] + \hat{n}_t = 0, \quad t = 1, \ldots, T. \tag{24}
\]

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

\[
\varphi(x) + K \text{dist}(x, X \cap F^{-1}(Y));
\]

see [6, Prop. 2.4.3]. The system (15)–(16) is subregular with some constant \( \hat{C} \) by virtue of Theorem 4.2. Consequently, \( \hat{x} \) is a local minimum of the function

\[
\varphi(x) + K \hat{C} \left( \text{dist}(F(x), Y) + \text{dist}(x, X) \right).
\]

This type of argument is discussed in detail in [14,4,15]. We use Clarke’s necessary conditions of optimality for Lipschitz continuous functions:

\[
0 \in \partial \varphi(\hat{x}) + K \hat{C} \partial \left[ \text{dist}(F(\cdot), Y) \right](\hat{x}) + K \hat{C} \partial \left[ \text{dist}(\cdot, X) \right](\hat{x}).
\]

The Clarke-subdifferential of the function \( \text{dist}(F(\cdot), Y) \) is calculated in Theorem 4.3

\[
\partial \Phi(\hat{x}) = \left[ F'(\hat{x}) \right]^* \left( N_Y(F(\hat{x})) \cap \mathbb{B}_y \right)
\]

The subdifferential of \( \text{dist}(\hat{x}, X) \) is \( N_X(\hat{x}) \cap \mathbb{B}_x \). 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

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

We can derive a more explicit form of the vector \( \left[ F'(\hat{x}) \right]^* \hat{\psi} \). Due to the decomposability of \( X_t \), we can apply Lemmas 24 and 25 to obtain that the normal cone \( N_{X_t}(x) \) is composed of elements which are selectors of \( N_{X_t}(x) \); we have \( N_{X_t}(\hat{x}_t)(\omega) = N_{X_t}(\omega)(\hat{x}_t(\omega)) \) a.s. Using the same argument and the causality of \( F_t \), we obtain

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

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

\[
\langle A^* \hat{\psi}, h \rangle = \langle \hat{\psi}, A h \rangle = \sum_{t=1}^T \langle \hat{\psi}_t, A_t h_t \rangle = \sum_{t=1}^T \sum_{t'=t}^T \langle \hat{\psi}_{t'}, A_{t',t} h_t \rangle = \sum_{t'=1}^T \sum_{t=1}^T \langle A_{t,t'}^* \hat{\psi}_{t'}, h_t \rangle. \tag{25}
\]

It follows that \( A_{t,t'}^* \hat{\psi}_{t'} = \mathbb{E}_{t} \left[ A_{t,t'}^* \hat{\psi}_{t'} \mid \mathcal{F}_t \right] \). This yields the equations (24). \qed
4.2 Nonanticipativity Constraints

A different situation arises with the use of nonanticipativity constraints. The fundamental idea reflected in this formulation, due to \[26\], is to consider extended spaces \( \mathcal{X}_t = \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n) \), \( t = 1, \ldots, T \) and a relaxed policy

\[
x_t = (x_1, \ldots, x_T) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_T = \mathcal{X}.
\]

In order to enforce that the relaxed policy can be identified with an element of the space \( \mathcal{X} \), we impose the following requirement known as nonanticipativity constraint:

\[
x_t = \mathbb{E}[x_t | \mathcal{F}_t], \quad t = 1, \ldots, T.
\]

The equations \[26\] define a closed subspace \( \mathcal{N} \) in \( \mathcal{X} \). This subspace can be identified with the space \( \mathcal{X} \) in the original problem.

In what follows, we use the notation \( \mathbb{E}_t[x_t] \) for \( \mathbb{E}[x_t | \mathcal{F}_t] \).

