Stability in multistage stochastic programming*

Olga Fiedler** and Werner Römisch
Humboldt-Universität Berlin, Institut für Mathematik, D-10099 Berlin, Germany

Multistage stochastic programs are regarded as mathematical programs in a Banach space X of summable functions. Relying on a result for parametric programs in Banach spaces, the paper presents conditions under which linearly constrained convex multistage problems behave stably when the (input) data process is subjected to (small) perturbations. In particular, we show the persistence of optimal solutions, the local Lipschitz continuity of the optimal value and the upper semicontinuity of optimal sets with respect to the weak topology in X. The linear case with deterministic first-stage decisions is studied in more detail.

Keywords: Multistage stochastic programs, optimization in Banach spaces, stability, approximation.

1. Introduction

Multistage stochastic programs arise in the modelling of finite horizon sequential optimization processes, in which a decision is made at stage t ($1 \leq t \leq T$) based only on information available at time t. In most practical situations, the available information becomes more refined with the passing of time. In our case, the information flow is generated by random variables ξ_t at each stage t ($1 \leq t \leq T$) defined on some probability space (Ω, Σ, μ). The (random) decision x_t (at stage t) should then depend only on the information (or data) $\xi_1, \dots, \xi_t$, i.e., x_t should be measurable with respect to the σ-algebra $\Sigma_t \subseteq \Sigma$, which is generated by the random vector $(\xi_1, \dots, \xi_t)$. If the decision process $x = (x_1, \dots, x_T)$ is adapted to the data process $\xi = (\xi_1, \dots, \xi_T)$ in this way, we call the process x nonanticipative (cf. [27, 33]).

We shall be concerned with the following multistage stochastic program:

$$\text{Minimize } E[f_0(x_1(\omega), \dots, x_T(\omega))] \text{ subject to the constraints} \tag{1.1}$$

$$\left.\begin{array}{l} A_t x_t(\omega) = g_t(x_1(\omega), \dots, x_{t-1}(\omega), \xi_t(\omega)) \\ x_t(\omega) \in C_t \\ x_t \text{ is } \Sigma_t\text{-measurable} \end{array}\right\} \quad \mu\text{-a.s.}, t = 1, \dots, T. \tag{1.2}$$

* This research has been supported by the Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuerung" of the Deutsche Forschungsgemeinschaft.

** Present address: Freie Universität Berlin, Institut für Mathematik, Arnimallee 2-6, D-14195 Berlin, Germany.

Here, f_0 is a real valued measurable function defined on $\times_{t=1}^{T}\mathbb{R}^{n_t}$, C_t is a nonempty subset of $\mathbb{R}^{n_t}$, A_t is an (s_t, n_t)-matrix and g_t is a measurable function from $(\times_{i=1}^{t-1}\mathbb{R}^{n_i}) \times \mathbb{R}^{r_t}$ into $\mathbb{R}^{s_t}$ for every $t = 1, \ldots, T$. E denotes the expectation with respect to (Ω, Σ, μ), and μ-a.s. means that the constraints are required to hold with μ-probability one.

Further hypotheses on the data in (1.1)–(1.2) are presented at the beginning of section 3. They enable, in particular, the formulation of (1.1)–(1.2) as a linearly constrained convex programming problem in an L_p-space. This L_p-space (or vector) formulation of multistage stochastic programs has been studied in several papers (e.g. [4, 8, 14, 18, 27, 28]). For further information on multistage models the reader is referred to [9, 12, 34] and to the recent overview [7]. We also mention applications of multistage stochastic programs in mathematical economics [1], stochastic scheduling [3], finance [6], resource management [10] and power generation [20].

In this paper, we analyze the effect of perturbations of the data process ξ on the optimal value and optimal decision set to (1.1)–(1.2). We show that, under reasonable assumptions, the optimal value and optimal decisions enjoy certain continuity properties if ξ is subjected to (small) perturbations with respect to some metric data space P_d. Stability results of this kind are available for certain two-stage models (see [15, 23, 29–31] and the survey [5]).

