On Stability of Multistage Stochastic Programs

Christian Küchler*

Humboldt–Universität zu Berlin, Unter den Linden 6,
D–10099 Berlin, Germany, ckuechler@math.hu-berlin.de

Abstract

We study the quantitative stability of linear multistage stochastic programs under perturbations of the underlying stochastic processes. It is shown that the optimal values behave Lipschitz continuously with respect to an L^p-distance. In order to establish continuity of the recourse function with respect to the current state of the stochastic process, we assume continuity of the conditional distributions in terms of a Fortet-Mourier metric. The main stability result holds for nonanticipative approximations of the underlying process and thus represents a rigorous justification of established discretization techniques.

AMS 2000 subject classification: 90C15, 90C31

Key words: stochastic programming, multistage, stability, probability metrics

Introduction

Many stochastic optimization problems of practical interest do not allow for an analytic solution and numerical approaches require the underlying probability measure to have finite support. Whenever the initial probability measure does not meet these demands, it has to be approximated by an auxiliary measure. Thereby, it is reasonable to choose the approximating measure such that the optimal value and the set of optimal decisions of the auxiliary problem are close to those of the originial problem. Consequently, perturbation and stability analysis of stochastic programs is necessary for the development of reliable techniques for discretization and scenario reduction. While stability properties are well understood for non-dynamic chance constrained and two-stage problems, cf. the recent survey by Römisch (2003), it turned out that the multistage case is more intricate. Recently, the latter situation has been studied by a variety of authors, thus the

*Support by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) and by the Bundesministerium für Bildung und Forschung (BMBF) under the grant 03SF0312E is gratefully acknowledged.

following references should not be considered to be exhaustive. Statistical bounds have been provided by Shapiro (2003). Pennanen (2005) established asymptotic stability of specific approximations for a general class of convex multistage problems in terms of epi-convergence. In doing so, he noticed that such quantitative results, as we discuss in this paper, require stronger assumptions. Indeed, the restriction on models with continuous decisions allowed Mirkov and Pflug (2007) to establish such a quantitative stability result for their tree approximations. Heitsch, Römisch, and Strugarek (2006) did not require regularity conditions on decisions and underlying processes. Consequently, their quantitative stability result, obtained by considering arbitrary perturbations of the underlying process, incorporates a term measuring the distance of the filtrations induced by the initial and the auxiliary process respectively. Vanishing in the two-stage case, this filtration distance reflects the relevance of the information structure and of the nonanticipativity constraints for multistage decision problems. We refer also to Barty (2004) who studied the role of information in stochastic optimization problems and introduced and reviewed several concepts of distances between filtrations.

The recent approach of Heitsch and Römisch (2007) aims to incorporate filtration distances into the construction of scenario trees. However, this requires some extra effort and, to the best of our understanding, these distances are not taken into account by a variety of established techniques. Thus, the main purpose of this paper is to provide general conditions under which these somewhat delicate terms may be omitted.

One of the main difficulties seems to be that without additional assumptions neither the recourse function nor an optimal decision depend continuously on the current state of the underlying process in general. Rockafellar and Wets (1974) showed that under weak conditions, the optimal value can be approximated by *continuous decisions*. However, while this allows one to deduce convergence results, such as those due to Pennanen (2005), it does not lead to quantitative estimates. For deriving *continuity of the recourse function* and bounds based on a barycentric approximation scheme, Kuhn (2005) required the underlying processes to be autoregressive. He also indicated, that *the key element in any scenario tree construction is the discretization of the conditional probabilities*. In particular, continuous dependency of these probabilities on the current state of the underlying process is necessary for potential continuity of the recourse function and can be seen as *continuity of the available information with respect to the current state*. It is illustrated by Example 2.6 of Heitsch, Römisch, and Strugarek (2006) that the latter property is indispensable in order to omit any filtration distances and to obtain a good approximation of the initial process by usual techniques which are based on stagewise clustering. Thus, we impose Lipschitz continuity of the conditional distributions to verify the same regularity for the recourse function in Theorem 1. With this at hand, we estimate in Theorem 2 the gap between the optimal value and the costs of a decision that is locally *calm*. This leads to our main result, Theorem 3, which provides an upper bound for the perturbation of the optimal value.

Acknowledgement. The author is grateful to Prof. Werner Römisch for his help and encouragement, to Stefan Vigerske for being perpetually willing for helpful discussions, and to Thomas Surowiec for proofreading the manuscript. The author also would like to thank the anonymous referees whose careful comments helped to clarify the presentation.

Notation and Conventions. Random variables are denoted by bold letters, for example $\boldsymbol{\xi}$ or $\boldsymbol{x}$, in contrast to their realizations (i.e., elements of their support) which are denoted by ξ or x, respectively. The notation $\boldsymbol{\xi}^t$ is used for the vector $(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_t)$, $\|\cdot\|$ denotes the maximum norm on $\mathbb{R}^n$ for the respective value $n \in \mathbb{N}$, and we set $\|\xi^t\| \triangleq \max_{i \le t} \|\xi_i\|$.

1 Problem Formulation

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we consider an $\mathbb{R}^s$-valued stochastic process $\boldsymbol{\xi} = (\boldsymbol{\xi}_t)_{t=1}^T$ with time horizon $T \in \mathbb{N}$ and the associated filtration $(\mathcal{F}_t)_{t=1}^T$ defined through $\mathcal{F}_t \triangleq \sigma(\boldsymbol{\xi}^t)$ for $t = 1, \dots, T$. We assume that $\mathcal{F}_1 = \{\Omega, \emptyset\}$, $\boldsymbol{\xi} \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for every $p \in [1, +\infty)$, and set

$$\mathbb{P}^t \triangleq \mathbb{P}\left[\boldsymbol{\xi}^t \in \cdot\ \right] \quad \text{and} \quad \Xi^t \triangleq \operatorname{supp} \mathbb{P}^t \subset \mathbb{R}^{s \cdot t} \quad \text{for} \quad t = 1, \dots, T.$$

Furthermore, we consider the *costs* $b_t(\cdot)$, the *technology matrices* $A_{t,1}(\cdot)$, and the *right-hand sides* $h_t(\cdot)$, which all are assumed to depend affinely on $\xi_t \in \Xi_t$ for $t = 1, \dots, T$. Together they define the set-valued mappings

$$\begin{aligned} M_t : X_{t-1} \times \Xi_t &\rightrightarrows X_t, \\ M_t(x_{t-1}, \xi_t) &\triangleq \{x_t \in X_t \ :\ A_{t,0} x_t + A_{t,1}(\xi_t) x_{t-1} = h_t(\xi_t)\} \end{aligned}$$

where $X_t \subset \mathbb{R}^m$ are certain nonempty, closed, and polyhedral sets, with $t = 1, \dots, T$. We assume complete recourse, i.e., $M_t(x_{t-1}, \xi_t)$ is nonempty for every $x_{t-1} \in X_{t-1}$ and every $\xi_t \in \Xi_t$. The objective function is given by

$$\begin{aligned} \varphi : \mathbb{R}^{m \cdot T} \times \Xi^T &\to \mathbb{R}, \\ \varphi(x_1, \dots, x_T, \xi^T) &\triangleq \sum_{t=1}^T \langle b_t(\xi_t), x_t \rangle. \end{aligned}$$

A tuple $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$ of Borel-measurable mappings $\boldsymbol{x}_t : \Xi^t \to X_t$, $t = 1, \dots, T$, is called a *feasible decision* with respect to $\boldsymbol{\xi}$, if the recourse condition

$$\boldsymbol{x}_t(\boldsymbol{\xi}^t) \in M_t(\boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1}), \boldsymbol{\xi}_t) \tag{1}$$

is fulfilled $\mathbb{P}-$a.s. for $t = 2, \dots, T$. The class of feasible decisions $\boldsymbol{x}$ will be denoted by $\mathcal{S}(\boldsymbol{\xi})$ and, for the sake of notational convenience, we set $x_0 = 1$.

We study the following multistage stochastic optimization problem:

$$v(\boldsymbol{\xi}) \triangleq \inf_{\boldsymbol{x} \in \mathcal{S}(\boldsymbol{\xi})} \mathbb{E}\left[\varphi(\boldsymbol{x}(\boldsymbol{\xi}), \boldsymbol{\xi})\right], \tag{2}$$

and aim to establish an upper bound for the perturbation of $v(\boldsymbol{\xi})$ when $\boldsymbol{\xi}$ is replaced by another process $\tilde{\boldsymbol{\xi}}$.

Complete recourse and the polyhedral form of M_t allow one to conclude, see Example 9.35 of Rockafellar and Wets (1998), that M_t is Lipschitz continuous on $X_{t-1} \times \Xi_t$ with respect to the Pompeiu-Hausdorff distance $\mathbb{d}$ in the following sense. There exists a constant $M \geq 0$ with

$$\begin{aligned} \mathbb{d}\left(M_t(x_{t-1}, \xi_t), M_t(\hat{x}_{t-1}, \xi_t)\right) &\leq M \cdot \max\{1, \|\xi_t\|\} \cdot \|\hat{x}_{t-1} - x_{t-1}\| \quad \text{and} \\ \mathbb{d}\left(M_t(x_{t-1}, \xi_t), M_t(x_{t-1}, \hat{\xi}_t)\right) &\leq M \cdot \max\{1, \|x_{t-1}\|\} \cdot \|\hat{\xi}_t - \xi_t\|, \end{aligned}$$

for every $(x_{t-1}, \xi_t), (\hat{x}_{t-1}, \hat{\xi}_t) \in X_{t-1} \times \Xi_t$. We recall that the Pompeiu-Hausdorff distance between two sets $A, B \subset \mathbb{R}^m$ is defined by

$$\mathbb{d}(A, B) \triangleq \max\left\{\sup_{a \in A} \operatorname{dist}(a, B), \sup_{b \in B} \operatorname{dist}(b, A)\right\}.$$

Remark 1.1. *Throughout this paper, the linearity of M_t is used only to obtain the claimed Lipschitz continuity. Analogously, we assume linear costs $\langle b_t(\xi_t), x_t\rangle$ only to ensure the existence of a constant $B \geq 0$ with*

$$\begin{aligned} \|\langle b_t(\xi_t), x_t\rangle - \langle b_t(\hat{\xi}_t), x_t\rangle\| &\leq B\|\xi_t - \hat{\xi}_t\|\|x_t\| \quad \text{and} \\ \|\langle b_t(\xi_t), x_t\rangle - \langle b_t(\xi_t), \hat{x}_t\rangle\| &\leq B \max\{1, \|\xi_t\|\}\|x_t - \hat{x}_t\|. \end{aligned}$$

The integrability condition on $\boldsymbol{\xi}$ is assumed for notational simplicity. Actually, it suffices to have $\boldsymbol{\xi} \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for a sufficiently large $p \in \mathbb{R}_+$.
Furthermore, all results remain valid if M_t, h_t, and b_t depend on $\boldsymbol{\xi}^t$ instead of $\boldsymbol{\xi}_t$.