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

\[
\tilde{\varphi}(x_1, x_2, \ldots, x_T) = \varphi(\mathbb{E}_1[x_1], \mathbb{E}_2[x_2], \ldots, \mathbb{E}_T[x_T]).
\]

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

\[
\tilde{X}_t = \{ x_t \in \mathcal{X}_t : x_t(\omega) \in X_t(\omega) \text{ a.s.} \},
\]

\[
\tilde{Y}_t = \{ y_t \in \mathcal{Y}_t : y_t(\omega) \in Y_t(\omega) \text{ a.s.} \}, \quad t = 1, \ldots, T.
\]

Notice that the sets \( \tilde{X}_t \) and \( \tilde{Y}_t \) 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:

\[
\min \tilde{\varphi}(x_1, \ldots, x_T)
\]

\[
\text{s.t. } x_t - \mathbb{E}_t x_t = 0 \quad \text{a.s.,} \quad t = 1, \ldots, T,
\]

\[
F_t(x_{1:t}) \in \tilde{Y}_t \quad \text{a.s.,} \quad t = 1, \ldots, T,
\]

\[
x_t \in \tilde{X}_t \quad \text{a.s.,} \quad t = 1, \ldots, T.
\]

Simplified versions of this problem are considered in \[21,12\], under the assumption that \( \varphi(x_1, \ldots, x_T) = \mathbb{E}\left[ \sum_{t=1}^T c_t(x_t(\omega), \omega) \right] \), with \( c_t(\cdot, \cdot) \) being convex normal integrands. The authors use the space \( \mathcal{L}_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^n) \) to allow for the interior point conditions for the sets \( \tilde{X}_t \), 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 \( \mathcal{L}_\infty \) spaces (see, \[20\] and \[5\] Ch. VII) allow for the restriction of the dual elements to \( \mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R}^n) \).

Our approach is different. We work in the space \( \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n) \), 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 \( 1, \ldots, T \) with the optimality conditions for the entire system.

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

**Theorem 4.4.** If the system (17)–(18) admits complete subregular recourse, then the system (28)–(30) is subregular at any feasible point \( \tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_T) \).

**Proof.** Let \( u = (u_1, \ldots, u_T) \in \tilde{X} \) be fixed. We shall construct a solution \( \bar{x} \) of (28)–(30) which is close to \( u \), with an appropriate error bound.

For \( t = 1, \ldots, T \), we consider the following system in the space \( \tilde{\mathcal{X}}_t \):

\[
F_t(\tilde{x}_{t-1}, x_t) \in \tilde{Y}_t,
\]

\[
x_t - E_t[x_t] = 0,
\]

\[
x_t \in \tilde{X}_t.
\]

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

\[
\| F_t(\tilde{x}_{t-1}, E_t[u_t]) \| \leq \| F_t(u_{t-1}) \| + L(\| \tilde{x}_{t-1} - u_{t-1} \| + \| u_t - E_t[u_t] \|).
\]

(31)

We define a multifunction \( \mathcal{G} : \Omega \to \mathbb{R}^n \) by the relations

\[
\mathcal{G}(\omega) = \left\{ \xi : f_t(\tilde{x}_{t-1}(\omega), \bar{x}, \omega) \in Y_t(\omega), \quad \xi \in \mathcal{X}_t(\omega),
\right. \\
\left. \| \xi - E_t[u_t](\omega) \| \leq C \left( d(f_t(\tilde{x}_{t-1}(\omega), E_t[u_t](\omega), \omega), Y_t(\omega)) + d(E_t[u_t](\omega), \mathcal{X}_t(\omega)) \right) \right\}.
\]

We observe that both distance terms on the right hand side are \( \mathcal{F}_t \)-measurable by [2] Corollary 8.2.13. Therefore, the multifunction \( \mathcal{G} \) is \( \mathcal{F}_t \)-measurable. It has non-empty images due to Definition 4.1 applied with \( \eta = E_t[u_t](\omega) \) and \( \zeta_{t-1} = \tilde{x}_{t-1}(\omega) \). Hence, an \( \mathcal{F}_t \)-measurable selection \( \bar{x} \) of \( \mathcal{G} \) exists (cf. [17]). From the construction of \( \mathcal{G} \),

\[
\| \bar{x}(\omega) - E_t[u_t](\omega) \| \leq C \left( d(f_t(\tilde{x}_{t-1}(\omega), E_t[u_t](\omega), \omega), Y_t(\omega)) + d(E_t[u_t](\omega), \mathcal{X}_t(\omega)) \right).
\]

(32)

We view both sides of this inequality as nonnegative elements of the space \( L_p(\Omega, \mathcal{F}_t, P) \). 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:

\[
\| \bar{x} - E_t[u_t] \| \leq C \left( \text{dist}(F_t(\tilde{x}_{t-1}, E_t[u_t]), Y_t) + \text{dist}(E_t[u_t], \mathcal{X}_t) \right).
\]

