Journal of Convex Analysis
Volume 30 (2023), No. 3, 897–916

# Dynamic Programming For Data Independent Decision Sets

**Paul Dommel, Alois Pichler**
*Faculty of Mathematics, University of Technology, Chemnitz, Germany*
*alois.pichler@math.tu-chemnitz.de*

*Dedicated to Roger J-B Wets on the occasion of his 85th birthday.*

Received: May 30, 2022
Accepted: March 31.2023

Multistage stochastic optimization problems are oftentimes formulated informally in a pathwise way. These formulations are appealing in a discrete setting and suitable when addressing computational challenges, for example. But the pathwise problem statement does not allow an analysis with mathematical rigor and is therefore not appropriate.
R. T. Rockafellar and R. J.-B. Wets [*Nonanticipativity and $L^1$-martingales in stochastic optimization problems*, Math. Programming Study 6 (1976) 170–187] address the fundamental measurability concern of the value functions in the case of convex costs and constraints. This paper resumes these foundations. The contribution is a proof that there exist measurable versions of intermediate value functions, which reveals regularity in addition. Our proof builds on the Kolmogorov continuity theorem.
It is demonstrated that verification theorems allow stating traditional problem specifications in the novel setting with mathematical rigor. Further, we provide dynamic equations for the general problem. The problem classes covered include Markov decision processes, reinforcement learning and stochastic dual dynamic programming.

*Keywords:* Multistage stochastic optimization, stochastic processes, measurability.

*2010 Mathematics Subject Classification:* 90C08, 90C15, 60G07.

## 1. Introduction

Stochastic optimization problems frequently occur in finance, energy management and operations research where it is essential and of primary interest to develop efficient algorithms and to provide access to fast decisions. Many of these algorithms build on finite models in discrete space. Multistage stochastic problems build on stochastic processes in discrete time or on decision trees, cf. Philpott et al. [29], Girardeau et al. [13], or Maggioni and Pflug [20] among many others.

This paper aims at presenting a rigorous mathematical framework for multistage stochastic optimization problems in general, finite or continuous state spaces. It exploits measurability systematically in conditional expectations and involves the proper conditional infimum. We develop value processes and show their relation to the genuine stochastic optimization problem. Building on the Kolmogorov continuity theorem, our central result ensures adequate regularity of the inherent value functions.

Supported by DFG, German Research Foundation, Project-ID 416228727, SFB 1410.

ISSN 0944-6532 / $ 2.50 

Multistage stochastic optimization involves optimization based on partial realizations, which are partially observed trajectories. It is a major difficulty in multistage stochastic optimization that individual realizations or trajectories have probability zero. But the problems are stated naturally in this pathwise way. It is hence essential to avoid difficulties that arise with this pathwise, or $\omega$-by-$\omega$ considerations and to address measurability carefully.

The conditional expectation, the conditional probability and the conditional infimum constitute main and major difficulties in multistage stochastic optimization. Early publications capture measurability (cf. Rockafellar and Wets [33, 35], Rockafellar [32], Olsen [21, 22, 23], Evstigneev [8] or also Bertsekas and Shreve [4]) in a dynamic context. The infimum in the optimization formulation and the conditional expectations need to be interchanged at subsequent stages to exploit computational advantages, cf. Carpentier et al. [5] or Pflug and Pichler [27]. Indeed, a recourse decision is based on a partial realization of a stochastic outcome, but has to be considered already at the very beginning of decision-making. Considering every outcome separately (i.e., $\omega$-by-$\omega$) is only possible for finite states, so that a tree describes the evolution of the stochastic process and the evolution of the decision process as well. In a multistage environment, however, the computational burden grows exponentially with the branching structure and this approach thus is clearly not advisable. The catchphrase *curse of dimensionality* is often associated with this phenomenon in multistage stochastic optimization.

The central result presented is a proof that there exists a measurable version of the intermediate value process which also provides regularity, which is not present in preceding publications. We present dynamic equations even for the general, non-Markovian setting. The general verification theorems presented are characterizations as martingales.

We elaborate the theory in full generality and elaborate on problem settings, which are of particular importance in applications and increasingly popular in stochastic optimization. They include dynamic programming (the references [3] and [9] include considerations on measurability as well), stochastic dual dynamic programming, the Bellman principle and reinforcement learning, which has grown to outstanding importance in machine learning or data science. For a recent tutorial including also computational aspects we refer to Shapiro [41].

Investigations on foundations have been started by Pichler and Shapiro [31] with a focus on the *distributionally robust* aspect of multistage stochastic optimization and also in Zhen et al. [46] and Pichler et al. [30]. This paper enhances, complements and continues these investigations on foundations, but now addressing the genuine problem statement itself.

An important special case of multistage stochastic optimization, as it is presented in this paper, is dynamic optimization. Dedicated algorithms have been developed for this special case and papers as Lan and Zhou [18] address convergence of dynamic stochastic approximation, e.g., Caprpentier et al. [6] collect recent theoretical results for the special case of dynamic optimization again.

Applications of multistage stochastic optimization are widespread over many economic and managerial disciplines. We pick Löhndorf et al. [19] to represent and demonstrate importance of multistage stochastic optimization for example in energy and Shapiro et al. [43] to exemplify computational limitations and Ruszczyński [37] to point to extensions involving risk.

**Outline.** We address the general multistage problem formulation in Section 3, after introducing the informal description and the mathematical setting. An essential component to manage the evolution of the underlying stochastic process and the decisions is the value process, introduced in Section 4. Particular situations as dynamic problems, additive cost functions, Markovian processes and SDDP (stochastic dual dynamic programming) appear frequently in applications. Important simplifications, dedicated complexity and convergence issues are essential to solve these problems. We address these particular problem formulations in Section 5.

## 2. Mathematical setting

Stochastic optimization builds on random variables, while multistage stochastic optimization generalizes the problem setting to stochastic processes on adequate probability spaces. In what follows we address the informal, pathwise setting and prepare the mathematical stage to discuss the optimization problem with mathematical rigor.

### 2.1. Informal description

The multistage optimization problem, as it is the basis for implementations, is

$$\inf_{u_0} \mathbb{E}_{X_1} \dots \mathbb{E}_{X_t} \inf_{u_t} \underbrace{\mathbb{E}_{X_{t+1}} \underbrace{\inf_{u_{t+1}} \dots \mathbb{E}_{X_T} \inf_{u_T} v(X_1, \dots, X_T, u_0, \dots u_T)}_{v_t(x_{1:t}, u_{0:t})}}_{V_t(x_{1:t}, u_{0:t-1})}. \tag{1}$$

Here, $v$ is the random objective of the optimization problem, $(X_1, \dots, X_T)$ are the consecutive random observations and $u_t$ the decisions made after each partial realization $X_t$ at each stage $t$. The functions $v_t$ and $V_t$ are the intermediate value functions, which are given intuitively in (1) in an $\omega$-by-$\omega$ or pathwise context, which is often comfortable in a discrete framework; Definition 4.1 below provides the rigorous definitions.

The problem statement (1) exhibits the following difficulties:

(i) The expectation at stage $t$ is a conditional expectation ($\mathbb{E}_{X_t}$), conditional on the preceding observations $(X_1, \dots, X_t)$. On a continuous probability space, this trajectory has probability 0 and the conditional expectation must not be considered in a pathwise specification as (1) does.

(ii) The infimum with respect to $u_t$ at stage $t$ depends on preceding observations. As above, this is a conditional infimum and further, the infimum of uncountable many functions is not measurable a priori.

(iii) The intermediate value functions $v_t$ and $V_t$ aggregate the entire future. As functions, defined on observed partial realizations, they are not necessarily measurable.

Nonetheless, the work instruction (1) provides a straightforward illustration of the optimization problem, indicating the progression of successive optimization and random realizations. While (i) and (ii) are fixed with standard means, interchanging the infimum with expectations requires clarification. The issue (iii) emerges specifically in multistage optimization. We resolve this problem with the help of Kolmogorov's continuity theorem.

In what follows we provide a rigorous mathematical problem statement of (1) first and then discuss derived variants.