2 Continuity of the Recourse Function

Let $V_t : \Xi^t \times X_{t-1} \to \mathbb{R}$ be the recourse function at time t, which is defined recursively by $V_{T+1} \triangleq 0$ and the Dynamic Programming Equation

$$V_t(\xi^t, x_{t-1}) \triangleq \inf_{x_t \in M_t(x_{t-1}, \xi_t)} \langle b_t(\xi_t), x_t\rangle + \mathbb{E}\left[V_{t+1}\left(\boldsymbol{\xi}^{t+1}, x_t\right) \middle| \boldsymbol{\xi}^t = \xi^t\right] \quad \text{for } t = T, \ldots, 1,$$

where the mapping $(x_t, \xi^t) \mapsto \mathbb{E}\left[V_{t+1}\left(\boldsymbol{\xi}^{t+1}, x_t\right) \middle| \boldsymbol{\xi}^t = \xi^t\right]$ denotes the *regular conditional expectation* of $V_{t+1}(\cdot, \cdot)$ relative to $\mathcal{F}_t$. The value $V_t(\xi^t, x_{t-1})$ represents the minimal achievable expected future costs after having chosen $\boldsymbol{x}_{t-1} = x_{t-1}$, having observed $\left\{\boldsymbol{\xi}^t = \xi^t\right\}$, and before deciding on x_t. In particular, we have the identity $v(\boldsymbol{\xi}) = V_1(\boldsymbol{\xi}_1, x_0)$ and complete recourse implies that $V_t < +\infty$ holds true on $\Xi^t \times X_{t-1}$. It was proved by Evstigneev (1976) that V_t is well defined and measurable under the following

Assumption 2.1.

(i) There exists an integrable random variable Q such that $\varphi(x,\boldsymbol{\xi}) \geq Q$ holds $\mathbb{P}-$a.s. for every $x \in \mathbb{R}^{m\cdot T}$.

(ii) For each $c \in \mathbb{R}$ the random level set $\{x \in \mathbb{R}^{m\cdot T} : \ \varphi(x,\boldsymbol{\xi}) \leq c\}$ is compact $\mathbb{P}-$a.s.

A decision $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{\xi})$ is optimal if and only if the equality

$$V_t(\xi^t, \boldsymbol{x}_{t-1}(\xi^{t-1})) = \langle b_t(\xi_t), \boldsymbol{x}_t(\xi^t)\rangle + \mathbb{E}\left[V_{t+1}\left(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t(\xi^t)\right) \middle|\, \boldsymbol{\xi}^t = \xi^t\right] \tag{3}$$

holds for $\mathbb{P}^t-$almost every $\xi^t \in \Xi^t$ and $t = 1, \ldots, T$. Moreover, for every Borel measurable mapping $\boldsymbol{x}_{t-1} : \Xi^{t-1} \to X_{t-1}$ there exists a measurable $\boldsymbol{x}_t : \Xi^t \to X_t$ such that relation (3) holds true for $\mathbb{P}^t-$almost every $\xi^t \in \Xi^t$. Actually, Evstigneev (1976) allows to show that the $\mathbb{P}^t$-null sets on which the latter property does not hold coincide for all measurable $\boldsymbol{x}_{t-1}$. Indeed, the following corollary is an immediate consequence of applying Lemma 4 of Evstigneev (1976) within the proof of his Theorem 2.

Corollary 2.2. *For $t = 1, \ldots, T$, there is a Borel set $A^{t'} \subset \Xi^t$ with $\mathbb{P}^t[A^{t'}] = 1$ such that the following property holds.*
For every Borel measurable mapping $\boldsymbol{x}_{t-1} : \Xi^{t-1} \to X_{t-1}$ there exists a measurable $\boldsymbol{x}_t : \Xi^t \to X_t$ such that identity (3) *holds true for every $\xi^t \in A^{t'}$.*

We assume the decision $\boldsymbol{x}_t$ can be chosen to fulfill a certain growth condition:

Assumption 2.3. *For $t = 1, \ldots, T$, there is a Borel set $A^{t'} \subset \Xi^t$ with $\mathbb{P}^t[A^{t'}] = 1$ and a constant $L \geq 1$ such that the following property holds.*
For every Borel measurable mapping $\boldsymbol{x}_{t-1} : \Xi^{t-1} \to X_{t-1}$ there exists a measurable $\boldsymbol{x}_t : \Xi^t \to X_t$ such that identity (3) *and the growth condition*

$$\|\boldsymbol{x}_t(\xi^t)\| \leq L \cdot \max\left\{1, \|\boldsymbol{x}_{t-1}(\xi^{t-1})\|\right\} \cdot \max\left\{1, \|\xi^t\|\right\} \tag{4}$$

hold true for every $\xi^t \in A^{t'}$.

Remark 2.4. *Unfortunately, the existence of decisions which are bounded in the above sense may be hard to verify, in general. However,* (4) *holds true for* every $x_t \in M_t(x_{t-1}, \xi_t)$ *if X_t is bounded, or, more general, whenever the projection of X_t onto the kernel of the recourse matrix $A_{t,0}$ is bounded.*
Furthermore, the linear growth condition (4) *could be relaxed to polynomial growth, then the growth rate in ξ^t of the Lipschitz constant in Theorem 1 and the subsequent results would change accordingly.*

Assumptions 2.1 and 2.3 imply the existence of an optimal decision $\boldsymbol{x}$ satisfying

$$\|\boldsymbol{x}_t(\boldsymbol{\xi}^t)\| \leq L^t \cdot \max\left\{1, \|\boldsymbol{\xi}^t\|\right\}^{t-1} \qquad \mathbb{P} - \text{a.s. for } t = 1, \ldots, T. \tag{5}$$

Indeed, a tuple $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$ of mappings with (1) and (3)–(5) can be constructed by recursion, and from Theorem 14.37 of Rockafellar and Wets (1998) it follows that every $\boldsymbol{x}_t$ can be chosen to be measurable. Consequently, $\boldsymbol{x}$ is an optimal decision. Decisions fulfilling (4) and (5) will be denoted as *bounded* in the following.

To establish a quantitative stability result, we will study the continuity of V_t. Thereby, regularity properties of the mapping $x_{t-1} \mapsto V_t(\xi^t, x_{t-1})$ are well-known. We refer to Birge and Louveaux (1997) and Ruszczyński and Shapiro (2003) who derived convexity as well as piecewise linearity for the case of finite Ξ^T and to Kuhn (2005) who proved continuity under compactness assumptions on Ξ^T and $X_1, \dots, X_T$. Thus, the following Proposition can be seen as an adaption of these results to our Lipschitz continuous framework.

Proposition 2.5. *The recourse function V_t is Lipschitz continuous with respect to the decision x_{t-1} in the following sense. For $t = 1, \dots, T$, there exists a constant $\bar{M} > 0$ and a Borel set $A^{t\prime\prime} \subset \Xi^t$ with $\mathbb{P}^t[A^{t\prime\prime}] = 1$, such that for every $\xi^t \in A^{t\prime\prime}$ the relation*

$$\left| V_t(\xi^t, x_{t-1}) - V_t(\xi^t, \hat{x}_{t-1}) \right| \le [V_t]^x_{Lip}(\xi^t) \cdot \| x_{t-1} - \hat{x}_{t-1} \| \tag{6}$$

holds true for every $x_{t-1}, \hat{x}_{t-1} \in X_{t-1}$ with a (random) Lipschitz constant $[V_t]^x_{Lip}(\xi^t)$ satisfying

$$[V_t]^x_{Lip}(\xi^t) \le \bar{M} \cdot \mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^T\|\}^{2+T-t} \;\middle|\; \boldsymbol{\xi}^t = \xi^t \right]. \tag{7}$$

Proof. The assertion is true for $V_{T+1} \equiv 0$. Assume it is true also for $s = t+1, \dots, T$ with Lipschitz constants $[V_s]^x_{Lip}$ and that the difference on the left side of (6) is negative. Then, due to (3), there exists an $\boldsymbol{x}^*_t(\xi^t) \in M_t(x_{t-1}, \xi_t)$, such that the left side of (6) coincides for $\mathbb{P}^t$-a.e. ξ^t with

$$\begin{aligned} &-\langle b_t(\xi_t), \boldsymbol{x}^*_t(\xi^t) \rangle - \mathbb{E}\left[V_{t+1}\left(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}^*_t(\xi^t) \right) \middle| \boldsymbol{\xi}^t = \xi^t \right] \\ &\quad + \inf_{\hat{x}_t \in M_t(\hat{x}_{t-1}, \xi_t)} \left\{ \langle b_t(\xi_t), \hat{x}_t \rangle + \mathbb{E}\left[V_{t+1}\left(\boldsymbol{\xi}^{t+1}, \hat{x}_t \right) \middle| \boldsymbol{\xi}^t = \xi^t \right] \right\}. \end{aligned} \tag{8}$$

Moreover, it follows from Corollary 2.2 that we may assume that the $\mathbb{P}^t(d\xi^t)$-null sets on which this identity does not hold coincide for all $x_{t-1} \in X_{t-1}$. Due to Theorem 14.37 of Rockafellar and Wets (1998) we can choose a measurable $\hat{\boldsymbol{x}}^*_t$ with

$$\hat{\boldsymbol{x}}^*_t(\xi^t) \in \arg \min_{z \in M_t(\hat{x}_{t-1}, \xi_t)} \left\| z - \boldsymbol{x}^*_t(\xi^t) \right\|$$

to estimate (8) from above by

$$|\langle b_t(\xi_t), \boldsymbol{x}^*_t(\xi^t) - \hat{\boldsymbol{x}}^*_t(\xi^t) \rangle| + |\mathbb{E}\left[V_{t+1}\left(\boldsymbol{\xi}^{t+1}, x^*_t \right) - V_{t+1}\left(\boldsymbol{\xi}^{t+1}, \hat{x}^*_t \right) \middle| \boldsymbol{\xi}^t = \xi^t \right]|.$$