Those results assert exclusively stability for the deterministic first-stage decisions and rely on finite-dimensional parametric optimization, whereas our approach necessarily utilizes stability results for programs in Banach spaces. Stability investigations for stochastic programs have a twofold motivation. Their outcome forms a theoretical basis for dealing with approximation schemes and incomplete information on the data.

Our paper is organized as follows. In section 2 we present a stability result (proposition 2.1) for a parametric program in a reflexive Banach space X under mild assumptions (e.g. without differentiability conditions). The result relies on a certain interaction of continuity and closedness properties (for the objective and the constraint set) with respect to strong and weak convergence in X as well as on a boundedness condition for level sets. As a conclusion we study a convex parametric program where the parameter only appears in linear constraints (corollary 2.2), which allows for a direct application to multistage models. Section 3 begins with the formulation of (1.1)–(1.2) as a mathematical program in a suitable L_p-space. Then the main stability result (theorem 3.1) is presented and followed by a discussion of the crucial condition that a certain level set is bounded. More precisely, we show (corollary 3.3) that a linear multistage model is stable if the optimal solution set of its "dynamic" formulation is nonempty and bounded. Finally, we discuss the special case of two-stage problems and the case of discrete approximations to linear multistage programs.

2. A stability result for programs in Banach spaces

In this section we present a perturbation result for infinite optimization problems which serves as a prerequisite for our stability analysis of multistage stochastic programs.

Let X be a Banach space with norm $\|\cdot\|$ and (P, d) be a metric space. We study the parametric optimization problem

$$\min\{f(x): x \in M(p)\} \quad (p \in P), \tag{2.1}$$

where f is a mapping from X into $\mathbb{R}$ and M a set-valued mapping from P into X. We define the optimal value $\varphi(p) := \inf\{f(x): x \in M(p)\}$ and the set of optimal solutions $\psi(p) := \{x \in M(p): f(x) = \varphi(p)\}$ of (2.1). For some fixed parameter $p_0 \in P$ we refer to (2.1) with $p = p_0$ as the "original program" and to the case of $p \neq p_0$ as the "perturbed program".

Our stability result relies on a certain interaction of properties (for the data of (2.1)) with respect to strong (norm) and weak convergence in X. Weak convergence in X is denoted by "$\rightharpoonup$". For the formulation of the result we recall the following notions. The function f is called weakly lower semicontinuous on X if, for each $x \in X$, $x_n \in X$ $(n \in \mathbb{N})$ with $x_n \rightharpoonup x$, we have $\liminf_{n\to\infty} f(x_n) \geq f(x)$. M is said to be weakly closed at p_0 if for each pair of sequences (p_n) in P and (x_n) in X with the properties $p_n \to p_0$, $x_n \in M(p_n)$, $x_n \rightharpoonup x_0$, it holds $x_0 \in M(p_0)$. M is called Hausdorff-continuous at p_0 if $d_H(M(p), M(p_0)) \to_{p\to p_0} 0$, where $d_H(A, B) := \max\{\Delta(A, B), \Delta(B, A)\}$ is the (extended) Hausdorff distance between two subsets A, B of X and $\Delta(A, B) := \sup_{x\in A} \inf_{y\in B} \|x - y\|$ is the excess of A on B. (Note that $d_H(A, B)$ may be equal to $+\infty$ when A or B is unbounded or empty.)

PROPOSITION 2.1

Consider the program (2.1), fix some $p_0 \in P$ and assume that

(a) X is reflexive;

(b) f is uniformly continuous and weakly lower semicontinuous;

(c) $M(p_0)$ is nonempty, M is weakly closed at p_0 and its values are weakly closed subsets of X, and M is Hausdorff-continuous at p_0;

(d) the level set $l_c(p_0) := \{x \in M(p_0): f(x) \leq c\}$ is bounded for each $c \in \mathbb{R}$.

The following holds:

(i) φ is continuous at p_0,

(ii) $\psi(p) \neq \emptyset$ for all p belonging to some neighbourhood U of p_0 in P ("persistence"),

(iii) ψ is weakly upper semicontinuous at p_0, i.e., for all sequences (p_n) in P and (x_n) in X having the properties $p_n \to p_0$, $x_n \in \psi(p_n)$, there exists a subsequence of (x_n) which converges weakly in X to an element of $\psi(p_0)$.