### 2.2. Mathematical exposition

Let $(\Omega, \mathcal{F}, P)$ be a probability space. We may refer to Kallenberg [15, Lemma 1.13] or Shriyaev [44, Theorem II.4.3] for the following Doob-Dynkin lemma.

**Lemma 2.1.** (Doob-Dynkin) *Suppose the random variable $U$ with values in $\mathbb{R}^t$ is measurable with respect to the $\sigma$-algebra $\sigma(X)$ generated by the random variable $X$ with values in $\mathbb{R}^d$. Then there is a (Borel-) measurable function $\varphi\colon \mathbb{R}^d \to \mathbb{R}^t$ so that*

$$U = \varphi \circ X.$$

The essential infimum of a set of random variables is defined in Dunford and Schwartz [7]. We want to highlight Föllmer and Schied [12, Appendix A.5] for a compelling proof regarding existence and mention that there is a relation to measurable majorants, cf. van der Vaart and Wellner [45, Section 1.2].

**Definition 2.2.** (Essential infimum) Let $\mathcal{U}$ be a family of $\mathbb{R}$-valued random variables. The random variable $Y$ is the *essential infimum* of $\mathcal{U}$, if

(i) $Y \le U$ a.e. for all $U \in \mathcal{U}$ and

(ii) whenever $Z \le U$ a.e. for all $U \in \mathcal{U}$, then $Z \le Y$ a.e.

We shall write $\operatorname{ess\,inf}_{U\in\mathcal{U}} U := Y$ for the essential infimum of $\mathcal{U}$.

Further, for a $\sigma$-algebra $\mathcal{G} \subset \mathcal{F}$ we shall write $Y = \operatorname{ess\,inf}_{U\in\mathcal{U}} (U|\, \mathcal{G})$, if

(iii) $Y$ and $Z$ are measurable with respect to $\mathcal{G}$ in (i) and (ii), in addition.

**Remark 2.3.** Provided that there is any measurable lower bound $X$ satisfying (i), the essential infimum exists and is unique, cf. Föllmer and Schied [12, Appendix A.5] or Karathas and Shreve [16, Appendix A]. If $\mathcal{U}$ is closed under pairwise minimization[1], i.e, $\min(U, V) \in \mathcal{U}$ for $U, V \in \mathcal{U}$, then there is a nonincreasing sequence $U_n \in \mathcal{U}$ such that $U_n \to \operatorname{ess\,inf}_{U\in\mathcal{U}} U$ a.s., as $n \to \infty$.

For $X$ measurable, the random variable $\operatorname{ess\,inf}_{U\in\mathcal{U}} \big(U|\, \sigma(X)\big)$ is measurable with respect to $\sigma(X)$, the sigma algebra generated by $X$. By the Doob-Dynkin lemma (Lemma 3.4) there is a measurable $\varphi(\cdot)$ so that $\operatorname{ess\,inf}_{U\in\mathcal{U}} \big(U|\, \sigma(X)\big) = \varphi(X)$. We shall denote this function by $\operatorname{ess\,inf}_{U\in\mathcal{U}} (U|\, X) := \varphi$.

**Remark 2.4.** We shall also address the conditional essential infimum for a singleton $\mathcal{U} = \{U\}$. In this case, the random variable $\operatorname{ess\,inf}_{U\in\mathcal{U}} (U|\, X)$ is the $\sigma(X)$ measurable envelope of $U$ for which we shall write $\operatorname{ess\,inf} (U|\, X)$.

**Remark 2.5.** (Caveat) The term essential infimum is occasionally also used for the largest *number* $c \in \mathbb{R}$ smaller than the random variable $X$, $c \le X$ a.s. This is $\operatorname{ess\,inf}_{U\in\mathcal{U}} (U|\, \{\emptyset, \Omega\})$ in the notation introduced above, where $\{\emptyset, \Omega\}$ is the trivial sigma algebra.

[1] The set $\mathcal{U}$ is said to be *directed downwards* in Föllmer and Schied [12].

### 2.3. Functional optimization

The prevailing perspective in practice of multistage stochastic optimization is not a measure theoretic perspective but rather a functional view: we shall develop and address this perspective as the *informal*, *$\omega$-by-$\omega$* or *pathwise description.* Throughout, we will give the stochastic process perspective first and then complement the informal perspective as well. While the first one provides expressions with mathematical rigor, the latter, intuitive problem statement is perhaps better to understand, well-established and more practical for concrete numerical implementations. This is essential for both the governing stochastic process and the decision process.

Stochastic optimization generally involves an observation $x \in \mathbb{R}^d$, an associated control $u(x)$ as well as assigned costs $v(x, u(x))$. Although this setting is clear from an informal perspective, it needs to be treated carefully from a mathematical point of view. Indeed, assuming that $u$ and $v$ are (Lebesque-)measurable functions is mathematically not *sufficient* for the problem of interest. This is due to the fact that the composition $v(x, u(x))$ is not necessarily measurable, even for measurable $v$ and $u$. To overcome this issue, we restrict the class of possible value functions $v$ to the following class.

**Definition 2.6.** A function $v\colon \mathbb{R}^d \times \mathbb{R}^t \to \mathbb{R}$ *is a Carathéodory function if*

(i) $x \mapsto v(x, u)$ is measurable for every $u \in \mathbb{R}^t$ and

(ii) $u \mapsto v(x, u)$ is continuous on $\mathbb{R}^t$ for each $x \in \mathbb{R}^d$.

If $v$ is a *Carathéodory* function and $u$ measurable, the composition $v(x, u(x))$ is measurable. This is an immediate consequence of the following condition on joint measurability from Gowrisankaran [14, Theorem 2]; we state the result in full mathematical beauty, although we do not need this most general variant.

**Theorem 2.7.** *Let $(X, \tau)$ be a measurable space and $Y$ a Suslin space. Let $\mathcal{B}$ be the Borel $\sigma$-algebra of all measurable subsets for a locally finite measure $\lambda$ on the Borel $\sigma$-algebra of $Y$. Then, a function $v\colon X \times Y \to A$ with values in a separable metrizable space $A$ with*

(i) *$x \mapsto v(x, y)$ is $\tau$-measurable for every $y \in Y$ and*

(ii) *$y \mapsto v(x, y)$ is continuous on $Y$ for each $x \in X$*

*is $\tau \otimes \mathcal{B}$-measurable on $X \times Y$.*

**Remark 2.8.** Pennanen and Perkkiö [24, 25] elaborate measurability of the important special case, where the cost function $v$ is convex. To resolve the issue above, i.e., the super positional measurability of $v(x, u(x))$, they assume that $v$ as a *normal integrand* (see Rockafellar [32, Theorem 2.A]). This is a slightly weaker assumption than the Carathéodory property of Definition 2.6 (Pennanen and Perkkiö [25, Exampel 1.10]). However, in the setting of this paper, where $v$ takes only finite values, both properties are equivalent (cf. Rockafellar [32, Proposition 2C]).

We select the controls $u$ from a decision set $\mathcal{U}$ of measurable functions. Our analysis partially builds on decomposable decisions as defined below.

**Definition 2.9.** (Decomposable functions) Let $\sigma(\mathcal{U}) := \sigma(u\colon u \in \mathcal{U})$ be the sigma algebra generated by the functions $u\colon \mathbb{R}^t \to \mathbb{R}^d$ contained in $\mathcal{U}$. We shall say that the class of functions $\mathcal{U}$ is *decomposable*, if $u_A \in \mathcal{U}$, where

$$u_A(x) := \begin{cases} u_1(x) & \text{if } x \in A, \\ u_2(x) & \text{else} \end{cases}$$

whenever $A \in \sigma(\mathcal{U})$ and $u_1,\, u_2 \in \mathcal{U}$.

Traditional formulations of the interchangeability principle require that the infimum is measurable (cf. Shapiro [40] or normal integrands in Rockafellar and Wets [35, Theorem 14.60] or Rockafellar and Wets [34]). By involving the essential infimum, the following proposition establishes the interchangeability principle without requesting measurability explicitly.