From the linear growth of b_t and the Lipschitz continuity of V_{t+1} with respect to x_t one concludes that this term is not greater than

$$\left(B \max\{1, \|\xi_t\|\} + \mathbb{E}\left[[V_{t+1}]^x_{Lip}(\boldsymbol{\xi}^{t+1}) \middle| \boldsymbol{\xi}^t = \xi^t \right] \right) \cdot \left\| \boldsymbol{x}^*_t(\xi^t) - \hat{\boldsymbol{x}}^*_t(\xi^t) \right\|,$$

again $\mathbb{P}^t(d\xi^t)$-a.s. for every $x_{t-1}, \hat{x}_{t-1} \in X_{t-1}$. By definition of $\hat{\boldsymbol{x}}_t^*$ and Lipschitz continuity of M_t, the latter term is bounded from above by

$$\Big(MB\max\{1, \|\xi_t\|^2\} + M\max\{1, \|\xi_t\|\} \cdot \mathbb{E}\left[[V_{t+1}]^x_{Lip}(\boldsymbol{\xi}^{t+1}) \;\Big|\; \boldsymbol{\xi}^t = \xi^t\right]\Big) \cdot \|x_{t-1} - \hat{x}_{t-1}\|.$$

An analoguous estimate holds whenever the difference on the left side of (6) is positive. Hence, $[V_t]^x_{Lip}(\xi^t)$ is given by the term in parentheses, from where we conclude by recursion that we can put

$$[V_t]^x_{Lip}(\xi^t) \triangleq B\sum_{i=t}^{T} M^{i-t+1}\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}_i\|^2\} \cdot \prod_{k=t}^{i-1}\max\{1, \|\boldsymbol{\xi}_k\|\} \;\middle|\; \boldsymbol{\xi}^t = \xi^t\right].$$

Thus, the asserted bound for $[V_t]^x_{Lip}$ results from a straightforward estimate. □

Establishing continuity of $\xi^t \mapsto V_t(\xi^t, x_{t-1})$ is more subtle since, unlike the decision variable x_{t-1}, the observation ξ^t impacts not only the Lipschitz continuous time coupling constraints at time t, but also the expectations about future realizations of $\boldsymbol{\xi}$. Therefore, one can hardly expect V_t to be Lipschitz continuous with respect to ξ^t without having that *the conditional distribution of* $(\boldsymbol{\xi}_s)_{s=t+1}^T$ *under* $\{\boldsymbol{\xi}^t = \xi^t\}$ *depends continuously on* ξ^t with respect to some appropriate measure of distance. It is illustrated by Example 2.6 of Heitsch, Römisch, and Strugarek (2006) that without such a continuous dependency stability of optimal values in terms of an L^p-distance does not hold in general. Thus, for establishing recursively the continuity of V_t, we need that continuity of V_{t+1} with respect to ξ^{t+1} is passed down to the mapping $\xi^t \mapsto \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, x_t)\middle| \boldsymbol{\xi}^t = \xi^t\right]$. To this end, we introduce for $p \geq 1$ and a given Borel set $A^{t+1} \subset \Xi^{t+1}$ with $\mathbb{P}^{t+1}[A^{t+1}] = 1$ the class of functions

$$\mathbb{F}_p^{A^{t+1}}(\Xi^{t+1}) \triangleq \left\{f : \Xi^{t+1} \to \mathbb{R} : (9) \text{ holds for } \xi^{t+1}, \hat{\xi}^{t+1} \in A^{t+1}\right\}$$

along with the Lipschitz condition

$$|f(\xi^{t+1}) - f(\hat{\xi}^{t+1})| \leq \max\{1, \|\xi^{t+1}\|, \|\hat{\xi}^{t+1}\|\}^{p-1}\|\xi^{t+1} - \hat{\xi}^{t+1}\|. \tag{9}$$

We consider the following distance between probability measures P, Q on Ξ^{t+1}:

$$\zeta_p^{A^{t+1}}(P, Q) \triangleq \sup_{f \in \mathbb{F}_p^{A^{t+1}}(\Xi^{t+1})} \left| \int_{\Xi^{t+1}} f(\xi^{t+1})P(d\xi^{t+1}) - \int_{\Xi^{t+1}} f(\xi^{t+1})Q(d\xi^{t+1})\right|.$$

Recall that with the exception that we disregard the $\mathbb{P}^{t+1}$-null set $\Xi^{t+1} \setminus A^{t+1}$ within the definition of $\mathbb{F}_p^{A^{t+1}}$, the functional $\zeta_p^{A^{t+1}}$ corresponds to the *p-th order Fortet-Mourier distance*; see Rachev (1991) and Römisch (2003). Using this notation, the claimed continuity of the conditional distributions is specified by the following

Assumption 2.6. *There exist constants $W, K > 0$ and $r \geq 0$, such that with*

$$m_t \triangleq 1 + (T-t)(1+r) \quad \text{for } t = 1,\dots,T, \tag{10}$$

the following conditions are fulfilled.

(i) For every $t = 1,\dots,T-1$, every Borel set $A^{t+1} \subset \Xi^{t+1}$ with $\mathbb{P}^{t+1}[A^{t+1}] = 1$, and $\mathbb{P}^t$-a.e. $\xi^t, \hat{\xi}^t \in \Xi^t$

$$\zeta^{A^{t+1}}_{m_{(t+1)}+1}\left(\mathbb{P}\left[\boldsymbol{\xi}^{t+1} \in \cdot \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right], \mathbb{P}\left[\boldsymbol{\xi}^{t+1} \in \cdot \,\middle|\, \boldsymbol{\xi}^t = \hat{\xi}^t\right]\right)$$
$$\leq K \max\left\{1, \|\xi^t\|, \|\hat{\xi}^t\|\right\}^{m_t - 1} \|\xi^t - \hat{\xi}^t\|.$$

(ii) For every $t = 1,\dots,T-1$ and $\mathbb{P}^t$-a.e. $\xi^t \in \Xi^t$

$$\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^T\|\}^{1+T-t} \;\middle|\; \boldsymbol{\xi}^t = \xi^t\right] \leq W \cdot \max\{1, \|\xi^t\|\}^{m_t}.$$

Since the above assumption is crucial for the following continuity and stability results, it is discussed by the following remark.

Remark 2.7. *Condition (i) is related to terms usually related to Markov processes, namely the* coefficient of ergodicity *and the* Feller property*; see, e.g., Dobrushin (1956) and Dynkin (1965), respectively. A similar assumption has been made by Bally, Pagès, and Printems (2005) to ensure stability of an optimal-stopping problem in a Markovian framework and by Mirkov and Pflug (2007) for their study of consistency of tree approximations. It is also made implicitly by Kuhn (2005) by focusing on autoregressive processes. The more involved formulation of Assumption 2.6, allowing for polynomially growing Lipschitz constants, is due to the fact that neither $\langle b_t(\xi_t), x_t\rangle$ nor M_{t+1} are uniformly Lipschitz continuous in ξ_t and x_t, unless both the support Ξ^T and the sets $X_t, t = 1,\dots,T$, are bounded. Indeed, under such a boundedness condition (i) may be significantly simplified; see Remark 2.8 below.*

Lemma A.1 in the Appendix provides conditions on $\boldsymbol{\xi}$, under which both (i) and (ii) hold true. In particular, this is the case if Ξ^T is finite. In the latter case one sees that ζ_p is the optimal value of a linear optimization problem that can be solved numerically to determine the constants K and r.

The following Theorem shows that Assumption 2.6 indeed provides Lipschitz continuity of V_t with respect to ξ^t. We also refer to Proposition 2.7 of Kuhn (2005) which represents a corresponding continuity result.

Theorem 1. *Suppose the Assumptions 2.1, 2.3, and 2.6 are fulfilled. For every $t = 1,\dots,T$ there is a constant $C_t > 0$ and a Borel set $A^t \subset \Xi^t$ with $\mathbb{P}^t[A^t] = 1$ such that*

$$\frac{1}{C_t \max\{1, \|x_{t-1}\|\}} V_t(\,\cdot\,, x_{t-1}) \in \mathbb{F}^{A^t}_{m_t+1}(\Xi^t),$$

holds true for every $x_{t-1} \in X_{t-1}$.

Proof. The assertion holds true for $V_{T+1} \equiv 0$, we show that it follows recursively for $t \leq T$. To this end, we proceed as in the proof of Proposition 2.5 and choose a measurable $\boldsymbol{x}_t^*$ with $\boldsymbol{x}_t^*(\xi^t) \in M_t(x_{t-1}, \xi_t)$, that fulfills (3) and $\|\boldsymbol{x}_t^*(\xi^t)\| \leq L \cdot \max\{1, \|x_{t-1}\|\} \cdot \max\{1, \|\xi^t\|\}$. Thus, we obtain

$$\begin{aligned} &|V_t(\xi^t, x_{t-1}) - V_t(\hat{\xi}^t, x_{t-1})| \\ &= |\langle b_t(\xi_t), \boldsymbol{x}_t^*(\xi^t)\rangle + \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(\xi^t)) \middle| \boldsymbol{\xi}^t = \xi^t\right] \\ &\quad - \inf_{\hat{x}_t \in M_t(x_{t-1}, \hat{\xi}_t)} \left\{\langle b_t(\hat{\xi}_t), \hat{x}_t\rangle + \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{x}_t) \middle| \boldsymbol{\xi}^t = \hat{\xi}^t\right]\right\}|, \end{aligned} \tag{11}$$

which holds, due to Assumption 2.3, for every $\xi^t, \hat{\xi}^t \in A^{t\prime}$ with $\mathbb{P}^t[A^{t\prime}] = 1$ for all $x_{t-1} \in X_{t-1}$. We consider the case when the term under the norm is negative and choose a measurable $\hat{\boldsymbol{x}}_t^*$ with

$$\hat{\boldsymbol{x}}_t^*(\hat{\xi}^t) \in \operatorname{argmin}_{z \in M_t(x_{t-1}, \hat{\xi}_t)} \|z - \boldsymbol{x}_t^*(\xi^t)\|,$$

to obtain the following upper bound for (11):

$$\begin{aligned} &-\langle b_t(\xi_t), \boldsymbol{x}_t^*(\xi^t)\rangle - \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(\xi^t)) \middle| \boldsymbol{\xi}^t = \xi^t\right] \\ &\quad + \langle b_t(\hat{\xi}_t), \hat{\boldsymbol{x}}_t^*(\hat{\xi}^t)\rangle + \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{\boldsymbol{x}}_t^*(\hat{\xi}^t)) \middle| \boldsymbol{\xi}^t = \hat{\xi}^t\right]. \end{aligned} \tag{12}$$