Proof

Let $\epsilon > 0$ be fixed. Since f is uniformly continuous, there exists a $\delta > 0$ such that x, $\tilde{x} \in X$ and $\|x - \tilde{x}\| < \delta$ implies $|f(x) - f(\tilde{x})| < \epsilon$. Let $x_0 \in M(p_0)$. Then there exists a neighbourhood U of p_0 such that $d_H(M(p), M(p_0)) < \delta$ for all p belonging to U. In particular, there exists a mapping $x: U \to X$ with the properties $x(p) \in M(p)$ and

$$\|x_0 - x(p)\| < \delta, \quad \text{for all } p \in U.$$

Let $p \in U$ and $y \in M(p)$ with $f(y) \leq f(x(p))$. We obtain

$$f(y) \leq f(x(p)) \leq f(x_0) + |f(x(p)) - f(x_0)| < f(x_0) + \epsilon.$$

Furthermore, there exists an $\bar{x} \in M(p_0)$ such that $\|y - \bar{x}\| < \delta$. This implies $f(\bar{x}) \leq f(y) + \epsilon \leq f(x_0) + 2\epsilon$, i.e., y belongs to the bounded set $B := \{x \in X: \inf_{v \in l_c(p_0)} \|x - v\| \leq \delta\}$ with $c := f(x_0) + 2\epsilon$.

Hence, $\{y \in M(p): f(y) \leq f(x(p))\} \subseteq B$ for all $p \in U$. Since $M(p)$ is weakly closed and f is weakly lower semicontinuous, the level set $l_{f(x(p))}(p)$ is weakly compact. Assertion (ii) now follows from Weierstrass' theorem.

Although the proof of (i) and (iii) parallels those of classical stability results (e.g. theorem 4.2.2 in [2] and theorem 3.1 in [16]), we will outline the proofs for the reader's convenience.

To prove (i), let $\epsilon > 0$ be chosen arbitrarily and let $x_0 \in \psi(p_0)$. Again, we choose $\delta > 0$ from the uniform continuity of f as at the beginning of the proof. Then there exists a neighbourhood V of p_0 such that for each $p \in V$ there is an element $\bar{x}(p) \in M(p)$ with $\|x_0 - \bar{x}(p)\| < \delta$. Hence, $\varphi(p) - \varphi(p_0) \leq f(\bar{x}(p)) - f(x_0) < \epsilon$ for all $p \in V$. Analogously one shows $\varphi(p_0) - \varphi(p) < \epsilon$ for all $p \in V$, and the proof of (i) is complete.

Finally, let (p_n) be a sequence in P which converges to p_0 and let $x_n \in \psi(p_n)$ for all $n \in \mathbb{N}$. Since x_n belongs to B for large n, we may assume without loss of generality that (x_n) converges weakly to some limit $x_0 \in X$. (c) implies that $x_0 \in M(p_0)$. To show the optimality of x_0, we choose $x^* \in \psi(p_0)$ and a sequence (x_n^*) with $x_n^* \in M(p_n)$ for each $n \in \mathbb{N}$ such that $x_n^* \to x^*$. Then we obtain $\varphi(p_0) \leq f(x_0) \leq \liminf_{n\to\infty} f(x_n) = \liminf_{n\to\infty} f(x_n^*) = f(x^*) = \varphi(p_0)$. This shows that $x_0 \in \psi(p_0)$ and completes the proof. □

For our application to stochastic programming we are interested in the particular example of (2.1), where the constraint-set-mapping M is of the following form

$$M(p) := \{x \in C: A(p)x = b(p)\} \quad (p \in P), \tag{2.2}$$

where $C \subseteq X$ is closed convex, A is a continuous mapping from P to the normed space $L(X, Y)$ of linear bounded operators from X to another linear normed space Y and b is a continuous mapping from P to Y.