**Proposition 2.10.** (Interchangeability principle) *Let $\mathcal{U}$ be a class of measurable functions, let $v\colon \mathbb{R}^t \times \mathbb{R}^d \to \mathbb{R}$ be a (measurable) function bounded from below and $X\colon \Omega \to \mathbb{R}^t$ a random variable. It holds that*

$$\mathbb{E} \operatorname*{ess\,inf}_{u\in\mathcal{U}} v\big(X, u(X)\big) \le \inf_{u\in\mathcal{U}} \mathbb{E}\, v\big(X, u(X)\big). \tag{2}$$

*Equality holds in* (2) *if $\mathcal{U}$ is decomposable and $X$ is measurable with respect to $\sigma(\mathcal{U})$.*

**Proof.** For every $x$ we have that $\inf_{u\in\mathcal{U}} v\big(x, u(x)\big) \le v\big(x, u(x)\big)$ and thus

$$\operatorname*{ess\,inf}_{u\in\mathcal{U}} v\big(X, u(X)\big) \le v\big(X, u(X)\big) \quad \text{a.e.}$$

Taking expectations first and then the infimum reveals (2).

For the remaining assertion recall from Remark 2.3 (or Karatzas and Shreve [16, Appendix A]) that there is a sequence $u_j$ so that

$$\min_{j=1,\dots,n} v\big(X, u_j(X)\big) \to \operatorname*{ess\,inf}_{u\in\mathcal{U}} v\big(X, u(X)\big)$$

almost surely, as $n \to \infty$. Define

$$A_i := \Big\{x \in \mathbb{R}^t\colon v\big(x, u_i(x)\big) = \min_{j=1,\dots,n} v\big(x, u_j(x)\big)\Big\}, \qquad \tilde{A}_i := A_i \backslash \bigcup_{j<i} A_j$$

and set $\tilde{u}_n := \sum_{i=1}^n u_i \cdot \mathbb{1}_{\tilde{A}_i}$. As $\mathcal{U}$ is decomposable we have that $\tilde{u}_n \in \mathcal{U}$ and $v\big(x, \tilde{u}_n(x)\big) = \min_{i=1,\dots,n} v\big(x, u_i(x)\big)$. Employing Beppo Levi's monotone convergence theorem we conclude $\mathbb{E}\, v\big(X, \tilde{u}_n(X)\big) \to \mathbb{E} \operatorname{ess\,inf}_{u\in\mathcal{U}} v\big(X, u(X)\big)$ as $n \to \infty$ and hence the assertion. □

**Proposition 2.11.** *Suppose that $u \mapsto v(x, u)$ is non-decreasing for every $x$, i.e., $v(x, u_1) \le v(x, u_2)$ whenever $u_1 \le u_2$ in every component and* $\min(u_1, u_2) \in \mathcal{U}$ *for $u_1,\, u_2 \in \mathcal{U}$. Then interchangeability* (2) *holds with equality.*

**Proof.** By monotonicity of $v$ we have with $u := \min_{i=1,\dots,n} u_i \in \mathcal{U}$ that

$$\min_{j=1,\dots,n} v\big(X, u_j(X)\big) = v\Big(X, \min_{j=1,\dots,n} u_j(X)\Big) = v\big(X, u(X)\big).$$

The assertion follows along the proof of Proposition 2.10. □

## 3. General multistage optimization problems

The general multistage optimization problem involves a stochastic process instead of a simple random variable. Let $X = (X_1, \dots, X_t)$ be a stochastic process with stages $t = 1, \dots, T$ and, without loss of generality, with marginals $X_t \in \mathbb{R}$. For convenience, the stochastic process $X$ is occasionally also augmented with a deterministic starting value $X_0 = x_0$ a.s. so that $X = (X_0, X_1, \dots, X_T)$.

**Definition 3.1.** (Nonanticipativity) The stochastic process $U = (U_0, \dots, U_T)$ is *adapted* to $X$, if $U_t$ is measurable with respect to $\sigma(X_0, \dots, X_t)$ for every stage $t = 0, \dots, T$. We shall write $U \lhd X$, if $U$ is adapted to $X$.

In stochastic optimization, the synonymous term *nonanticipative* is more common than adapted.

**Definition 3.2.** (The natural filtration) The stochastic process $X = (X_0, \dots, X_T)$ is adapted to the *natural filtration*, if we have $X_t(\omega) = X_t(\omega_1, \dots, \omega_t)$ (that is, $X_t(\omega) = \tilde{X}_t(\omega_1, \dots, \omega_t)$ for some random variable $\tilde{X}$ which we identify with $X_t$).

Multistage stochastic optimization considers classes $\mathcal{U}$ of stochastic control processes. To not run into difficulties regarding a governing measure we assume that there is a control $U_0$ such that

$$U \lhd U_0 \text{ (nonanticipative) for all } U \in \mathcal{U}.$$

A particular situation arises for the class $\mathcal{U}$ of stochastic processes adapted to $X$, $\mathcal{U} \subset \{U : U \lhd X\}$. In this case one may choose $U_0 = X$ as governing process.

We consider the following, general multistage stochastic optimization problem.

**Definition 3.3.** (Multistage optimization problem) Let

$$v : \mathbb{R}^{T+1} \times \mathbb{R}^{T+1} \to \mathbb{R}, \qquad (x, u) \mapsto v(x, u) \tag{3}$$

be a *Carathéodory* function. For a class $\mathcal{U}$ of feasible controls, the general multistage stochastic optimization problem is

$$\inf_{U \in \mathcal{U}} \mathbb{E}\, v(X, U), \tag{4}$$

where the infimum is among all feasible control policies $U \in \mathcal{U}$ adapted to $X$ for which the expectation exists. The function $v$ is the (stochastic) *objective function* and the set $\mathcal{U}$ is the set of *admissible controls*, *decisions* or *policies*. Note that the decision space is $\mathbb{R}^{T+1}$ in (3), that is, at each stage $t \in \{0, \dots, T\}$ a decision in $\mathbb{R}$ is made; this setting is chosen for convenience of presentation.

In what follows we shall assume that the infimum in (4) is finite. A somewhat stronger assumption, although not necessary, is that $v$ is uniformly bounded from below (i.e., $v \geq C > -\infty$) so that the expectation in (4) is well-defined for every $U \in \mathcal{U}$.

### 3.1. Equivalent problem statements

For $u_t(x_1, \dots, x_t)$ measurable it is evident that $u_t(X_1, \dots, X_t)$ is measurable with respect to $\sigma(X_1, \dots, X_t)$. For this,

$$U := u(X_1, \dots, X_T) \tag{5}$$

is a nonanticipative process with respect to $X$, provided that

$$u(x_1, \dots, x_T) = \begin{pmatrix} u_0 \\ u_1(x_1) \\ \vdots \\ u_T(x_1, \dots, x_T) \end{pmatrix}. \tag{6}$$

The Doob-Dynkin lemma (Lemma 2.1) ensures that every process $U \in \mathcal{U}$ adapted to $X$ has the particular form (5) with (6).

**Lemma 3.4.** (Doob-Dynkin lemma, extended) *Let $X = (X_0, \dots, X_T)$ be a stochastic process in discrete time with marginal states $X_t \in \mathbb{R}^d$ and $U \lhd X$. There are measurable functions $u_t$ so that $U_t = u_t(X_1, \dots, X_t)$ for $t = 0, \dots, T$ a.s. and $U = \varphi_U \circ X$, where $\varphi_U = u$ is given by* (6).

**Functional optimization perspective.** The optimization problem (4) employs a fixed stochastic process $X$. In view of the Doob-Dynkin lemma, the problem (4) thus can be stated as an optimization problem among stochastic processes, or equivalently also as optimization problem among functions, each of the specific form (6). The multistage stochastic optimization problem thus can be classified as a *functional optimization problem*, because solving it means finding unknown *functions* as (6).

The equivalence between measurable functions and processes is given by $U \mapsto \varphi_U$, where $\varphi_U$ is the function from the extended Doob-Dynkin lemma (Lemma 3.4), while the inverse is the map

$$u \mapsto U = u(X)$$

given in (5). Further, this equivalence allows extending the notion of decomposable to stochastic processes.