(33)

For every \( \bar{x} \in \tilde{\mathcal{X}}_t \), Jensen inequality implies that

\[
\| E_t[u_t] - E_t[\bar{x}] \| \leq \| E_t[u_t] - \bar{x} \|
\]

and \( E_t[\bar{x}] \in \mathcal{X}_t \) by convexity. Therefore, \( \text{dist}(E_t[u_t], \mathcal{X}_t) = \text{dist}(E_t[u_t], \tilde{X}_t) \). Using a similar argument, we have \( \text{dist}(F_t(\tilde{x}_{t-1}, E_t[u_t]), Y_t) = \text{dist}(F_t(\tilde{x}_{t-1}, E_t[u_t]), \tilde{Y}_t) \).
We observe that both distances above are finite because \( \|f_t(\bar{x}_{1:t-1}(\cdot), \mathbb{E}_t[u](\cdot), \cdot)\| \) has a finite \( \mathcal{L}_p \)-norm by virtue of (31) and the term \( d(\mathbb{E}_t[u](\cdot), X_0(\cdot)) \) is bounded from above by \( \|\mathbb{E}_t[u](\cdot) - \bar{z}(\cdot)\| \), which has a finite \( \mathcal{L}_p \)-norm by assumption.

Combining these observations with inequalities (33) and (31), we infer that

\[
\|\bar{z}_t - \mathbb{E}_t[u]\| \leq C \left( \text{dist}(F_t(u_{1:t}), \bar{Y}_t) + L(\|\bar{x}_{1:t-1} - u_{1:t-1}\| + \|u_{t} - \mathbb{E}_t[u]\|) + \text{dist}(\mathbb{E}_t[u], \bar{X}_t) \right).
\]

Since \( \text{dist}(\mathbb{E}_t[u], \bar{X}_t) \leq \text{dist}(u_{t}, \bar{X}_t) + \|u_{t} - \mathbb{E}_t[u]\| \), we conclude that

\[
\|\bar{z}_t - u_t\| \leq (1 + C + CL)\|u_{t} - \mathbb{E}_t[u]\| + C \left( \text{dist}(F_t(u_{1:t}), \bar{Y}_t) + L(\|\bar{x}_{1:t-1} - u_{1:t-1}\| + \text{dist}(u_{t}, \bar{X}_t) \right).
\]

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

\[
\|\bar{z}_t - u_t\| \leq \bar{C}_t \sum_{\ell=1}^{t} \left( \|u_{\ell} - \mathbb{E}_\ell[u]\| + \text{dist}(F_\ell(u_{1:\ell}), \bar{Y}_\ell) + \text{dist}(u_{\ell}, \bar{X}_\ell) \right).
\]

For \( t = 1 \), the result follows from (32), because the term \( \|\bar{x}_{1:t-1} - u_{1:t-1}\| \) is not present. Supposing it is true for \( t - 1 \), we verify it for \( t \) using (34). The last relation for \( t = T \) establishes the subregularity of the system (28)–(30).

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

\[
F_{t}^{'}(\bar{x}_{1:t}) = A_{t} = (A_{t,1}, \ldots, A_{t,T}), \quad t = 1, \ldots, T,
\]

referring to the partial Jacobians \( A_{t,\ell} : \tilde{X}_{t} \to \tilde{X}_{t} \), which are defined by (22)–(23), but are acting as linear operators between larger spaces.

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

**Theorem 4.5.** Suppose the system (17)–(18) admits complete subregular recourse. If a policy \( \hat{x} \) is a local minimum of problem (27)–(30) then a subgradient \( \bar{g}_t \in \partial \bar{\varphi}(\hat{x}) \), multipliers \( \lambda_t \in \tilde{X}_{t}^{*} \), \( \bar{\psi}_t \in N_{\tilde{X}_{t}}(F_t(\hat{x}_{1:t})) \), \( t = 1, \ldots, T \), and normal elements \( \bar{n}_t \in N_{\tilde{X}_{t}}(\hat{x}_t) \), \( t = 1, \ldots, T \), exist, such that for \( P \)-almost all \( \omega \in \Omega \) we have:

\[
\bar{g}_t + \lambda_t + \sum_{\ell=1}^{T} A_{t,\ell}^{'} \bar{\psi}_\ell + \bar{n}_t = 0, \quad t = 1, \ldots, T,
\]