Using linearity of b_t and Lipschitz continuity of M_t, the difference of the scalar product terms can be estimated by

$$\begin{aligned} &|\langle b_t(\xi_t), \boldsymbol{x}_t^*(\xi^t)\rangle - \langle b_t(\hat{\xi}_t), \boldsymbol{x}_t^*(\xi^t)\rangle| + |\langle b_t(\hat{\xi}_t), \boldsymbol{x}_t^*(\xi^t)\rangle - \langle b_t(\hat{\xi}_t), \hat{\boldsymbol{x}}_t^*(\hat{\xi}^t)\rangle| \\ &\leq B\|\xi_t - \hat{\xi}_t\| \cdot L \cdot \max\{1, \|x_{t-1}\|\} \max\left\{1, \|\xi^t\|\right\} \\ &\quad + B\max\{1, \|\hat{\xi}_t\|\} \cdot M \max\{1, \|x_{t-1}\|\}\|\xi_t - \hat{\xi}_t\| \\ &\leq B(L+M)\max\{1, \|x_{t-1}\|\} \max\left\{1, \|\xi^t\|, \|\hat{\xi}^t\|\right\} \|\xi_t - \hat{\xi}_t\|, \end{aligned} \tag{13}$$

The difference of the conditional expectations in (12) is bounded by

$$\begin{aligned} &|\mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(\xi^t)) \;\middle|\; \boldsymbol{\xi}^t = \xi^t\right] - \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(\xi^t)) \;\middle|\; \boldsymbol{\xi}^t = \hat{\xi}^t\right]| \\ &\quad + |\mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(\xi^t)) \;\middle|\; \boldsymbol{\xi}^t = \hat{\xi}^t\right] - \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{\boldsymbol{x}}_t^*(\hat{\xi}^t)) \;\middle|\; \boldsymbol{\xi}^t = \hat{\xi}^t\right]| \\ &\leq C_{t+1}\max\left\{1, \|\boldsymbol{x}_t^*(\xi^t)\|\right\} \zeta^{A^{t+1}}_{m_{(t+1)}+1}\left(\mathbb{P}\left[\boldsymbol{\xi}^{t+1} \in \cdot \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right], \mathbb{P}\left[\boldsymbol{\xi}^{t+1} \in \cdot \,\middle|\, \boldsymbol{\xi}^t = \hat{\xi}^t\right]\right) \\ &\quad + \mathbb{E}\left[[V_{t+1}]^x_{Lip}(\boldsymbol{\xi}^{t+1}) \;\middle|\; \boldsymbol{\xi}^t = \hat{\xi}^t\right] M \max\{1, \|x_{t-1}\|\} \cdot \|\xi^t - \hat{\xi}^t\|, \end{aligned} \tag{14}$$

whereby the last inequality follows from the assertion for V_{t+1}, Proposition 2.5, and the Lipschitz continuity of M_t. This estimate holds true for every $\xi^t, \hat{\xi}^t \in A^{t\prime\prime}$ for all $x_{t-1} \in X_{t-1}$, where $A^{t\prime\prime}$ denotes the sets on which the assertions of Proposition 2.5 hold. Applying now condition (i) of Assumption 2.6 and the estimate (7), we see that the sum (14) does not exceed

$$\begin{aligned} &KC_{t+1}\max\left\{1, \|\boldsymbol{x}_t^*(\xi^t)\|\right\} \max\left\{1, \|\xi^t\|, \|\hat{\xi}^t\|\right\}^{m_t - 1} \|\xi^t - \hat{\xi}^t\| \\ &\quad + \bar{M} \cdot \mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^T\|\}^{1+T-t} \;\middle|\; \boldsymbol{\xi}^t = \xi^t\right] M \max\{1, \|x_{t-1}\|\} \cdot \|\xi^t - \hat{\xi}^t\| \end{aligned}$$

for every $\xi^t, \hat{\xi}^t \in A^{t'''}$. Thereby, $A^{t'''}$ denotes the set of $\mathbb{P}^t$-probability one on which Assumption 2.6 holds.
From condition (ii) of Assumption 2.6 and the boundedness of $\|\boldsymbol{x}_t^*\|$, we conclude that the latter sum is again bounded from above by

$$(15) \qquad \big(KC_{t+1}L + \bar{M}WM\big) \max\{1, \|x_{t-1}\|\} \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^{m_t} \cdot \|\xi^t - \hat{\xi}^t\|.$$

The upper bounds (13) and (15) remain valid if the term under the norm in (11) is positive. Piecing this all together, the assertion for V_t follows with $A^t \triangleq A^{t'} \cap A^{t''} \cap A^{t'''}$ and the Lipschitz constant C_t can be chosen by collecting the constants from (13) and (15), i.e.,

$$C_t \triangleq B(M + L) + KC_{t+1}L + \bar{M}WM.$$

□

In the following remark we sketch what can be simplified whenever the sets X_t and Ξ^T are bounded.

Remark 2.8. *The constant m_t is chosen to be equal to the growth rate of the term $\max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}$ within an upper bound of* (14). *Assuming boundedness of the sets X_t for $t = 1, \ldots, T$ allows one to estimate the term $\max\{1, \|\boldsymbol{x}_t^*(\xi^t)\|\}$ in the first summand of* (14) *by some constant instead of estimating it by $\max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}$. Consequently, one can allow the growth rate of the ζ-terms in* (14) *and in Assumption 2.6 to increase from $m_t - 1$ to m_t.*

If the set Ξ^T is bounded as well, then $[V_{t+1}]^x_{Lip}(\boldsymbol{\xi}^{t+1})$ is bounded by a constant, condition (i) of Assumption 2.6 may be simplified to

$$\zeta_1^{A^{t+1}}\left(\mathbb{P}\left[\boldsymbol{\xi}^{t+1} \in \cdot \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right], \mathbb{P}\left[\boldsymbol{\xi}^{t+1} \in \cdot \,\middle|\, \boldsymbol{\xi}^t = \hat{\xi}^t\right]\right) \leq K\|\xi^t - \hat{\xi}^t\|,$$

condition (ii) of Assumption 2.6 may be omitted, and the assertion of Theorem 1 can be written as $(1/C_t)\, V_t(\,\cdot\,, x_{t-1}) \in \mathbb{F}_1^{A^t}(\Xi^t)$, i.e., $(\xi^t, x_{t-1}) \mapsto V_t(\xi^t, x_{t-1})$ is uniformly Lipschitz continuous.

The optimality and boundedness conditions (3)–(5) as well as the continuity properties claimed in Assumption 2.6 and Theorem 1 hold on some Borel set $A \subset \Xi^T$ with $\mathbb{P}^T[A] = 1$. Since an approximation $\tilde{\boldsymbol{\xi}}$ may have its support in the set $\Xi^T \setminus A$, it is reasonable to modify the considered random variables on this $\mathbb{P}^T$-null set to appropriate versions which fulfill the claimed properties for *every* $\xi^T \in \Xi^T$. To this end, we assume without loss of generality that the set A is dense in Ξ^T. For every $\hat{\xi}^T \in \Xi^T \setminus A$ we then consider a sequence $(\xi^T_{(n)})_{n\in\mathbb{N}} \subset A$ converging to $\hat{\xi}^T$ as n goes to infinity. The recourse function and the regular conditional distributions are modified in $\hat{\xi}^T$ by setting $V_t(\hat{\xi}^t, x_{t-1}) \triangleq \lim_{n\to\infty} V_t(\xi^t_{(n)}, x_{t-1})$ and $\mathbb{E}[g(\boldsymbol{\xi}^{t+1})|\boldsymbol{\xi}^t = \hat{\xi}^t] \triangleq \lim_{n\to\infty} \mathbb{E}[g(\boldsymbol{\xi}^{t+1})|\boldsymbol{\xi}^t = \xi^t_{(n)}]$ for every Lipschitz continuous mapping g. A bounded optimal solution $\boldsymbol{x}^*$ can be appropriately modified in $\hat{\xi}^t$ by considering a subsequence $(\xi^T_{(n_k)})_{k\in\mathbb{N}}$ such that $\boldsymbol{x}_t^*(\xi^t_{(n_k)})$ converges toward some $z_t(\hat{\xi}^t) \in X_t$ for

$t = 1, \ldots, T$. Then we put $\boldsymbol{x}_t^*(\hat{\xi}^t) \triangleq z_t(\hat{\xi}^t)$ and we obtain that the above stated conditions and properties indeed hold for every $\xi^T \in \Xi^T$.

3 Approximations

Whenever an auxiliary process $\tilde{\boldsymbol{\xi}}$ is expected to approximate $\boldsymbol{\xi}$ with regard to the optimization problem (2), it is indispensable that $\tilde{\boldsymbol{\xi}}$ is nonanticipative with respect to $\boldsymbol{\xi}$. This is illustrated, for the sake of completeness, by Example A.3 in the Appendix. Nonanticipativity is ensured in the following by

Definition 3.1. *A stochastic process $\tilde{\boldsymbol{\xi}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called an* approximation of $\boldsymbol{\xi}$*, if there exist Borel-measurable mappings*

$$f_t : \Xi^t \to \Xi_t, \text{ for } t = 1, \ldots, T,$$

fulfilling the following conditions:

(i) $\tilde{\boldsymbol{\xi}}_t = f_t(\boldsymbol{\xi}^t)$ *for* $t = 1, \ldots T$,

(ii) $f^T(\Xi^T) \subset \Xi^T$,

(iii) $f_1(\xi_1) = \xi_1$ *for every* $\xi_1 \in \Xi_1$, *and*

(iv) $f^T(\boldsymbol{\xi}) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ *for every* $p \in [1, \infty)$.

Thereby, $f^t(\xi^t)$ denotes the vector $(f_i(\xi^i))_{i=1}^t \in \mathbb{R}^{s \cdot t}$, for $t = 1, \ldots, T$. In the following, we use the notation f for the mapping $f^T(\cdot)$.

Remark 3.2. *The nonanticipativity condition (i) is equivalent to $\sigma(\boldsymbol{\xi}^t)$-measurability of the random variable $\tilde{\boldsymbol{\xi}}_t$. Condition (ii) ensures that f^T maps onto realizations $\xi^T \in \Xi^T$ of the initial process and thus implies that the restriction of a decision $\boldsymbol{x}(\cdot) \in \mathcal{S}(\boldsymbol{\xi})$ on the set $f^T(\Xi^T)$ is a $\tilde{\boldsymbol{\xi}}$-feasible decision. The integrability condition (iv) is assumed again for the sake of simplicity. For the following results, it suffices that $f^T(\boldsymbol{\xi}) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for a constant $p \in \mathbb{R}_+$ that is sufficiently large.*

The following proposition relies heavily on the continuity of the recourse function stated in Theorem 1. It is shown that, although an optimal decision $\boldsymbol{x}^*(\cdot)$ is not continuous in general, its expected costs can be approximated by the decision $\boldsymbol{x}^*(f(\cdot))$ (which is piecewise constant whenever $\tilde{\boldsymbol{\xi}}$ has finite support). Although $\boldsymbol{x}^*(f(\cdot))$ may fail to fulfill the time-coupling constraints (1) with respect to $\boldsymbol{\xi}$, it can be used to construct a feasible decision. This will be carried out in the next section.