**Definition 3.5.** (Decomposable processes) The class $\mathcal{U}$ of stochastic processes is decomposable, if each function in $\{\varphi_U \colon U \in \mathcal{U}\}$ is decomposable in the sense of Definition 2.9, where $\varphi_U$ is given in Lemma 3.4.

### 3.2. Special cases of the general problem setting

The conventional stochastic optimization problem and the stochastic optimization problem with recourse are special cases of the multistage stochastic optimization problem.

**Example 3.6.** ($T = 0$) Consider the set of policies with

$$\mathcal{U} \subset \{U \colon U_t \lhd X_0 \text{ for all } t \geq 0\}$$

(or $T = 0$), so that each component $u_t$ is deterministic, i.e., nonrandom.

The corresponding optimization problem

$$\inf_{u\in\mathbb{R}^{T+1}} \mathbb{E}\, v(X,u) \tag{7}$$

is a conventional stochastic optimization problem, as it is sufficient to treat $X$ as a random vector in (7). Here, it is not essential that $X$ is a stochastic process, the time component is suppressed.

**Example 3.7.** ($T=1$) Consider the feasible policies

$$\mathcal{U} \subset \{u\colon u_0 \lhd X_0 \text{ and } u_t \lhd X_1 \text{ for all } t\ge 1\}$$

(or $T=1$). With (6), the problem simplifies to

$$\inf_{(u_0,u_1(X))\in\mathcal{U}} \mathbb{E}\, v\big(X_1,u_0,u_1(X_1)\big).$$

Here, the decision $u_0$ is deterministic, i.e., does not depend on the random components of $X$; $u_1(\cdot)$ is called the *random recourse decision* in the literature (cf. Shapiro et al. [42]).

## 4. The value process

It is an important conceptual element in stochastic optimization to consider the problem sequentially in time, so that any new observation $X_t$ triggers a subsequent new decision $u_t(X_1\dots,X_t)$, which itself is based on the past. Shapiro [39] depicts the consecutive transitions via the chain in Figure 4.1. The transitions can be started with $X_0$ equally well.

$$u_0 \rightsquigarrow X_1 \rightsquigarrow u_1 \rightsquigarrow \cdots \rightsquigarrow X_t \rightsquigarrow u_t \rightsquigarrow X_{t+1} \rightsquigarrow \cdots \rightsquigarrow X_T \rightsquigarrow u_T$$

Figure 4.1: The progression of random observations and decisions

In what follows we develop a similar decomposition of the optimization problem (4) and present our main result in Theorem 4.2 below. For notational convenience we introduce the abbreviation $x_{t:t'} := (x_t, x_{t+1},\dots,x_{t'})$ ($0\le t,\ t'\le T$) for subvectors. We also write $X_{:t} := (X_0,\dots,X_t)$ for the initial and $U_{t:} := (U_t,\dots,U_T)$ for the final (trailing) substrings. Recall that $U$ is a non-anticipative process if there is a control $u$ so that $U = u(X_1,\dots,X_T)$, as well as the functions $u_{t:}$ with $U_{t:} = u_{t:}(X)$. By $\mathcal{U}_{t:} = \{u_{t:}\colon U\in\mathcal{U}\}$ we denote the set of functions including the final decisions of all control processes.

### 4.1. Existence of the intermediate value functions

A common way to solve the initial problem (4) is to decompose it into a sequence of subproblems. We specify these subproblems by introducing the value process in the following considerations. Let $u_{:t}\in\mathbb{R}^{t+1}$ and a function $\tilde u_{t+1:T}\in\mathcal{U}_{t+1:T}$ be given. As a consequence of the Doob-Dynkin lemma (Lemma 2.1) there is measurable mapping $v_{t,u_{:t}}^{\tilde u_{t+1:T}}\colon \mathbb{R}^{t+1}\to\mathbb{R}$ such that

$$v_{t,u_{:t}}^{\tilde u_{t+1:T}}(X_{:t}) = \mathbb{E}\big(v\left(X,u_{:t},\tilde u_{t+1:T}(X)\right)\big|\, X_{:t}\big). \tag{8}$$

These conditional expectations constitute the building block for the intermediate value functions.

**Definition 4.1.** The *(intermediate) value functions* are, where $t = 0, \dots, T$,

$$v_t(x_{:t}, u_{:t}) := \operatorname*{ess\,inf}_{\tilde{u}_{t+1:T} \in \mathcal{U}_{t+1:T}} v_{t,u_{:t}}^{\tilde{u}_{t+1:T}}(x_{:t}) \text{ and} \tag{9}$$

$$V_t(x_{:t}, u_{:t-1}) := \operatorname*{ess\,inf}_{\tilde{u}_t \in \mathcal{U}_t} v_t\big(x_{:t}, (u_{:t-1}, \tilde{u}_t(x_{:t}))\big). \tag{10}$$

These value functions are functions on $\mathbb{R}^{(t+1)\times(t+1)}$ ($\mathbb{R}^{(t+1)\times t}$, resp.) and the essential infima are with respect to these spaces. These functions are generally not unique as there are multiple functions satisfying the Doob-Dynkin lemma. The value functions $V_t$ and $v_t$ are defined pointwise (and well-defined on each point), but they are not necessarily measurable. Hence, additional conditions on $v$ need to be imposed to ensure measurability.

The following statement establishes existence of a measurable version of the intermediate value functions and reveals additional regularity. The proof builds on *Kolmogorov's continuity theorem*, also known as Kolmogorov-Chentsov theorem.

**Theorem 4.2.** (Existence of a measurable version of the value function) *Assume that $v$ is locally Hölder continuous with exponent $\alpha > 0$ in $u$, i.e.,*

$$|v(x, u_1) - v(x, u_2)| \le C \,\|u_1 - u_2\|^\alpha \qquad \text{for } x \in \mathbb{R}^t \text{ and } \|u_1 - u_2\| \le \delta, \tag{11}$$

*where $\delta > 0$ is sufficiently small. Then there exists a version of $v_t$ of the intermediate value function which is measurable with respect to $\mathcal{B}(\mathbb{R}^{t+1}) \otimes \mathcal{B}(\mathbb{R}^{t+1})$ and locally Hölder continuous with exponent $\tilde{\alpha} \in \big(0, \frac{\alpha}{t+2}\big)$.*

**Proof.** We shall employ Theorem 2.7. Consider the function

$$v_t^{\tilde{u}_{t+1:T}}(x_{:t}, u_{:t}) := v_{t,u_{:t}}^{\tilde{u}_{t+1:T}}(x_{:t})$$

(cf. (8)), where $\tilde{u}_{t+1:T} \in \mathcal{U}_{t+1:T}$ is fixed. Measurability follows from the definition of the function $v_t$ in (9) and general measurability of the essential infimum and thus the condition (i) of Theorem 2.7.

In order to employ Theorem 2.7 we need to show continuity (i.e., (ii)) of $u_{:t} \mapsto v_t^{\tilde{u}_{t+1:T}}(x_{:t}, u_{:t})$. To this end, consider the stochastic process $(Z_u)_{u \in \mathbb{R}^{t+1}}$, indexed by $u \in \mathbb{R}^{t+1}$ and defined by

$$Z_u := \mathbb{E}\big(v\,(X, u, \tilde{u}_{t+1:T}(X)) \big|\, X_{:t}\big).$$

Further, let $u_0 \in \mathbb{R}$ and $u \in U_\delta(u_0)$ for $\delta > 0$ sufficiently small be given. Set $\tilde{\alpha} := \frac{t+2}{\alpha}$, $\beta := 1$, and by employing the Hölder condition (11) we have that

$$\begin{aligned}
\mathbb{E}\,|Z_u - Z_{u_0}|^{\tilde{\alpha}} &= \mathbb{E}\,\big|\,\mathbb{E}\,(v\,(X, u, \tilde{u}_{t+1:T}(X)) - v\,(X, u_0, \tilde{u}_{t+1:T}(X))|\, X_{:t})\,\big|^{\tilde{\alpha}} \\
&\le \mathbb{E}\,(C \cdot \|u - u_0\|^\alpha)^{\tilde{\alpha}} \\
&\le C\,\|u - u_0\|^{t+2} = C\,\|u - u_0\|^{t+1+\beta}
\end{aligned}$$

for some $C < \infty$, independent of $\tilde{u}_{t+1:T}$.