COROLLARY 2.2

Let X be reflexive, f be convex and uniformly continuous on C. Consider (2.2) with general assumptions as above and assume that $M(p_0)$ is nonempty and M is Hausdorff continuous at p_0. Let the level set $l_c(p_0)$ be bounded for some $c > \varphi(p_0)$. Then the assertions (i)–(iii) of proposition 2.1 hold, too.

Proof

Since f is convex and continuous on C, it is also weakly lower semicontinuous on C. An inspection of the proof of proposition 2.1 shows in addition that condition (b) is only needed to hold on C. By definition the values of M are closed and convex, and, hence, weakly closed subsets of X. In order to show that M is weakly closed at p_0, let (p_n) be a sequence in P converging to p_0 and let $x_n \in M(p_n)$, for each $n \in \mathbb{N}$, such that (x_n) converges weakly to some $x_0 \in X$. Then $x_0 \in C$ and $b(p_n) = A(p_n)x_n \to b(p_0)$, since b is continuous. The general assumptions for A imply that the sequence $(A(p_0)x_n)$ converges weakly to $A(p_0)x_0$ (in Y) and that $\lim_{n\to\infty} \| A(p_n)x_n - A(p_0)x_n \| = 0$.

Thus, we conclude that $(A(p_n)x_n)$ converges weakly to $A(p_0)x_0$, too, and obtain $A(p_0)x_0 = b(p_0)$. Hence, condition (c) is satisfied and it remains to appeal to corollary 4D in [24], which yields that $l_c(p_0)$ is bounded for every $c \in \mathbb{R}$ if it is bounded for some $c > \varphi(p_0)$. □

Remark 2.3

If, in proposition 2.1 and corollary 2.2, the uniform continuity property for f is replaced by the Lipschitz continuity and if M is (even) Hausdorff Lipschitz continuous at p_0, then the optimal value satisfies even the Lipschitz condition $|\varphi(p) - \varphi(p_0)| \leq L_\varphi d(p, p_0)$ for p belonging to some neighbourhood of p_0 and for L_φ being the product of the Lipschitz constants for f and M.

Corollary 2.2 becomes incorrect if the assumption that the level set $l_c(p_0)$ is bounded for some $c > \varphi(p_0)$ is replaced by the (weaker) condition that $\psi(p_0)$ is nonempty and bounded. This is illustrated by the next example, which is essentially due to Bernd Kummer.

EXAMPLE 2.4

Let $X := \{(x_n) : \sum_{n=1}^{\infty} x_n^2 < \infty\}$ be the classical Hilbert space l^2 of real sequences with the norm $\|(x_n)\| := \left(\sum_{n=1}^{\infty} x_n^2\right)^{1/2}$. Consider the mapping

$f((x_n)) := \sum_{n=1}^{\infty}(1/n)|x_n|$, which is convex and Lipschitz continuous on X. Let $P := \mathbb{R}_+$, $p_0 := 0$ and define the set-valued mapping M from P to X by $M(p) := \{(x_n) \in X: \sum_{n=1}^{\infty}(1/n)x_n = p\}$ $(p \in P)$. Then $\varphi(p) := \inf\{f((x_n)): (x_n) \in M(p)\} = p$, for every $p \in P$, $\psi(0) = \{0\}$ and $d_H(M(p), M(0)) = p$.

Furthermore, the sequences $(x_n^{(k)})$,

$$x_n^{(k)} := \begin{cases} 0, & n \neq k \\ pn, & n = k \end{cases} \quad (k \in \mathbb{N})$$

belong to $\psi(p)$. Hence, the optimal sets $\psi(p)$ are not bounded for every $p > 0$ and, thus, do not satisfy property (iii) of proposition 2.1.

The reflexivity of the Banach space X is indispensable even for the solvability of the original and perturbed programs (2.1). Consequently, we consider the stability of multistage stochastic programs in L_p-spaces (with $p \in (1, +\infty)$) of the underlying random quantities. Studying multistage stochastic programs as programs in L_p-spaces was first proposed in [8] and [18]. This approach has also been used to derive necessary optimality conditions [4, 11, 14]. At first glance, a drawback of this functional approach might be that abstract constraint qualifications in Banach spaces (e.g. [16, 22, 36]) can hardly be fulfilled in L_p-spaces. To verify the Hausdorff-continuity of M, however, we do not rely on results in functional spaces (e.g. those in [22]), but work naturally in terms of each realization in finite dimensions.