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

**Proof.** We follow a similar line of argument as in Theorem 4.3 Using the Lipschitz continuity of \( \bar{\varphi}(\cdot) \) about \( \hat{x} \) with some Lipschitz constant \( L_{\bar{\varphi}} \), we infer that, for every \( K > L_{\bar{\varphi}} \), the point \( \hat{x} \) is a local minimum of the function

\[
\bar{\varphi}(x) + K \text{dist}(x, \tilde{X} \cap \tilde{F}^{-1}(\bar{Y}) \cap \mathcal{N}).
\]

We define the linear operator \( \Pi : \tilde{X} \to \tilde{X} \), by

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

(37)

Theorem 4.4 implies that the system (28)–(30) is metrically subregular with some constant \( \bar{C} \). Consequently, \( \hat{x} \) is a local minimum of the function

\[
\bar{\varphi}(x) + K\bar{C}(\text{dist}(F(x), \bar{Y}) + \text{dist}(x, \tilde{X}) + \|x - \Pi x\|).
\]
We use necessary conditions of optimality for Lipschitz continuous functions:

\[ 0 \in \partial \varphi(\hat{x}) +KC \partial_\psi \left[ \text{dist}(F(\hat{x}),\bar{Y}) \right] + KC \partial \left[ \text{dist}(\hat{x}, \bar{X}) \right] + KC \partial \| \hat{x} - \Pi \hat{x} \|, \]

By virtue of Theorem 3.3, the subdifferential of the function \( \text{dist}(F(\cdot),\bar{Y}) \) it is equal to \( \left[ F'(\hat{x}) \right]^* (N_{\bar{Y}}(F(\hat{x})) \cap \mathbb{B}_{\bar{Y}}) \). The subdifferential of \( \text{dist}(\hat{x}, \bar{X}) \) is \( N_{\bar{X}}(\hat{x}) \cap \mathbb{B}_{\bar{X}} \). The subdifferential of the last term is \( (I - \Pi^*) \mathbb{B}_{\bar{X}} \).

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,

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

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

\[ \hat{g} + \lambda + \left[ F'(\hat{x}) \right]^* \bar{\psi} + \hat{n} = 0. \]

The condition \( \lambda \in \ker(\Pi^*) \) is equivalent to \( 36 \). Equations \( 35 \) can now be derived as in the proof of Theorem 4.3 using the block-triangular form of \( A = F'(\hat{x}) \), and equation \( 25 \) for any \( h \in \bar{X} \). Since both spaces, \( \bar{X}^* \) and \( \bar{X}_T^* \), are defined with the use of the full \( \sigma \)-algebra \( \bar{F} \), we simply have \( A_{t,T}^* = A_{t,t}^T \). That is why no conditional expectation appears in \( 35 \). \[ \square \]

It may be of interest to explore the relations of two sets of optimality conditions of Theorems 3.3 and 4.3.

**Corollary 4.6.** The subgradient \( \hat{g} \in \partial \varphi(\hat{x}) \) given by \( \hat{g}_t = \mathbb{E}_t[\hat{g}_t], t = 1, \ldots, T \), together with the multipliers \( \bar{\psi}_t = \mathbb{E}_t[\bar{\psi}_t], t = 2, \ldots, T \), and normal vectors \( \hat{n}_t = \mathbb{E}_t[\hat{n}_t] \) satisfy the optimality conditions \( 24 \).

**Proof.** We take the conditional expectation of both sides of a typical relation in \( 35 \), first with respect to with respect to \( \bar{F} \). Since \( \mathbb{E}_t[\pi_t] = 0 \), using the tower property and \( \bar{F}_t \)-measurability of \( A_{t,T} \), we obtain

\[ 0 = \mathbb{E}_t[\hat{g}_t] + \mathbb{E}_t \left[ \sum_{t'=t}^T A_{t,t}^T \bar{\psi}_{t'} \right] + \mathbb{E}_t[\hat{n}_t] = \mathbb{E}_t[\hat{g}_t] + \mathbb{E}_t \left[ \sum_{t'=t}^T A_{t,t}^T \mathbb{E}_t[\bar{\psi}_{t'}] \right] + \mathbb{E}_t[\hat{n}_t]. \]