Hence, by the Kolmogorov continuity theorem (cf. Klenke [17, p. 453]), there is a process $Z_u$ such that $Z_u = Z(\cdot, u) = \mathbb{E}\left(v\left(X, u, \tilde{u}_{t+1:T}(X)\right) | X_{:t}\right)$ and $Z(\omega, \cdot)$ is Hölder continuous with exponent $\frac{\beta}{\tilde{\alpha}} = \frac{\alpha}{t+2}$ for almost every $\omega \in \Omega$. It follows that the corresponding functions $v_t^{\tilde{u}_{t+1:T}}$ are continuous with respect to $u$. This proves (ii) and hence verifies the measurability of $v_t$. To see Hölder continuity of $v_t$, note that the Hölder constants $\alpha, C$ of $v_t^{\tilde{u}_{t+1:T}}$ do not depend on the underlying policy $\tilde{u}_{t+1:T}$. Thus, the essential infimum retains the Hölder property, which is the assertion. □

**Remark 4.3.** The Hölder condition (11) on $v$ is crucial for two different arguments in the proof above. It primarily serves as a prerequisite for the Kolmogorov-Chentsov theorem, which provides continuous versions of the intermediate functions $v_t^{\tilde{u}_{t+1:T}}$. These versions ensure the required measurability of the essential infimum $v_t$. The condition is further crucial for the regularity of $v_t$. Uniformly bounding the Hölder constants of $v_t^{\tilde{u}_{t+1:T}}$, guarantees Hölder continuity of the essential infimum. Without this uniform bound, $v_t$ might not even be continuous.

**Remark 4.4.** (Lipschitz continuity) It is evident that measurable versions of (9) and (10) exist for uniformly Lipschitz continuous objective functions $v$.

**Remark 4.5.** The $\omega$-by-$\omega$ setting in (1) describes the setting for discrete probability measures. In this case, the essential infimum is the usual infimum and the conditional expectation deduces from elementary events, each with strictly positive probability. The value functions in Definition 4.1 generalize this discrete setting.

### 4.2. The value process

In what follows, we define the value processes substituting $x_{:t}$ and $u_{:t}$ by their stochastic counterparts $X_{:t}$ and $U_{:t}$.

**Definition 4.6.** (Value process) Assume $v$ satisfies the Hölder condition imposed in Theorem 4.2 and $U \in \mathcal{U}$ is a nonanticipative stochastic process ($U \lhd X$). The *general value processes* are

$$\boldsymbol{v}_t^U := v_t(X_{:t}, U_{:t}) \text{ and} \tag{12}$$

$$\boldsymbol{V}_t^U := V_t(X_{:t}, U_{:t-1}), \tag{13}$$

where $v_t$ and $V_t$ are the intermediate value functions, cf. Definition 4.1.

**Remark 4.7.** The functions $v_t$ and $V_t$ (cf. (9) and (10)) are defined on

$$V_t \colon \mathbb{R}^{t+1} \times \mathbb{R}^t \to \mathbb{R} \text{ and } v_t \colon \mathbb{R}^{t+1} \times \mathbb{R}^{t+1} \to \mathbb{R}.$$

We employ bold letters to indicate random variables, i.e., functions on $\Omega$ given by

$$\boldsymbol{v}_t^U(\omega) = v_t\big(X_{:t}(\omega), U_{:t}(\omega)\big) \text{ and}$$
$$\boldsymbol{V}_t^U(\omega) = V_t\big(X_{:t}(\omega), U_{:t-1}(\omega)\big).$$

Figure 4.2 depicts the domain and the range of these functions and random variables.

$$\begin{tikzcd}
\Omega \arrow[rr, "\boldsymbol{v}_t"] \arrow[dr, "{(X,U)}"] \arrow[ddr, "{(X_{:t},U_{:t})}"'] & & \mathbb{R} \\
& \mathbb{R}^{T+1}\times\mathbb{R}^{T+1} \arrow[ur, "v"'] \arrow[d] & \\
& \mathbb{R}^{t+1}\times\mathbb{R}^{t+1} \arrow[uur, "{v_t \leq \mathbb{E}\left( v(X,U)\middle|\, X_{:t}, U_{:t}\right)}"'] &
\end{tikzcd}$$

Figure 4.2: Domain and range of the objective and the general value process

**Remark 4.8.** (Pathwise, or $\omega$-by-$\omega$ description) The functions $V_t$ and $v_t$, which describe the value processes (12) and (13) can be given explicitly and directly – but intuitively – as

$$V_t(x_{:t}, u_{:t-1}) = \inf_{u_{t:}(\cdot)} \mathbb{E}\left( v\big(X_{:T}, u_{:t-1}, u_{t:T}(X_{:T})\big)\middle|\, X_{:t} = x_{:t}, U_{:t} = u_{:t-1}\right) \tag{14}$$

and

$$v_t(x_{:t}, u_{:t}) = \inf_{u_{t+1:}(\cdot)} \mathbb{E}\left( v\big(X_{:T}, u_{:t}, u_{t+1:T}(X_{:T})\big)\middle|\, X_{:t} = x_{:t}, U_{:t} = u_{:t}\right), \tag{15}$$

where the infima are among functions

$$u_{t:}(x_1,\dots,x_T) = \begin{pmatrix} u_t(x_1,\dots,x_t) \\ \vdots \\ u_T(x_1,\dots,x_t,\dots,x_T) \end{pmatrix}$$

with $u(X) \in \mathcal{U}$.

Note, however, that the expressions (14) and (15) are not necessarily well-defined, as they may depend explicitly on the choice of the control process $U_{:t}$. They further face a delicate measurability problem, as the pointwise infimum is not measurable, in general. Hence, (14) and (15) can*not* be used as definitions. Our definitions (9) and (10), together with (12) and (13), resolve this problem by addressing $u_{:t}$ as a parameter and passing over to the essential infimum, which has a measurable version by the main theorem, Theorem 2.7.

## 4.3. Relation to the multistage problem

In what follows we derive the equations interconnecting the value functions introduced in the preceding section. To this end observe first that

$$V_0 = \inf_{U\in\mathcal{U}} \mathbb{E}\, v(X,U) \tag{16}$$

by definition (10), so that $V_0$ is the optimal value of the initial problem. Further, we have with (9) that

$$v_T = v, \tag{17}$$

which is the starting point of the optimization problem at the final stage.

The following statements interconnect the value functions at intermediate stages.

**Theorem 4.9.** *Let $U \in \mathcal{U}$ be a feasible policy. It holds that*

$$V_t(X_{:t}, U_{t-1}) = \operatorname*{ess\,inf}_{\tilde{u}_t \in \mathcal{U}_t} v_t(X_{:t}, U_{:t-1}, \tilde{u}_t(X_{:t}))$$

*and*

$$v_t(X_{:t}, U_{:t}) \geq \mathbb{E}\left( V_{t+1}\left(X_{:t+1}, U_{:t}\right) \middle| X_{:t}\right). \tag{18}$$

*Equality holds in* (18)*, if $\mathcal{U}$ is decomposable.*

**Proof.** The first equation follows directly from the definition of $V_t$ and $v_t$. The second follows by Proposition 2.10 from

$$\begin{aligned}\mathbb{E}\Big( V_{t+1}\left(X_{:t+1}, U_{:t}\right) \Big| X_{:t}\Big) &= \mathbb{E}\Big( \operatorname*{ess\,inf}_{\tilde{u}_{t+1:T} \in \mathcal{U}_{t+1:T}} \mathbb{E}\left( v(X, U_{:t}, \tilde{u}_{t+1:T}(X)) \middle| X_{:t+1}\right) \Big| X_{:t}\Big) \\ &\leq \operatorname*{ess\,inf}_{\tilde{u}_{t+1:T} \in \mathcal{U}_{t+1:T}} \mathbb{E}\Big( \mathbb{E}\left( v(X, U_{:t}, \tilde{u}_{t+1:T}(X)) \middle| X_{:t+1}\right) \Big| X_{:t}\Big) \\ &= \operatorname*{ess\,inf}_{\tilde{u}_{t+1:T} \in \mathcal{U}_{t+1:T}} \mathbb{E}\Big( v(X, U_{:t}, \tilde{u}_{t+1:T}(X)) \Big| X_{:t}\Big)\end{aligned}$$

and the tower property of the conditional expectation. Equality holds, by Proposition 2.10 again, for decomposable controls and hence the assertion. □