Proposition 3.3. *Consider an optimal decision $\boldsymbol{x}^*$ which is bounded in the sense of* (5) *and an approximation mapping f according to Definition 3.1.*
Then there exists a constant $D > 0$ such that the following estimate holds:

$$|\varphi(\boldsymbol{x}^*(\boldsymbol{\xi}), \boldsymbol{\xi}) - \varphi(\boldsymbol{x}^*(f(\boldsymbol{\xi})), \boldsymbol{\xi})| \leq D\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}\|, \|f(\boldsymbol{\xi})\|\}^{m_1} \cdot \|\boldsymbol{\xi} - f(\boldsymbol{\xi})\|\right], \tag{16}$$

where the constant m_1 is defined by (10).

Proof. Due to $f_1(\boldsymbol{\xi}_1) = \boldsymbol{\xi}_1$, we have to estimate

$$\left|\mathbb{E}\left[\sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t^*(\boldsymbol{\xi}^t)\rangle\right] - \mathbb{E}\left[\sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t^*(f^t(\boldsymbol{\xi}^t))\rangle\right]\right|.$$

By optimality of $\boldsymbol{x}^*$, the first expectation is equal to $\mathbb{E}\left[V_2(\boldsymbol{\xi}^2, \boldsymbol{x}_1^*)\right]$ and it follows from Theorem 1 and boundedness of $\boldsymbol{x}_1^*$ (and $\boldsymbol{x}_0^* \triangleq 1$) that

$$\mathbb{E}\left[\left|V_2(\boldsymbol{\xi}^2, \boldsymbol{x}_1^*) - V_2(f^2(\boldsymbol{\xi}^2), \boldsymbol{x}_1^*)\right|\right] \leq LC_2\mathbb{E}\left[\max\left\{1, \|\boldsymbol{\xi}^2\|, \|f^2(\boldsymbol{\xi}^2)\|\right\}^{m_2} \|\boldsymbol{\xi}^2 - f^2(\boldsymbol{\xi}^2)\|\right] \tag{17}$$

Thus, it remains to estimate

$$\left|\mathbb{E}\left[V_2(f^2(\boldsymbol{\xi}^2), \boldsymbol{x}_1^*) - \sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t^*(f^t(\boldsymbol{\xi}^t))\rangle\right]\right|. \tag{18}$$

To this end, we consider the following inequality

$$\left|\mathbb{E}\left[V_2(f^2(\boldsymbol{\xi}^2), \boldsymbol{x}_1^*) - \sum_{s=2}^{t-1}\langle b_s(\boldsymbol{\xi}_s), \boldsymbol{x}_s^*(f^s(\boldsymbol{\xi}^s))\rangle - V_t(f^t(\boldsymbol{\xi}^t), \boldsymbol{x}_{t-1}^*(f^{t-1}(\boldsymbol{\xi}^{t-1})))\right]\right| \leq D_t, \tag{19}$$

whose left side coincides with (18) for $t = T + 1$. It holds trivially for $t = 2$ with $D_2 = 0$ and we assume that it is also true for some $t \in \{2, \dots, T\}$ with a constant $D_t \geq 0$. To prove it recursively for $t + 1$, we have to find an upper bound for

$$\left|\mathbb{E}\left[V_t(f^t(\boldsymbol{\xi}^t), \boldsymbol{x}_{t-1}^*(f^{t-1}(\boldsymbol{\xi}^{t-1}))) - \langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t^*(f^t(\xi^t))\rangle - V_{t+1}(f^{t+1}(\boldsymbol{\xi}^{t+1}), \boldsymbol{x}_t^*(f^t(\xi^t)))\right]\right|. \tag{20}$$

To this end, we use again $\boldsymbol{x}^*$'s optimality to expand the first summand:

$$\begin{aligned} &\mathbb{E}[V_t(f^t(\boldsymbol{\xi}^t), \boldsymbol{x}_{t-1}^*(f^{t-1}(\boldsymbol{\xi}^{t-1})))] \\ &= \int \langle b_t(f_t(\xi^t)), \boldsymbol{x}_t^*(f^t(\xi^t))\rangle + \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(f^t(\xi^t)))\middle|\, \boldsymbol{\xi}^t = f^t(\xi^t)\right] \mathbb{P}^t(d\xi^t). \end{aligned} \tag{21}$$

Now, to estimate (20), we have to replace $b_t(f_t(\xi^t))$ by $b_t(\xi^t)$. The Lipschitz continuity of $b_t(\cdot)$ implies

$$|\langle b_t(f_t(\xi^t)), \boldsymbol{x}_t^*(f^t(\xi^t))\rangle - \langle b_t(\xi_t), \boldsymbol{x}_t^*(f^t(\xi^t))\rangle| \leq B \cdot \|\boldsymbol{x}_t^*(f^t(\xi^t))\| \cdot \|\xi^t - f^t(\xi^t)\|.$$

To estimate the difference of the V_{t+1}-terms in (20) and (21), we add and subtract the term $\mathbb{E}\left[V_{t+1}(f^{t+1}(\boldsymbol{\xi}^{t+1}), \boldsymbol{x}_t^*(f^t(\xi^t))) \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right]$ and use the triangle inequality to estimate

$$\begin{aligned}
&\left|\mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(f^t(\xi^t))) \,\middle|\, \boldsymbol{\xi}^t = f^t(\xi^t)\right] - \mathbb{E}\left[V_{t+1}(f^{t+1}(\boldsymbol{\xi}^{t+1}), \boldsymbol{x}_t^*(f^t(\xi^t))) \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right]\right| \\
&\leq \left|\mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(f^t(\xi^t))) \,\middle|\, \boldsymbol{\xi}^t = f^t(\xi^t)\right] - \mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(f^t(\xi^t))) \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right]\right| \\
&\quad + \left|\mathbb{E}\left[V_{t+1}(\boldsymbol{\xi}^{t+1}, \boldsymbol{x}_t^*(f^t(\xi^t))) \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right] - \mathbb{E}\left[V_{t+1}(f^{t+1}(\boldsymbol{\xi}^{t+1}), \boldsymbol{x}_t^*(f^t(\xi^t))) \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right]\right|.
\end{aligned}$$

By applying Theorem 1 and Assumption 2.6 we conclude that this term is bounded for $\mathbb{P}^t$-almost every ξ^t by

$$\begin{aligned}
&KC_{t+1} \max\left\{1, \|\boldsymbol{x}_t^*(f^t(\xi^t))\|\right\} \max\{1, \|\xi^t\|, \|f^t(\xi^t)\|\}^{m_t-1} \|\xi^t - f^t(\xi^t)\| \\
&+ C_{t+1} \max\left\{1, \|\boldsymbol{x}_t^*(f^t(\xi^t))\|\right\} \\
&\quad \cdot \mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^{t+1}\|, \|f^{t+1}(\boldsymbol{\xi}^{t+1})\|\}^{m_{t+1}} \|\boldsymbol{\xi}^{t+1} - f^{t+1}(\boldsymbol{\xi}^{t+1})\| \,\middle|\, \boldsymbol{\xi}^t = \xi^t\right] \\
&\leq (K+1) C_{t+1} L^t \mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^{t+1}\|, \|f^{t+1}(\boldsymbol{\xi}^{t+1})\|\}^{m_t+t-1} \|\boldsymbol{\xi}^{t+1} - f^{t+1}(\boldsymbol{\xi}^{t+1})\| | \boldsymbol{\xi}^t = \xi^t\right],
\end{aligned}$$

where the inequality follows from boundedness of $\boldsymbol{x}_t^*$ and the relation $m_{t+1} \leq m_t - 1$. Integration with respect to $\mathbb{P}^t(d\xi^t)$ and combining these estimates with (21) entails that (20) does not exceed

$$(22) \quad (B + (K+1)C_{t+1}) L^t \mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^{t+1}\|, \|f^{t+1}(\boldsymbol{\xi}^{t+1})\|\}^{m_t+t-1} \cdot \|\boldsymbol{\xi}^{t+1} - f^{t+1}(\boldsymbol{\xi}^{t+1})\|\right],$$

Hence, (19) holds for $t+1$ with D_{t+1} being equal to the sum of D_t and (22).
Due to the fact that both $m_t + t - 1$ and m_2 are smaller than m_1, the sum of (17) and (18) does not exceed

$$D\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}\|, \|f(\boldsymbol{\xi})\|\}^{m_1} \cdot \|\boldsymbol{\xi} - f(\boldsymbol{\xi})\|\right]$$

with $D \triangleq LC_2 + D_{T+1}$. This completes the proof. □

4 Calm Decisions

One of the main difficulties in establishing the stability of the optimal value $v(\boldsymbol{\xi})$ with respect to perturbations of $\boldsymbol{\xi}$ is that optimal solutions do not depend continuously on the realization of $\boldsymbol{\xi}$, in general. Furthermore, the gap between $v(\boldsymbol{\xi})$ and the minimal expected costs which can be realized by, e.g., Lipschitz continuous solutions may be hard to estimate. In this section we shall introduce specific *calm* decisions and estimate the minimal expected costs realized by those decisions.

We consider an optimal decision $\boldsymbol{x}^*$ which is bounded in the sense of (5). The *calm modification of* $\boldsymbol{x}^*$ is defined by

$$\bar{\boldsymbol{x}}_1^* \triangleq \boldsymbol{x}_1^* \quad \text{and} \quad \bar{\boldsymbol{x}}_t^*(\xi^t) \in \operatorname{argmin}_{z \in M_t(\bar{\boldsymbol{x}}_{t-1}^*(\xi^{t-1}), \xi_t)} \|\boldsymbol{x}_t^*(f^t(\xi^t)) - z\| \quad \text{for } t = 2, \ldots, T,$$

where, again due to Theorem 14.37 of Rockafellar and Wets (1998), the latter mappings can be chosen to be measurable. Observe that $\boldsymbol{x}_t^*$ and $\bar{\boldsymbol{x}}_t^*$ coincide on the set $f^t(\Xi^t)$, i.e.,

$$\bar{\boldsymbol{x}}_t^*(f^t(\xi^t)) = \boldsymbol{x}_t^*(f^t(\xi^t)) \text{ for every } \xi^t \in \Xi^t. \tag{23}$$

Due to the Lipschitz continuity of M_t, the local variability of $\bar{\boldsymbol{x}}_t^*(\cdot)$ in ξ^t can be estimated recursively and $\bar{\boldsymbol{x}}_t^*(\cdot)$ is indeed *calm locally around* $f^t(\xi^t)$ *for every* $\xi^t \in \Xi^t$ in the following sense.