3. Stability of multistage stochastic programs

With the notations of section 1, we define $X_t := L_{p_t}(\Omega, \Sigma_t, \mu; \mathbb{R}^{n_t})$ for $p_t \in (1, +\infty)$ and $t = 1, \ldots, T$. As the space of decision processes we take $X := \times_{t=1}^{T} X_t$ equipped with the norm $\|x\| := \max_{t=1,\ldots,T}\|x_t\|$, where $\|\cdot\|_t$ denotes the norm in X_t. For the data we assume that ξ_t belongs to some normed space P_t of $\mathbb{R}^{r_t}$-valued random variables with the norm $\|\cdot\|_{*,t}$ and define $P_d := \times_{t=1}^{T} P_t$ to be the data space equipped with the norm $\|\xi\|_* := \max_{t=1,\ldots,T}\|\xi_t\|_{*,t}$. Later, we will specify conditions for the choice of P_t and for relations between $\|\cdot\|_t$ and $\|\cdot\|_{*,t}$ (cf. (A3)). The constraint set $M(\xi)$ for the decisions is defined as follows

$$M(\xi) := \{x \in X: x_t(\omega) \in C_t, A_t x_t(\omega) = g_t(x^{t-1}(\omega), \xi_t(\omega)), t = 1, \ldots, T, \mu\text{-a.s.}\}, \tag{3.1}$$

where we use the notation x^{t-1} for $(x_1, \ldots, x_{t-1})$. M may be considered as a set-valued mapping from the data space P_d to the decision space X. By putting the objective

$$f(x) := E[f_0(x_1(\omega), \dots, x_T(\omega))] \quad (x \in X), \tag{3.2}$$

problem (1.1)–(1.2) leads to the following parametric program in the decision space X:

$$\min\{f(x) : x \in M(\xi)\}. \tag{3.3}$$

Throughout this section, we make the following assumptions for the data in (3.1)–(3.2) (and (1.1)–(1.2), respectively).

(A1) For each $t \in \{1, \dots, T\}$, the set $C_t \subseteq \mathbb{R}^{n_t}$ is nonempty, convex and polyhedral.

(A2) The function $f_0 \colon \times_{t=1}^T \mathbb{R}^{n_t} \to \mathbb{R}$ is convex and Lipschitz continuous on $\times_{t=1}^T C_t$.

(A3) For each $t \in \{1, \dots, T\}$, the function $g_t \colon (\times_{i=1}^{t-1} \mathbb{R}^{n_i}) \times \mathbb{R}^{r_t} \to \mathbb{R}^{s_t}$ is affine linear in $x^{t-1} \in \times_{i=1}^{t-1} \mathbb{R}^{n_i}$ and $\xi_t \in \mathbb{R}^{r_t}$, respectively, and satisfies the estimates

$$\| g_t(x^{t-1}(\cdot), \xi_t(\cdot)) \|_{(t)} \le K_t \Big(1 + \max_{i=1,\dots,t-1} \| x_i \|_i \Big) (1 + \| \xi_t \|_{*,t}),$$

$$\| g_t(x^{t-1}(\cdot), \xi_t(\cdot)) - g_t(y^{t-1}(\cdot), \xi_t(\cdot)) \|_{(t)} \le K_t (1 + \| \xi_t \|_{*,t}) \max_{i=1,\dots,t-1} \| x_i - y_i \|_i,$$

$$\| g_t(x^{t-1}(\cdot), \xi_t(\cdot)) - g_t(x^{t-1}(\cdot), \eta_t(\cdot)) \|_{(t)} \le K_t \Big(1 + \max_{i=1,\dots,t-1} \| x_i \|_i \Big) \| \xi_t - \eta_t \|_{*,t},$$

with some constant $K_t > 0$, for all $x, y \in X$, $\xi, \eta \in P_d$. Here $\| \cdot \|_{(t)}$ denotes the norm in $Y_t := L_1(\Omega, \Sigma, \mu; \mathbb{R}^{s_t})$.