**Remark 4.10.** (Pathwise, or $\omega$-by-$\omega$ description) As above and employing the functions (14) and (15), the equations can be stated directly and explicitly by

$$\begin{aligned} V_t(x_{:t}, u_{:t-1}) &= \inf_{u_t} v_t(x_{:t}, u_{:t-1}, u_t) \quad \text{and} \\ v_t(x_{:t}, u_{:t}) &\geq \mathbb{E}_{X_{t+1}}\left( V_{t+1}(X_{:t+1}, u_{:t}) \middle| X_{:t} = x_{:t}\right) \\ &= \mathbb{E}_{X_{t+1}}\left( V_{t+1}(x_{:t}, X_{t+1}, u_{:t}) \middle| X_{:t} = x_{:t}\right). \end{aligned}$$

Equality holds, if $\mathcal{U}$ is decomposable.

These equations get to the point directly and explain the computational task at each stage ($t$) and at each node ($x_{:t}$, $u_{:t-1}$). Note again that stating the equations this way is not justified from a mathematical perspective, the equations suffer from measurability issues, in general. They are justified in the finite dimensional case if $P(X_{:t} = x_{:t} \text{ and } U_{:t-1} = u_{t-1}) > 0$.

The mutual relations above give rise to combining the components to the following dynamic equations.

**Corollary 4.11.** (Dynamic relations) *Let $U \in \mathcal{U}$ be a feasible control process. Then we have*

$$\boldsymbol{V}_t^U \geq \operatorname*{ess\,inf}_{U'_{:t} \in \mathcal{U}_{:t},\, U'_{:t-1} = U_{:t-1}} \mathbb{E}\left( \boldsymbol{V}_{t+1}^{U'} \middle| X_{:t}\right)$$

*and*

$$\boldsymbol{v}_t^U \geq \mathbb{E}\left( \operatorname*{ess\,inf}_{U'_{:t+1} \in \mathcal{U}_{:t+1},\, U'_{:t} = U_{:t}} \boldsymbol{v}_{t+1}^{U'} \middle| X_{:t}\right).$$

*Equality holds, if $\mathcal{U}$ is decomposable.*

**Proof.** The assertion is immediate by combining the defining equations (12) and (13) and the assertions of Theorem 4.9. $\square$

**Remark 4.12.** (Dynamic relations, pathwise description) It holds that

$$V_t(x_{:t}, u_{:t-1}) \geq \inf_{u_t} \mathbb{E}_{X_{t+1}} \left( V_{t+1}(x_{:t}, X_{t+1}, u_{:t}) \,|\, X_{:t} = x_{:t}, U_{:t-1} = u_{:t-1} \right)$$

and
$$v_t(x_{:t}, u_{:t}) \geq \mathbb{E}_{X_{t+1}} \Big( \inf_{u_{t+1}} v_{t+1}(x_{:t}, X_{t+1}, u_{:t+1}) \big| X_{:t} = x_{:t}, U_{:t} = u_{:t} \Big).$$

Equality holds, if $\mathcal{U}$ is decomposable.

### 4.4. Verification theorems

Verification theorems provide optimality conditions. Given these characterizations it is the purpose of verification theorems to allow verifying or checking, if a given policy is optimal or not. An interesting, early reference is Rockafellar and Wets [33], who study martingales associated with optimality conditions. Fleming and Soner [10] give verification theorems for dynamic (in particular Markovian) problems in continuous time. We shall address this particular situation further in more detail below.

The value process $\boldsymbol{v}_t^U$ is a stochastic process depending on an underlying policy $U$. A special situation occurs if the underlying policy $U$ is optimal, i.e., $U$ solves the initial problem (4). In what follows we examine this situation. We further provide a useful characterization of the optimizers of (4), relating the different concepts regarding optimization and probability theory.

**Theorem 4.13.** (Verification theorem) *Let $u \in \mathcal{U}$ be any policy. Then the stochastic processes*

$$\boldsymbol{v}_t^U = v_t\big(X_{:t}, u_{:t}(X_{:t})\big), \qquad t = 0, \dots, T,$$

*and*
$$\boldsymbol{V}_t^U = V_t\big(X_{:t}, u_{:t-1}(X_{:t-1})\big), \qquad t = 0, \dots, T$$

*are submartingales. They are martingales, if $\mathcal{U}$ is decomposable and if $u$ solves the initial problem* (4)*.*

*Conversely, if $\mathcal{U}$ is decomposable and $\boldsymbol{V}_t^U$, $\boldsymbol{v}_t^U$ are martingales, then $U$ is an optimizer of* (4)*.*

**Proof.** The first assertion is immediate from Corollary 4.11. For the second assume that $\boldsymbol{V}_t^{U^*}$, $\boldsymbol{v}_t^{U^*}$ are martingales for an underlying policy $U^*$. By employing (16), (17) and Theorem 4.9 it follows that

$$\inf_{U \in \mathcal{U}} \mathbb{E}\, v(X, U) = \boldsymbol{V}_0 = \boldsymbol{V}_0^{U^*} = \mathbb{E}\left( \boldsymbol{V}_1^{U^*} \big| X_0 \right) = \boldsymbol{v}_0^{U^*} = \mathbb{E}\left( \boldsymbol{v}_T^{U^*} \big| X_0 \right) = \mathbb{E}\left( v\left(X, U^*\right)\right)$$

and thus the assertion. $\square$

Theorem 4.13 allows identifying a policy $U = u(X)$ as optimal policy by checking, if the value processes constitute a martingale or not. Note that the verification theorem does not give a hint on where and how to improve the policy. Instead, it can be used ex post to check an existing, given policy with respect to optimality.

The verification theorem presented above notably works for every multistage stochastic optimization problem. We did not impose other conditions on the function $v$ except Hölder continuity, and we did not restrict the analysis to Markovian processes. From this mathematical perspective the statement is rather general.

## 5. Special cases, specific and further objective functions

Most common in optimal control, finance and reinforcement learning are value functions, which accumulate costs occurring at consecutive stages. We derive their intermediate value functions explicitly by exploiting the specific structure of the objective function. To this end we transform the equations for the general additive case first and relate the equations to a Markov model subsequently. The Markovian property, from probabilistic perspective, is essential for Bellman equations. As well, we derive the equations for stochastic dual dynamic programming (SDDP) from the general equations.

### 5.1. Lag-$\ell$ stochastic processes and additive objective functions

The particular value function which we consider here,

$$v(x_{:T}, u_{:T}) := \sum_{t=1}^{T} \gamma^{t-1}\, c_t(x_{t-\ell:t}, u_{t-\ell:t-1}), \tag{19}$$

adds consecutive costs at lag $\ell \geq 0$ (entries with negative stage indices are ignored, as $u_{-1:2} = u_{0:2}$, e.g., and the corresponding cost function is adjusted accordingly). The value function (19) is of fundamental importance in finance and in reinforcement learning, where $c_t$ is the cost associated with time $t$ and $\gamma$ is a discount factor. Note the very particular choice of arguments of the function $c_t$: the last input element is the observation $x_t$, but the subsequent decision $u_t$ is not taken into account. Figure 5.1 depicts the support of the cost component $c_t$ at stage $t$ (compare with Figure 4.1).

$$\cdots \rightsquigarrow \underbrace{X_{t-\ell} \rightsquigarrow u_{t-\ell} \rightsquigarrow \cdots \rightsquigarrow X_{t-1} \rightsquigarrow u_{t-1} \rightsquigarrow X_t}_{c_t} \rightsquigarrow \ldots$$

Figure 5.1: Arguments of the cost component $c_t$

The parameter $\gamma \in (-1, 1)$ in (19) is most typically interpreted as *discount factor*. To derive the dynamic equations we assume that the functions $c_t$ are Hölder continuous and assume that the stochastic process associated with the value function (1) has lag $\ell$ as well; that is,

$$\sigma\left(X_1, \ldots, X_t\right) = \sigma\left(X_{t-\ell}, \ldots, X_t\right) \quad \text{for all} \;\; t = \ell, \ldots, T.$$

Define the functions $\tilde{V}_t$ by

$$\tilde{V}_t(x_{:t}, u_{:t-1}) \cdot \gamma^t := V_t(x_{:t}, u_{:t-1}) - \sum_{i=1}^{t} \gamma^{i-1}\, c_i(x_{i-\ell:i}, u_{i-\ell:i-1})$$

so that $\tilde{V}_0 = V_0$.