Proposition 4.1. *For every* $t = 1, \dots, T$ *and every* $\xi^T \in \Xi^T$ *we have*

$$\|\bar{\boldsymbol{x}}_t^*(\xi^t) - \bar{\boldsymbol{x}}_t^*(f^t(\xi^t))\| \leq (T-1)M^{T-1} \max\{1, \|\xi^T\|, \|f^T(\xi^T)\|\}^{T-1} \|\xi^T - f^T(\xi^T)\|. \tag{24}$$

Proof. For $t = 1$, the difference on the left side of (24) vanishes. For $t > 1$ we use the identity (23) and the definition of $\bar{\boldsymbol{x}}_t^*(\xi^t)$ to write

$$\|\bar{\boldsymbol{x}}_t^*(\xi^t) - \bar{\boldsymbol{x}}_t^*(f^t(\xi^t))\| = \inf_{z \in M_t(\bar{\boldsymbol{x}}_{t-1}^*(\xi^{t-1}), \xi_t)} \|z - \bar{\boldsymbol{x}}_t^*(f^t(\xi^t))\|. \tag{25}$$

Using the inclusion $\bar{\boldsymbol{x}}_t^*(f^t(\xi^t)) \in M_t(\bar{\boldsymbol{x}}_{t-1}^*(f^{t-1}(\xi^{t-1})), f_t(\xi^t))$, we obtain that the right-hand side of (25) is not greater than the Pompeiu-Hausdorff distance of $M_t(\bar{\boldsymbol{x}}_{t-1}^*(\xi^{t-1}), \xi^t)$ and $M_t(\bar{\boldsymbol{x}}_{t-1}^*(f^{t-1}(\xi^{t-1})), f_t(\xi^t))$. We then apply the triangle inequality with respect to this metric and use the Lipschitz continuity of M_t to conclude that the right-hand side of (25) is bounded from above by

$$\begin{aligned} &M \max\{1, \|\xi^t\|\} \, \|\bar{\boldsymbol{x}}_{t-1}^*(\xi^{t-1}) - \bar{\boldsymbol{x}}_{t-1}^*(f^{t-1}(\xi^{t-1}))\| \\ &\quad + M \max\{1, \|\bar{\boldsymbol{x}}_{t-1}^*(f^{t-1}(\xi^{t-1}))\|\} \, \|f^t(\xi^t) - \xi^t\|. \end{aligned}$$

By boundedness of $\boldsymbol{x}_{t-1}$, the latter sum does not exceed

$$\begin{aligned} &M \max\{1, \|\xi^t\|\} \, \|\bar{\boldsymbol{x}}_{t-1}^*(\xi^{t-1}) - \bar{\boldsymbol{x}}_{t-1}^*(f^{t-1}(\xi^{t-1}))\| \\ &\quad + ML \max\{1, \|f^{t-1}(\xi^{t-1})\|\}^{t-1} \, \|f^t(\xi^t) - \xi^t\|. \end{aligned}$$

Recursively, we obtain that the left side of (24) is bounded by

$$L \sum_{i=2}^{t} M^{t+1-i} \max\{1, \|f^{i-1}(\xi^{i-1})\|\}^{i-1} \; \max\{1, \|\xi^t\|\}^{t-i} \; \|\xi^i - f^i(\xi^i)\|.$$

The assertion follows by a straightforward estimate. □

By combining Proposition 3.3 and Propostion 4.1 one immediately concludes the following theorem, which shows that the difference of the expected costs generated by $\boldsymbol{x}^*$ and $\bar{\boldsymbol{x}}^*$ can be estimated in terms of the deviation between $\boldsymbol{\xi}$ and $f(\boldsymbol{\xi})$.

Theorem 2. *Suppose the Assumptions 2.1, 2.3, and 2.6 are fulfilled. Consider an optimal decision* $\boldsymbol{x}^*$ *which is bounded in the sense of* (5) *and its calm modification* $\bar{\boldsymbol{x}}^*$*.*
Then there exists a constant $C > 0$ *such that the following estimate holds*

$$|\mathbb{E}\varphi(\boldsymbol{x}^*(\boldsymbol{\xi}), \boldsymbol{\xi}) - \mathbb{E}\varphi(\bar{\boldsymbol{x}}^*(\boldsymbol{\xi}), \boldsymbol{\xi})| \leq C\,\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}\|, \|f(\boldsymbol{\xi})\|\}^{m_1} \cdot \|\boldsymbol{\xi} - f(\boldsymbol{\xi})\|\right],$$

where the constant m_1 *is defined by* (10).

Proof. To proof the assertion, we have to estimate the following term:

$$\left|\mathbb{E}\left[\sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t^*(\boldsymbol{\xi}^t)\rangle - \sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \bar{\boldsymbol{x}}_t^*(\boldsymbol{\xi}^t)\rangle\right]\right|. \tag{26}$$

Recall that, by Proposition 3.3, $\boldsymbol{x}^*(\boldsymbol{\xi})$ and $\boldsymbol{x}^*(f(\boldsymbol{\xi}))$ produce comparable costs. On the other hand, the difference between $\boldsymbol{x}^*(f(\boldsymbol{\xi}))$ and $\bar{\boldsymbol{x}}^*(\boldsymbol{\xi})$ can be estimated due to the calmness of the latter decision. Thus, we add and subtract the term

$$\mathbb{E}\left[\sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t^*(f^t(\boldsymbol{\xi}^t))\rangle\right]$$

to the expectation within (26). Using then the triangle inequality as well as Proposition 3.3, we conclude that (26) is not greater than the sum of

$$\left|\mathbb{E}\left[\sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t^*(f^t(\boldsymbol{\xi}^t))\rangle - \sum_{t=2}^{T}\langle b_t(\boldsymbol{\xi}_t), \bar{\boldsymbol{x}}_t^*(\boldsymbol{\xi}^t)\rangle\right]\right| \tag{27}$$

and the right-hand side of (16). It thus remains to estimate (27). By applying identity (23) as well as the calmness property of $\bar{\boldsymbol{x}}^*$ from Proposition 4.1, we obtain the following upper bound:

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=2}^{T} B\max\{1, \|\boldsymbol{\xi}^t\|\}(T-1)M^{T-1}\max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^{T-1}\|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\|\right]\\ &\leq (T-1)^2 BM^{T-1}\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^T\|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\|\right].\end{aligned}$$

Finally, the sum of the latter term and the right-hand side of (16) is smaller than

$$C\,\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^{m_1} \cdot \|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\|\right],$$

with a constant $C \triangleq D + (T-1)^2\, B\, M^{T-1}$. □

5 Stability

In order to address the question of stability, we have to consider the following issue. Although we assume the existence of bounded optimal solutions to the initial problem (2), the perturbed problem may be unbounded. This is illustrated, for the sake of completeness, by Example A.4 in the Appendix. Heitsch, Römisch, and Strugarek (2006) avoid such unfavorable cases by their Assumption (A2) of level-boundedness of the objective, *locally around* $\boldsymbol{\xi}$. We proceed by assuming that $\tilde{\boldsymbol{\xi}}$ fulfills Assumption 2.3 too, i.e., the perturbed problem $v(\tilde{\boldsymbol{\xi}})$ admits a bounded optimal solution. We now state our main result.

Theorem 3. *Suppose the Assumptions 2.1, 2.3, and 2.6 are fulfilled. Let $\tilde{\boldsymbol{\xi}}$ be an approximation of $\boldsymbol{\xi}$ according to Definition 3.1, which fulfills Assumption 2.3, too, and consider the constant m_1 defined by* (10).
Then there exists a constant $\gamma > 0$, such that

$$\left|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})\right| \leq \gamma\, \mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}\|, \|\tilde{\boldsymbol{\xi}}\|\}^{m_1} \cdot \|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|\right]$$

holds.

Proof. We denote the approximation mapping corresponding to $\tilde{\boldsymbol{\xi}}$ by f and consider a bounded optimal decision $\boldsymbol{x}^* \in \mathcal{S}(\boldsymbol{\xi})$ and the corresponding calm modification $\bar{\boldsymbol{x}}^*$ from Section 4.
Applying Theorem 2 yields the following inequality

$$\begin{aligned} v(\tilde{\boldsymbol{\xi}}) - v(\boldsymbol{\xi}) &= v(\tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\boldsymbol{x}^*(\boldsymbol{\xi}), \boldsymbol{\xi}) \\ &\leq v(\tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\bar{\boldsymbol{x}}^*(\boldsymbol{\xi}), \boldsymbol{\xi}) + C\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}\|, \|\tilde{\boldsymbol{\xi}}\|\}^{m_1} \cdot \|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|\right]. \end{aligned} \tag{28}$$

Since the restriction of $\bar{\boldsymbol{x}}^*$ on $f(\Xi)$ is contained in $\mathcal{S}(\tilde{\boldsymbol{\xi}})$, we can write

$$\begin{aligned} v(\tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\bar{\boldsymbol{x}}^*(\boldsymbol{\xi}), \boldsymbol{\xi}) &\leq \mathbb{E}\varphi(\bar{\boldsymbol{x}}^*(\tilde{\boldsymbol{\xi}}), \tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\bar{\boldsymbol{x}}^*(\boldsymbol{\xi}), \boldsymbol{\xi}) \\ &= \sum_{t=2}^{T} \mathbb{E}\left[\langle b_t(\tilde{\boldsymbol{\xi}}_t) - b_t(\boldsymbol{\xi}_t), \bar{\boldsymbol{x}}_t^*(\tilde{\boldsymbol{\xi}}^t)\rangle + \langle b_t(\boldsymbol{\xi_t}), \bar{\boldsymbol{x}}_t^*(\tilde{\boldsymbol{\xi}}^t) - \bar{\boldsymbol{x}}_t^*(\boldsymbol{\xi}^t)\rangle\right] \\ &\leq B\sum_{t=2}^{T} \mathbb{E}\left[\|\tilde{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t\| \, \|\bar{\boldsymbol{x}}_t^*(\tilde{\boldsymbol{\xi}}^t)\| + \max\{1, \|\boldsymbol{\xi}_t\|\} \, \|\bar{\boldsymbol{x}}_t^*(\tilde{\boldsymbol{\xi}}^t) - \bar{\boldsymbol{x}}_t^*(\boldsymbol{\xi}^t)\|\right]. \end{aligned} \tag{29}$$