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

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

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

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

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

This proves that \( \hat{n}_t \in N_{\bar{X}_t}(\hat{x}_t), t = 1, \ldots, T \). In a similar way, we obtain \( \bar{\psi}_t \in N_{\bar{Y}_t}(F_t(\hat{x}_{t,t})) \) for \( t = 1, \ldots, T \). \[ \square \]
References

[1] J. Appell and P. P. Zabrejko. *Nonlinear Superposition Operators*. Cambridge University Press, 1990.

[2] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Springer Science & Business Media, 2009.

[3] S. Banach. *Théorie des Opérations Linéaires*. Monografje Matematyczne, 1932.

[4] J. V. Burke. Calmness and exact penalization. *SIAM Journal on control and optimization*, 29(2):493–497, 1991.

[5] C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions*. Springer, 1977.

[6] F. H. Clarke. *Optimization and Nonsmooth Analysis*, volume 5. SIAM, 1990.

[7] C. Corduneanu. *Functional Equations with Causal Operators*. CRC Press, 2005.

[8] N. D. Cuong and A. Y. Kruger. Nonlinear transversality of collections of sets: Primal space characterizations. *arXiv preprint arXiv:1902.06186*, 2019.

[9] A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings*. Springer, 2009.

[10] M. J. Eisner and P. Olsen. Duality for stochastic programming interpreted as lp in $l_p$-space. *SIAM Journal on Applied Mathematics*, 28(4):779–792, 1975.

[11] I. V. Evstigneev. Lagrange multipliers for the problems of stochastic programming. In M. Łoś, J. Łoś, and A. Wieczorek, editors, *Warsaw Fall Seminars in Mathematical Economics 1975*, pages 34–48. Springer, 1976.

[12] S. D. Flåm. Nonanticipativity in stochastic programming. *Journal of Optimization Theory and Applications*, 46(1):23–30, 1985.

[13] A. D. Ioffe. *Variational Analysis of Regular Mappings*. Springer, 2017.

[14] A D Ioffe. Necessary and sufficient conditions for a local minimum. 1: A reduction theorem and first order conditions. *SIAM Journal on Control and Optimization*, 17(2):245–250, 1979.

[15] D. Klatte and B. Kummer. Constrained minima and Lipschitzian penalties in metric spaces. *SIAM Journal on Optimization*, 13(2):619–633, 2002.

[16] A. Y. Kruger, D. R. Luke, and N. H. Thao. Set regularities and feasibility problems. *Mathematical Programming*, 168(1-2):279–311, 2018.

[17] K. Kuratowski and C. Ryll-Nardzewski. A general theorem on selectors. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 13(6):397–403, 1965.

[18] J. V. Outrata and W. Römisch. On optimality conditions for some nonsmooth optimization problems over lp spaces. *Journal of optimization theory and applications*, 126(2):411–438, 2005.

[19] R. R. Phelps. *Convex Functions, Monotone Operators and Differentiability*, volume 1364. Springer, 2009.
[20] R. Rockafellar. Integrals which are convex functionals. II. *Pacific Journal of Mathematics*, 39(2):439–469, 1971.

[21] R. T. Rockafellar and R. J.-B. Wets. Nonanticipativity and $l_1$-martingales in stochastic optimization problems. *Mathematical Programming Study, Stochastic Systems: Modeling, Identification and Optimization, II*, 6:170–187, 1976.

[22] R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: basic duality. *Pacific Journal of Mathematics*, 62(1):173–195, 1976.

[23] R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: relatively complete recourse and induced feasibility. *SIAM Journal on Control and Optimization*, 14(3):574–589, 1976.

[24] R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: singular multipliers and extended duality singular multipliers and duality. *Pacific Journal of Mathematics*, 62(2):507–522, 1976.

[25] A. Ruszczyński and A. Shapiro. Stochastic programming models. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, pages 1–64. Elsevier, 2003.

[26] R. J.-B. Wets. On the relation between stochastic and deterministic optimization. In *Control Theory, Numerical Methods and Computer Systems Modelling*, pages 350–361. Springer, 1975.