For additive cost functions, the schematic decomposition (1) now is

$$\inf_{u_0} c_0 + \mathbb{E}_{X_1} \inf_{u_1} c_1 + \cdots + \mathbb{E}_{X_t} \inf_{u_t} \underbrace{c_t + \underbrace{\mathbb{E}_{X_{t+1}} \inf_{u_{t+1}} c_{t+1} + \cdots + \mathbb{E}_{X_T} \inf_{u_T} c_T}_{\tilde{v}_t(x_{1:t},u_{0:t})}}_{\tilde{V}_t(x_{1:t},u_{0:t-1})}.$$

From (14) we conclude that

$$\begin{aligned}&\tilde{V}_t(x_{:t}, u_{:t-1})\\ &= \inf_{u_{t:}(\cdot)} \mathbb{E}\left( \sum_{i=t+1}^{T} \gamma^{i-1-t}\, c_i\big(X_{i-\ell:i}, u_{i-\ell:i-1}, u_{i:T}(X_{:T})\big) \,\middle|\, \begin{array}{l} X_{:t} = x_{:t}, \\ U_{:t-1} = u_{:t-1} \end{array} \right).\end{aligned} \tag{20}$$

The function inside the expectation is independent of $x_{:t-\ell}$ and the stochastic process $X$ has lag $\ell$. Further, the decision process $U$ is adapted to $X$ (cf. (4)) and thus has lag $\ell$ as well. With that it follows that (20) actually is

$$\begin{aligned}&\tilde{V}_t(x_{t-\ell+1:t}, u_{t-\ell+1:t-1})\\ &= \inf_{u_{t:}(\cdot)} \mathbb{E}\left( \sum_{i=t+1}^{T} \gamma^{i-1-t}\, c_i\big(X_{:i}, u_{i-\ell:i-1}, u_{i:T}(X_{:T})\big) \,\middle|\, \begin{array}{l} X_{t-\ell+1:t} = x_{t-\ell+1:t}, \\ U_{t-\ell+1:t-1} = u_{t-\ell+1:t-1} \end{array} \right).\end{aligned}$$

Employing Remark 4.12 we deduce the recursion

$$\begin{aligned}&\tilde{V}_t(x_{t-\ell+1:t}, u_{t-\ell+1:t-1})\\ &\geq \inf_{u_t} \mathbb{E}_{X_{t+1}} \left( \begin{array}{l} c_{t+1}(x_{t+1-\ell:t}, X_{t+1}, u_{t+1-\ell:t}) \\ \quad +\gamma\, \tilde{V}_{t+1}(x_{t-\ell+2:t}, X_{t+1}, u_{t-\ell+2:t}) \end{array} \,\middle|\, \begin{array}{l} X_{t-\ell+1:t} = x_{t-\ell+1:t}, \\ U_{t-\ell+1:t-1} = u_{t-\ell+1:t-1} \end{array} \right),\end{aligned} \tag{21}$$

where equality indicates optimality. This backwards recursion leads to the following discussion on Markov models.

## 5.2. Markov model

Markov models consider the cost functions (19) with lag $\ell = 1$, i.e.,

$$v(x_{:T}, u_{:T}) := \sum_{t=1}^{T} \gamma^{t-1}\, c_t(x_{t-1}, x_t; u_{t-1}) \tag{22}$$

and a process $X$ with the same lag $\ell = 1$, i.e., a Markovian process. With that, the recursion (21) collapses further to

$$\tilde{V}_t(x_t) = \inf_{u_t} \mathbb{E}\left( c_{t+1}(x_t, X_{t+1}, u_t) + \gamma\, \tilde{V}_{t+1}(X_{t+1}) \,\middle|\, X_t = x_t \right). \tag{23}$$

This recursion is also known as backward induction involving the Bellman principle (cf. Bellman [1, 2]), which is of fundamental importance in dynamic programming.

**Remark 5.1.** The Markov model above should not be confused with Markov decision processes (MDP). The MDP literature considers trajectories which are driven themselves by the control $u$ (the control is called *action* in the MDP literature). To recognize this dependency in addition we can restate the recursion as

$$\tilde{V}_t(x_t) = \inf_{u_t} \mathbb{E}_{u_t} \left( c_{t+1}(x_t, X_{t+1}, u_t) + \gamma\, \tilde{V}_{t+1}(X_{t+1}) \middle| X_t = x_t \right),$$

where $\mathbb{E}_u$ is the expectation with respect to the kernel $P_u(\cdot \mid x_t)$, which explicitly depends on the decision $u$. The general considerations of this paper do not cover MDPs.

### 5.3. Dynamic optimization and Bellman's principle of optimality

The cost function (22) is also considered on an infinite horizon, that is,

$$v(x_{:T}, u_{:T}) := \sum_{t=1}^{\infty} \gamma^{t-1} \cdot c_t(x_{t-1}, x_t; u_{t-1}), \tag{24}$$

where $\gamma \in (-1, 1)$ (although, most typically, $\gamma \in (0, 1)$). The value function (19) is bounded in the chosen setting, if the cost functions are uniformly bounded, that is, $|c_t| \leq K < \infty$.

A particularly interesting situation arises for cost functions which do not depend on the stage $t$, i.e., $c_t = c$ and decision satisfying $(X_t, X_{t+1}) \sim (X, X')$. Then, the value functions $\tilde{V}_t$ does not depend on $t$ neither and the following equation holds:

$$\tilde{V}(x) = \inf_{u} \mathbb{E} \left( c(x, X', u) + \gamma\, \tilde{V}(X') \middle| X = x \right). \tag{25}$$

This is a fixed point equation and Banach's fixed point theorem can be applied to prove existence and uniqueness of the value function $\tilde{V}$ in appropriate spaces. As well, the equation (25) specifies an iterative scheme to improve the value function $\tilde{V}$ in consecutive steps. As an example we state the following, where we refer to Fleten [11] for a proof in a similar situation.

**Theorem 5.2.** *Suppose that $c$ is continuous and $X \in K$ a.s. for some compact set $K \subset \mathbb{R}^n$ and $|\gamma| < 1$. Then the value function $\tilde{V}$ is continuous and $\tilde{V} \in C(K)$.*

### 5.4. SDDP

The problem setting of stochastic dual dynamic programming (SDDP) considers a stagewise independent stochastic process $X_t$ (i.e., $X_t$ is *independent* of all preceding $X_{t'}$, $t' < t$), which is a further simplification of all situations described above; it is just noise. With $X_t \sim X$, the dynamic equation reduces further to

$$\tilde{V}_t(x_t) = \inf_{u_t} \mathbb{E} \left( c_{t+1}(x_t, X_{t+1}, u_t) + \gamma\, \tilde{V}_{t+1}(X_{t+1}) \right). \tag{26}$$

This is the simplest situation from a statistical perspective, and it is not surprising that large and extensive problem settings are accessible for numerical computations.

The important algorithm for SDDP for solving the problem (26) efficiently originated in Pereira and Pinto [26].

We refer to Shapiro [38] for an extended analysis of the algorithm, to Römisch and Guigues [36] and to Girardeau et al. [13], Philpott and Guan [28] for convergence proofs of the algorithm.

## 6. Summary

Multistage stochastic optimization has many applications in varying areas, from finance to data science to mention just two. The problems are popular and typically stated conditioned on partial realizations. This pathwise, or $\omega$-by-$\omega$, perspective lacks mathematical rigor.