Due to the fact that $\bar{\boldsymbol{x}}^*$ and $\boldsymbol{x}^*$ coincide on the set $f(\Xi^T)$, see (23), we obtain that $\bar{\boldsymbol{x}}^*$ fulfills the boundedness condition (5) on $f(\Xi^T)$. Using this boundedness as well as the calmness of $\bar{\boldsymbol{x}}^*$, each of the $T-1$ summands in (29) can be estimated. Thus, (29) is bounded from above by

$$H\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}\|, \|\tilde{\boldsymbol{\xi}}\|\}^{T} \cdot \|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|\right], \tag{30}$$

with an appropriate constant $H > 0$ and we can use the relation $T \leq m_1$ to obtain

$$v(\tilde{\boldsymbol{\xi}}) - v(\boldsymbol{\xi}) \leq (C + H)\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}\|, \|\tilde{\boldsymbol{\xi}}\|\}^{m_1} \cdot \|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|\right].$$

Now, we consider a bounded optimal decision $\tilde{\boldsymbol{x}}^*$ of $v(\tilde{\boldsymbol{\xi}})$. Following exactly the construction of Section 4, we obtain a decision $\bar{\tilde{\boldsymbol{x}}}^* \in \mathcal{S}(\boldsymbol{\xi})$ that is calm in the sense of Proposition 4.1 and whose restriction on $f(\Xi^T)$ is optimal for $v(\tilde{\boldsymbol{\xi}})$. As in (29), it follows that

$$\begin{aligned} v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}}) &\leq \mathbb{E}\varphi(\bar{\tilde{\boldsymbol{x}}}^*(\boldsymbol{\xi}), \boldsymbol{\xi}) - \mathbb{E}\varphi(\bar{\tilde{\boldsymbol{x}}}^*(\tilde{\boldsymbol{\xi}}), \tilde{\boldsymbol{\xi}}) \\ &\leq B\sum_{t=2}^{T} \mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}_t\}\, \| \|\bar{\tilde{\boldsymbol{x}}}_t^*(\tilde{\boldsymbol{\xi}}^t) - \bar{\tilde{\boldsymbol{x}}}_t^*(\boldsymbol{\xi}^t)\| + \|\tilde{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t\| \, \|\bar{\tilde{\boldsymbol{x}}}_t^*(\tilde{\boldsymbol{\xi}}^t)\|\right] \\ &\leq H\mathbb{E}\left[\max\{1, \|\boldsymbol{\xi}^T\|, \|\tilde{\boldsymbol{\xi}}^T\|\}^T \cdot \|\boldsymbol{\xi}^T - \tilde{\boldsymbol{\xi}}^T\|\right]. \end{aligned}$$

Applying again $T \leq m_1$ and setting $\gamma \triangleq C + H$ completes the proof. □

Remark 5.1. *Since the purpose of this paper is to establish a stability result rather than the development of new approximation techniques, we restrict ourselves to refer to existing approaches based on conditional or unconditional clustering, that can be used to control the upper bound on* $|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})|$ *from Theorem 3. We mention here the recent approaches of Heitsch and Römisch (2007), Bally, Pagès, and Printems (2005), Hochreiter and Pflug (2007), and Pennanen (2007).*

Other approximation techniques, e.g., those proposed by Høyland and Wallace (2001) and Mirkov and Pflug (2007), are not based on projections of the initial process $\boldsymbol{\xi}$*. Consequently, neither the joint distribution of* $\boldsymbol{\xi}$ *and* $\tilde{\boldsymbol{\xi}}$ *nor the underlying probability spaces are necessarily specified. However, under some weak regularity conditions our results may be applied to such cases as well. Indeed, one may choose the sample space* $(\Xi^T \times \Xi^T, \mathcal{B}(\Xi^T) \otimes \mathcal{B}(\Xi^T), \mathbb{P}^T \otimes \tilde{\mathbb{P}}^T)$ *as underlying probability space for both processes and one may define a nonanticipative coupling mapping* $f : \Xi^T \to \Xi^T$ *by, e.g., using successive projections. Thereby,* $\mathcal{B}(\Xi^T)$ *denotes the Borel sets of* Ξ^T*, and* $\tilde{\mathbb{P}}^T$ *denotes the distribution of the approximating process* $\tilde{\boldsymbol{\xi}}$*. Determining such a mapping* f *is closely related to* L^p-minimal metrics *and to* mass transportation problems*, see also Remark 2.3 of Heitsch, Römisch, and Strugarek (2006).*

Appendix

The following lemma provides conditions under which the conditions of Assumption 2.6 hold true.

Lemma A.1. *Assume the dynamics of the process* $\boldsymbol{\xi}$ *are given by the following scheme:*

$$\boldsymbol{\xi}_{t+1} = g_t(\boldsymbol{\xi}^t, \boldsymbol{\varepsilon}_{t+1}), \tag{31}$$

where $\boldsymbol{\varepsilon}_{t+1}$ *is a* $\mathbb{R}^n$*-valued random variable that is independent of* $\boldsymbol{\xi}^t$ *and* g_t *are measurable mappings from* $\mathbb{R}^{s\cdot t} \times \mathbb{R}^n$ *to* $\mathbb{R}^s$ *which satisfy the following Lipschitz and linear growth conditions:*

(i) $\|g_t(\xi^t, \varepsilon) - g_t(\hat{\xi}^t, \varepsilon)\| \leq \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^r \, \|\xi^t - \hat{\xi}^t\| \, h(\|\varepsilon\|),$,

(ii) $\|g_t(\xi^t, \varepsilon)\| \leq \max\{1, \|\xi^t\|\} \, k(\|\varepsilon\|),$

for all $\varepsilon \in \mathbb{R}^n$ *and* $\xi^t, \hat{\xi}^t \in \mathbb{R}^{s\cdot t}$*, some constant* $r \geq 1$ *and Borel-measurable mappings* $h, k \geq 1$*, such that* $h(\|\boldsymbol{\varepsilon}_{t+1}\|)$ *and* $k(\|\boldsymbol{\varepsilon}_{t+1}\|)$ *are in* L^p *for every* $p \in [1, +\infty)$*.*
Then $\boldsymbol{\xi}$ *fulfills both conditions of Assumption 2.6 with the constants*

$$K \triangleq \mathbb{E}\left[k(\|\boldsymbol{\varepsilon}_{t+1}\|)^{m_1} h(\|\boldsymbol{\varepsilon}_{t+1}\|)\right] \quad \text{and} \quad W \triangleq \mathbb{E}\left[\prod_{i=t+1}^{T} k(\|\boldsymbol{\varepsilon}_i\|)^{1+T-t}\right].$$

Proof. Consider $f \in \mathbb{F}_{1+m_{t+1}}(\Xi^{t+1})$. Then we obtain

$$\begin{aligned}
&\left|\mathbb{E}\left[f(\boldsymbol{\xi}^{t+1})\middle|\, \boldsymbol{\xi}^t = \xi^t\right] - \mathbb{E}\left[f(\boldsymbol{\xi}^{t+1})\middle|\, \boldsymbol{\xi}^t = \hat{\xi}^t\right]\right| \\
&= \left|\mathbb{E}\left[f(g_t(\xi^t, \boldsymbol{\varepsilon}_{t+1}))\right] - \mathbb{E}\left[f(g_t(\hat{\xi}^t, \boldsymbol{\varepsilon}_{t+1}))\right]\right| \\
&\leq \mathbb{E}\left[\max\left\{1, \|g_t(\xi^t, \boldsymbol{\varepsilon}_{t+1})\|, \|g_t(\hat{\xi}^t, \boldsymbol{\varepsilon}_{t+1})\|\right\}^{m_{t+1}} \|g_t(\xi^t, \boldsymbol{\varepsilon}_{t+1}) - g_t(\hat{\xi}^t, \boldsymbol{\varepsilon}_{t+1})\|\right] \\
&\leq \mathbb{E}\left[\max\left\{1, \|g_t(\xi^t, \boldsymbol{\varepsilon}_{t+1})\|, \|g_t(\hat{\xi}^t, \boldsymbol{\varepsilon}_{t+1})\|\right\}^{m_{t+1}} h(\|\boldsymbol{\varepsilon}_{t+1}\|)\right] \\
&\quad \cdot \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^r \|\xi^t - \hat{\xi}^t\| \\
&\leq \mathbb{E}\left[\max\left\{1, \|\xi^t\|, \|\hat{\xi}^t\|\right\}^{m_{t+1}} k(\|\boldsymbol{\varepsilon}_{t+1}\|)^{m_{t+1}} h(\|\boldsymbol{\varepsilon}_{t+1}\|)\right] \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^r \|\xi^t - \hat{\xi}^t\| \\
&= \mathbb{E}\left[k(\|\boldsymbol{\varepsilon}_{t+1}\|)^{m_{t+1}} h(\|\boldsymbol{\varepsilon}_{t+1}\|)\right] \cdot \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^{r+m_{t+1}} \|\xi^t - \hat{\xi}^t\|.
\end{aligned}$$

Due to the identity $r + m_{t+1} = m_t - 1$, this entails condition (i) of Assumption 2.6. The asserted form of K follows from $m_1 \geq m_t$ for $t = 1, \ldots, T$.
Furthermore, we apply (31) recursively to obtain the following estimate:

$$\|\boldsymbol{\xi}^T\| \leq \max\{1, \|\boldsymbol{\xi}^t\|\} \prod_{i=t+1}^{T} k(\|\boldsymbol{\varepsilon}_i\|).$$

Raising both sides to the power of $1 + (T - t)$ and taking conditional expectations $\mathbb{E}[\ \cdot \ |\boldsymbol{\xi}^t = \xi^t]$ verifies condition (ii) of Assumption 2.6. □

The conditions of Lemma A.1 are fulfilled, e.g., by a variety of time-series models. We provide the following simple example.