(A4) Let P be the subset of those $\xi \in P_d$ satisfying the property $g_t(v_1, \dots, v_{t-1}, \xi_t(\omega)) \in A_t(C_t)$, μ-a.s., for all $v_i \in C_i$, $i = 1, \dots, t-1$, and $t = 1, \dots, T$. Let P be nonempty and consider P to be equipped with the metric induced by $\| \cdot \|_*$.

We mention here that (A3) describes growth and continuity conditions for the constraint functions g_t with respect to appropriate norms and that (A4) characterizes all admissible data processes. A discussion of both assumptions is given in remark 3.2.

We denote the optimal value and the set of optimal solutions to (3.3) by $\varphi(\xi)$ and $\psi(\xi)$, respectively. The following theorem presents our main stability result.

THEOREM 3.1

Suppose (A1)–(A4), $\xi^0 \in P$ and let the level set $l_c(\xi^0) := \{x \in M(\xi^0): f(x) \le c\}$ be bounded in X for some constant $c > \varphi(\xi^0)$. Then there exists a neighbourhood U of ξ^0 in P and a constant $L_\varphi > 0$ such that

(i) $|\varphi(\xi) - \varphi(\xi^0)| \le L_\varphi \|\xi - \xi^0\|_*$ for all $\xi \in U$,

(ii) $\psi(\xi) \ne \emptyset$ for all $\xi \in U$,

(iii) ψ is weakly upper semicontinuous at ξ^0 (in the sense of 2.1(iii)).

Proof

To prove the result we apply corollary 2.2 (together with remark 2.3). The space X is reflexive and (A1) implies that the subset $C := \{x \in X: x_t(\omega) \in C_t$ μ-a.s., $t = 1, \dots, T\}$ of X is closed convex. Because of (A2), the objective function f is well-defined on C and inherits the convexity and Lipschitz property (on C) from f_0. Since each function g_t $(t = 1, \dots, T)$ is affine linear, the constraint set $M(\xi)$ is of the following form $M(\xi) = \{x \in C: A_t x_t(\omega) + \sum_{i=1}^{t-1} A_{ti}(\xi_t(\omega)) x_i(\omega) = b_t(\xi_t(\omega))$ μ-a.s, $t = 1, \dots, T\}$, where $A_{ti}(\cdot)$ $(i = 1, \dots, t-1)$ and $b_t(\cdot)$ depend affine linearly on ξ_t $(t = 1, \dots, T)$. Hence, $M(\xi)$ has the form (2.2) and, by putting $Y := \times_{t=1}^{T} Y_t$ the general assumptions of (2.2) are implied by (A3).

It remains to show that $M(\xi^0)$ is nonempty and that M (from P to X) is Hausdorff–Lipschitzian at ξ^0. To this end, we introduce the following notation for every $t \in \{1, \dots, T\}$:

$$B_t(y_t) := \{v_t \in C_t: A_t v_t = y_t\} \quad (y_t \in \mathbb{R}^{s_t}).$$

By Hoffman's Theorem [22, p. 760] there exists a constant $L_t > 0$ such that the polyhedral set-valued mapping B_t from $\mathbb{R}^{s_t}$ into $\mathbb{R}^{n_t}$ has the Hausdorff–Lipschitz property

$$d_{H,t}(B_t(y_t), B_t(\tilde{y}_t)) \le L_t |y_t - \tilde{y}_t|_{s_t} \tag{3.4}$$

whenever $y_t, \tilde{y}_t \in \mathbb{R}^{s_t}$ and $B_t(y_t) \ne \emptyset$, $B_t(\tilde{y}_t) \ne \emptyset$. Here $d_{H,t}$ denotes the Hausdorff distance on nonempty subsets of $\mathbb{R}^{n_t}$ and $|\cdot|_{s_t}$ the Euclidean norm on $\mathbb{R}^{s_t}$.