This paper clarifies that multistage optimization problems, even if given in an informal, pathwise or $\omega$-by-$\omega$ way can be cast with mathematical rigor. We start by outlining the general problem and employ the Kolmogorov continuity theorem to verify that value functions are well-defined, even if conditioned on sets of measure zero. In addition, this approach provides regularity of the value functions.

Verification theorems can be employed to confirm that candidate policies are optimal. We further characterize optimal policies by involving martingales to characterize these optimal solutions.

Markov decision processes, the Bellman principle for reinforcement learning and stochastic dual dynamic programming are probably most well-known and common in practice of dynamic programming. We derive these problem settings as special cases and, in this way, provide rigorous mathematical foundations.

## References

[1] R. E. Bellman: *Dynamic Programming*, Princeton University Press, Princton (1957).

[2] R. E. Bellman: *Adaptive Control Processes*, Princeton Legacy Library 2045, Princeton University Press (1961).

[3] D. Bertsekas: *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont (2012).

[4] D. Bertsekas, S. E. Shreve: *Stochastic Optimal Control: the Discrete Time Case*, Academic Press, New York (1978).

[5] P. Carpentier, J.-P. Chancelier, G. Cohen, M. De Lara: *Stochastic Multi-Stage Optimization*, Springer, Cham (2015).

[6] P. Carpentier, J.-P. Chancelier, G. Cohen, M. De Lara, P. Girardeau: *Dynamic consistency For stochastic optimal control problems*, Ann. Oper. Res. 200/1 (2012) 247–263.

[7] N. Dunford, J. T. Schwartz: *Linear Operators. Part I: General Theory*, Wiley-Interscience, New York (1957).

[8] I. V. Evstigneev: *Measurable selection and dynamic programming*, Math. Oper. Res. 1/3 (1976) 267–272.

[9] E. A. Feinberg: *On measurability and representation of strategic measures in Markov decision processes*, in: Statistics, Probability and Game Theory, IMS Lecture Notes and Monograph Series Vol. 30 Institute of Mathematical Statistics (1996) 29–43.

[10] W. H. Fleming, H. M. Soner: *Controlled Markov Processes and Viscosity Solutions*, 2nd ed., Springer, New York (2006).

[11] S.-E. Fleten, E. Haugom, A. Pichler, C. J. Ullrich: *Structural estimation of switching costs for peaking power plants*, Eur. J. Oper. Res. 285/1 (2020) 23–33.

[12] H. Föllmer, A. Schied: *Stochastic Finance: An Introduction in Discrete Time*, De Gruyter Studies in Mathematics 27, De Gruyter, Berlin (2004.

[13] P. Girardeau, V. Leclère, A. B. Philpott: *On the convergence of decomposition methods for multistage stochastic convex programs*, Math. Oper. Res. 40/1 (2014) 1–16.

[14] K. Gowrisankaran: *Measurability of functions in product spaces*, Proc. Amer. Math. Soc. 31 (1972) 485–488.

[15] O. Kallenberg: *Foundations of Modern Probability*, Springer, New York (2002).

[16] I. Karatzas, S. E. Shreve: *Methods of Mathematical Finance*, Applications of Mathematics Vol. 39, Springer, Berlin (1998).

[17] A. Klenke: *Probability Theory*, Springer, London (2014).

[18] G. Lan, Z. Zhou: *Dynamic stochastic approximation for multi-stage stochastic optimization*, Math. Programming Ser. A 187/1-2 (2021) 487–532.

[19] N. Löhndorf, D. Wozabal, S. Minner: *Optimizing trading decisions for hydro storage systems using approximate dual dynamic programming*, Oper. Res. 61/4 (2013) 810–823.

[20] F. Maggioni, G. Pflug: *Guaranteed bounds for general non-discrete multistage risk-averse stochastic optimization programs*, SIAM J. Optimization 29/1 (2019) 454–483.

[21] P. Olsen: *Multistage stochastic programming with recourse as mathematical programming in an $l_p$ space*, SIAM J. Control Optimization 3 (1976) 528–537.

[22] P. Olsen: *Discretizations of multistage stochastic programming problems*, Math. Programming Study 6 (1976) 111–124.

[23] P. Olsen: *When is a multistage stochastic programming problem well-defined?*, SIAM J. Control Optimization 14/3 (1976) 518–527.

[24] T. Pennanen, A.-P. Perkkiö: *Dynamic programming in convex stochastic optimization*, J. Convex Analysis 30/4 (2023) 1241–1283.

[25] T. Pennanen, A.-P. Perkkiö: *Convex stochastic optimization*, preprint (2023).

[26] M. V. F. Pereira, L. M. V. G. Pinto: *Multi-stage stochastic optimization applied to energy planning*, Math. Programming 52/1-3 (1991) 359–375.

[27] G. Pflug, A. Pichler: *Multistage Stochastic Optimization*, Springer Series in Operations Research and Financial Engineering, Springer, Cham (2014).

[28] A. B. Philpott, Z. Guan: *On the convergence of stochastic dual dynamic programming and related methods*, Oper. Res. Letters 36/4 (2008) 450–455.

[29] A. B. Philpott, V. L. de Matos, E. Finardi: *On solving multistage stochastic programs with coherent risk measures*, Oper. Res. 61/4 (2013) 957–970.

[30] A. Pichler, R. P. Liu, A. Shapiro: *Risk-averse stochastic programming: Time consistency and optimal stopping*, Oper. Res. 70/4 (2022) 2439–2455.

[31] A. Pichler, A. Shapiro: *Mathematical foundations of distributionally robust multistage optimization*, SIAM J. Optimization 31/4 (2021) 3044–3067.

[32] R. T. Rockafellar: *Integral functionals, normal integrands and measurable selections*, in: *Nonlinear Operators and the Calculus of Variations*, Summer School Bruxelles 1975, Lecture Notes in Mathematics 543, Springer, Berlin (1976) 157–207.

[33] R. T. Rockafellar, R. J.-B. Wets: *Nonanticipativity and $L^1$-martingales in stochastic optimization problems*, Math. Programming Study 6 (1976) 170–187.

[34] R. T. Rockafellar, R. J.-B. Wets: *On the interchange of subdifferentiation and conditional expectations for convex functionals*, Stochastics 7/3 (1982) 173–182.

[35] R. T. Rockafellar, R. J.-B. Wets: *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften 317, Springer, Berlin (1998).

[36] W. Römisch, V. Guigues: *Sampling-based decomposition methods for multistage stochastic programs based on extended polyhedral risk measures*, SIAM J. Optimization 22/2 (2012) 286–312.

[37] A. Ruszczyński: *Risk-averse dynamic programming for Markov decision processes*, Math. Programming Ser. B 125 (2010) 235–261.

[38] A. Shapiro: *Analysis of stochastic dual dynamic programming method*, Eur. J. Oper. Res. 209/1 (2010) 63–72.

[39] A. Shapiro: *Time consistency of dynamic risk measures*, Oper. Res. Letters 40/6 (2012) 436–439.

[40] A. Shapiro: *Interchangeability principle and dynamic equations in risk averse stochastic programming*, Oper. Res. Letters 45/4 (2017) 377–381.

[41] A. Shapiro: *Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming*, Eur. J. Oper. Res. 288/1 (2021) 1–13.

[42] A. Shapiro, D. Dentcheva, A. Ruszczyński: *Lectures on Stochastic Programming. Modeling and Theory*, 3rd ed., MOS-SIAM Series on Optimization Vol. 28, Society for Industrial and Applied Mathematics, Philadelphia (2021).

[43] A. Shapiro, W. Tekaya, J. P. da Costa, M. Pereira Soares: *Risk neutral and risk averse stochastic dual dynamic programming method*, Eur. J. Oper. Res. 224/2 (2013) 375–391.

[44] A. N. Shiryaev: *Probability*, Springer, New York (1996).

[45] A. W. van der Vaart, J. A. Wellner: *Weak Convergence and Empirical Processes. With Applications to Statistics*, Springer Series in Statistics, Springer, New York (1996).

[46] J. Zhen, D. Kuhn, W. Wiesemann: *Mathematical foundations of robust and distributionally robust optimization*, arXiv: 2105.00760 (2021).