Example A.2. *Let* $\boldsymbol{\xi}$ *be a GARCH process defined by the following difference equations:*

$$\begin{aligned}
&\boldsymbol{\xi}_t = (\boldsymbol{w}_t, \boldsymbol{v}_t, \boldsymbol{\varepsilon}_t) \text{ with} \\
&\boldsymbol{v}_{t+1} \triangleq \sum_{i=0}^{s} (\beta_i v_{t-i} + \gamma_i \boldsymbol{\varepsilon}_{t-i}) \quad \text{and} \quad \boldsymbol{w}_{t+1} \triangleq \sum_{i=0}^{s} \alpha_i \boldsymbol{w}_{t-i} + \boldsymbol{v}_{t+1} \cdot \boldsymbol{\varepsilon}_{t+1}
\end{aligned}$$

for certain parameters $\alpha_i, \beta_i, \gamma_i \in \mathbb{R}$*. Thereby,* $\boldsymbol{v}$ *represents the stochastic volatility process of* $\boldsymbol{w}$ *and* $(\boldsymbol{\varepsilon}_t)_{t\geq 0}$ *is a sequence of i.i.d. random variables, following a standard normal distribution. It is easy to see that* $\boldsymbol{\xi}$ *fulfills the conditions of Lemma A.1 with* $r = 1$ *and* $h(\cdot), k(\cdot)$ *being affine functions.*

The following example shows that nonanticipativity with respect to the initial process is indispensable for an approximating process.

Example A.3. *Consider* $T = 3$ *and the process* $\boldsymbol{\xi}$ *that is given by* $\boldsymbol{\xi}_1 \equiv 0$ *and the two independent random variables* $\boldsymbol{\xi}_2$ *and* $\boldsymbol{\xi}_3$*, both uniformly distributed on* $[0, 1]$*. For* $n \in \mathbb{N}$ *and* $0 < \varepsilon < 1$ *we introduce the grids* $A^{(n)} \triangleq \{\frac{i}{n} : i = 1, \ldots, n\}$ *and the associated (right-continuous) projections* $\pi_{A^{(n)}} : [0, 1] \to A^{(n)}$*, defined by*

$$\pi_{A^{(n)}}(z) \triangleq \max\left\{\frac{i}{n} \in A^{(n)} : \left|z - \frac{i}{n}\right| \leq \left|z - \frac{j}{n}\right| \text{ for all } \frac{j}{n} \in A^{(n)}\right\}.$$

Furthermore, we define processes $\boldsymbol{\xi}^{(n)}, n \in \mathbb{N}$*, given by* $\boldsymbol{\xi}_1^{(n)} \equiv 0$*,* $\boldsymbol{\xi}_3^{(n)} \triangleq \pi_{A^{(n)}}\boldsymbol{\xi}_3$*, and*

$$\boldsymbol{\xi}_2^{(n)} \triangleq \begin{cases} \pi_{A^{(n)}}\boldsymbol{\xi}_2 & \text{if } \boldsymbol{\xi}_3 \leq 1/2, \\ (\pi_{A^{(n)}}\boldsymbol{\xi}_2) + \frac{\varepsilon}{n} & \text{if } \boldsymbol{\xi}_3 > 1/2. \end{cases}$$

We remark that by observing $\boldsymbol{\xi}_2^{(n)}$ *one knows whether* $\boldsymbol{\xi}_3 > \frac{1}{2}$ *or not. Furthermore, the sequence* $\boldsymbol{\xi}^{(n)}$ *can be seen as an approximation of* $\boldsymbol{\xi}$*, since* $\mathbb{E}\left[\|\boldsymbol{\xi} - \boldsymbol{\xi}^{(n)}\|\right] \leq \frac{1+2\varepsilon}{2n}$ *holds.*

We consider the following optimization problem

$$v(\boldsymbol{\xi}) \triangleq \min\left\{\mathbb{E}\left[\boldsymbol{x}_2 \cdot \boldsymbol{\xi}_2 + \boldsymbol{x}_3 \cdot \boldsymbol{\xi}_3\right] : \ \boldsymbol{x}_t \geq 0,\, \boldsymbol{x}_t \in \sigma(\boldsymbol{\xi}^t),\, t = 2, 3,\, \boldsymbol{x}_2 + \boldsymbol{x}_3 = 1 \text{ a.s.}\right\},$$

which is solved by $\boldsymbol{x}_2^* = \mathbb{1}_{\{\boldsymbol{\xi}_2 \leq 1/2\}}$ *and* $\boldsymbol{x}_3^* = 1 - \boldsymbol{x}_2^*$ *with optimal value* $v(\boldsymbol{\xi}) = 12/32$*. When replacing* $\boldsymbol{\xi}$ *by* $\boldsymbol{\xi}^{(n)}$*, we use the decisions*

$$\boldsymbol{x}_2^{(n)} = \mathbb{1}_{\{\boldsymbol{\xi}_2^{(n)} \leq 1/4\}} + \mathbb{1}_{\{\boldsymbol{\xi}_2^{(n)} \in]1/4,\, 3/4[\setminus A^{(n)}\}} \quad \text{and} \quad \boldsymbol{x}_3^{(n)} = 1 - \boldsymbol{x}_2^{(n)}$$

to obtain $\limsup_{n\to\infty} v(\boldsymbol{\xi}^{(n)}) \leq 11/32$*. Obviously, convergence of* $v(\boldsymbol{\xi}^{(n)})$ *to* $v(\boldsymbol{\xi})$ *does not hold since the processes* $\boldsymbol{\xi}^{(n)}$ *do not fulfill the nonanticipativity condition (i) of Definition 3.1. With regard to the results of Heitsch, Römisch, and Strugarek (2006) and Mirkov and Pflug (2007), convergence fails since the conditional distributions* $\mathbb{P}[\boldsymbol{\xi}_3^{(n)} \in \cdot \,|\boldsymbol{\xi}_2^{(n)} = z]$ *do not converge toward* $\mathbb{P}[\boldsymbol{\xi}_3 \in \cdot \,|\boldsymbol{\xi}_2 = z]$ *and the filtration distance between* $\boldsymbol{\xi}^{(n)}$ *and* $\boldsymbol{\xi}$ *does not converge toward* 0*, respectively.*

The following example shows that the perturbed problem $v(\tilde{\boldsymbol{\xi}})$ may be unbounded, even if the initial problem $v(\boldsymbol{\xi})$ admits a bounded optimal solution.

Example A.4. *Consider some* $\varepsilon \in (0, \frac{1}{4})$*,* $T = 2$*, and the* stock prices $\boldsymbol{\xi}_1 \equiv \frac{1}{2} + \varepsilon$ *and* $\boldsymbol{\xi}_2$*, where the latter is uniformly distributed on* $[0, 1]$*. The optimal investment problem*

$v(\boldsymbol{\xi}) = \min_{x \geq 0} x \cdot \boldsymbol{\xi}_1 - \mathbb{E}[x \cdot \boldsymbol{\xi}_2] = \min_{x \geq 0} x \cdot \varepsilon$ *is solved by* $x = 0$. *The process* $\tilde{\boldsymbol{\xi}}$, *defined by* $\tilde{\boldsymbol{\xi}}_1 \triangleq \boldsymbol{\xi}_1$ *and*

$$\tilde{\boldsymbol{\xi}}_2 \triangleq \begin{cases} 1 & \textit{if } \boldsymbol{\xi}_2 \geq \frac{1}{2} - 2\,\varepsilon, \\ 0 & \textit{else} \end{cases},$$

is an approximation of $\boldsymbol{\xi}$ *according to Definition 3.1. However, we see that* $\mathbb{E}[\tilde{\boldsymbol{\xi}}_2] = \frac{1}{2} + 2\,\varepsilon$ *and, consequently,* $v(\tilde{\boldsymbol{\xi}}) = \min_{x \geq 0} -x \cdot \varepsilon = -\infty$.

References

BALLY, V., G. PAGÈS, AND J. PRINTEMS (2005): "A quantization tree method for pricing and hedging multidimensional american options," *Mathematical Finance*, 15(1), 119–168.

BARTY, K. (2004): "Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique," Thèse de Doctorat, École Nationale des Ponts et Chaussées, Paris.

BIRGE, J., AND F. LOUVEAUX (1997): *Introduction to Stochastic Programming*, Springer Series in Operations Research. Springer-Verlag.

DOBRUSHIN, R. (1956): "Central limit theorem for non-stationary Markov chains I.," *Teor. Veroyatnost. i Primenen.*, 1(N1), 72–89.

DYNKIN, E. (1965): *Markov processes.* Springer, Berlin.

EVSTIGNEEV, I. (1976): "Measurable selection and dynamic programming," *Mathematics of Operations Research*, 1(3), 267–272.

HEITSCH, H., AND W. RÖMISCH (2007): "Scenario tree modelling for multistage stochastic programs," *Mathematical Programming,* to appear.

HEITSCH, H., W. RÖMISCH, AND C. STRUGAREK (2006): "Stability of multistage stochastic programs," *SIAM Journal on Optimization*, 17, 511–525.

HOCHREITER, R., AND G. PFLUG (2007): "Financial scenario generation for stochastic multi-stage decision processes as facility location problems," *Annals of Operations Research*, 152(1), 257–272.

HØYLAND, K., AND S. WALLACE (2001): "Generating scenario trees for multistage decision problems," *Management Science*, 47(2), 295–307.

Kuhn, D. (2005): *Generalized bounds for convex multistage stochastic programs*, vol. 548 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin.

Mirkov, R., and G. Pflug (2007): "Tree Approximations of Dynamic Stochastic Programs," *SIAM Journal on Optimization*, 18(3), 1082–1105.

Pennanen, T. (2005): "Epi-convergent Discretizations of Multistage Stochastic Programs," *Mathematics of Operations Research*, 30(1), 245–256.

——— (2007): "Epi-convergent discretizations of multistage stochastic programs via integration quadratures," *Mathematical Programming*, to appear.

Rachev, S. (1991): *Probability Metrics and the Stability of Stochastic Models*. Wiley.

Rockafellar, R., and R. J.-B. Wets (1974): "Continuous Versus Measurable Recourse in N-Stage Stochastic Programming," *Journal of Mathematical Analysis and Applications*, 48(3), 836–859.

——— (1998): *Variational Analysis*. Springer-Verlag, Berlin.

Römisch, W. (2003): *Stability of Stochastic Programming Problems,* chap. 8, pp. 483–554, in Ruszczyński and Shapiro (2003).

Ruszczyński, A., and A. Shapiro (eds.) (2003): *Stochastic Programming*, Handbooks in Operations Research and Management Science. Elsevier, Amsterdam.

Shapiro, A. (2003): "Inference of statistical bounds for multistage stochastic programming problems," *Mathematical Methods of Operations Research*, 58(1), 57–68.