Next, we show that $M(\xi)$ is nonempty for every $\xi \in P$. Let $\xi \in P$. By induction we show that there exists an $x_t \in X_t$ for each $t = 1, \dots, T$ such that $(x_1, \dots, x_T) \in M(\xi)$. First let $t = 1$. (A4) implies that $\hat{B}_1(\cdot) := B_1(g_1(\xi_1(\cdot)))$ is a set-valued mapping from Ω to $\mathbb{R}^{n_1}$ having closed and μ-a.s. nonempty values. By standard arguments (cf. e.g. theorem 2J in [26]), $\hat{B}_1$ is Σ_1-measurable and, hence, there exists a Σ_1-measurable mapping $x_1: \Omega \to \mathbb{R}^{n_1}$ such that $x_1(\omega) \in B_1(\omega)$ μ-a.s. Then (3.4) implies for some $v_1 \in C_1$,

$$|x_1(\omega) - v_1|_{n_1} \le L_1 |g_1(\xi_1(\omega)) - A_1 v_1|_{s_1} \quad \mu\text{-a.s., and}$$

$$\begin{aligned} \|x_1\|_1 &\le |v_1|_{n_1} + L_1 |A_1 v_1|_{s_1} + L_1 \|g_1(\xi_1(\cdot))\|_{(1)} \\ &\le \hat{K}_1(1 + \|\xi_1\|_{*,1}). \end{aligned}$$

Here $\hat{K}_1 > 0$ is some constant that exists according to (A3). Now, suppose that we have already determined $x_i \in X_i$, $i = 1, \ldots, t-1$, such that $x_i(\omega) \in B_i(g_i(x^{i-1}(\omega), \xi_i(\omega)))$, μ-a.s., $i = 1, \ldots, t-1$. The set-valued mapping $\hat{B}_t(\cdot) := B_t(g_t(x^{t-1}(\cdot), \xi_t(\cdot)))$ from Ω to $\mathbb{R}^{n_t}$ has again closed and μ-a.s. nonempty values. By analogous arguments there exists a Σ_t-measurable mapping $x_t: \Omega \to \mathbb{R}^{n_t}$ such that $x_t(\omega) \in \hat{B}_t(\omega)$ μ-a.s. and the following estimate holds according to (3.4) and (A3),

$$\|x_t\|_t \leq \hat{K}_t\left(1 + \max_{i=1,\ldots,t-1} \|x_i\|_i\right)(1 + \|\xi_t\|_{*,t}) \tag{3.5}$$

with some positive constant $\hat{K}_t$. Hence, we have $(x_1, \ldots, x_T) \in M(\xi)$.

We finally show that M is Hausdorff–Lipschitz continuous at ξ^0. In fact, we even show that M is Hausdorff–Lipschitzian at each ξ belonging to P. Because of symmetric arguments we only have to consider the one-sided Hausdorff distance

$$\Delta(M(\eta), M(\xi)) := \sup_{x \in M(\eta)} \inf_{y \in M(\xi)} \|x - y\| \quad (\xi, \eta \in P).$$

Let ξ, $\eta \in P$ and $x \in M(\eta)$. Since $M(\xi)$ is nonempty, closed and convex in the reflexive Banach space X, there exists a $z \in M(\xi)$ such that $\|x - z\| = \inf_{y \in M(\xi)} \|x - y\|$. For later use we introduce the following subsets of X_t depending on $z^{t-1} := (z_1, \ldots, z_{t-1})$ $(t = 1, \ldots, T)$:

$$M_t(z^{t-1}, \xi_t) := \{y_t \in X_t: y_t(\omega) \in B_t(g_t(z^{t-1}(\omega), \xi_t(\omega))), \mu\text{-a.s.}\}.$$

By definition we have $z_t \in M_t(z^{t-1}, \xi_t)$ and, moreover,

$$\|x_t - z_t\|_t = \inf_{y_t \in M_t(z^{t-1}, \xi_t)} \|x_t - y_t\|_t.$$

By appealing to theorem 2.2 in [13] and to (3.4) we obtain

$$\begin{aligned}
\|x_t - z_t\|_t^{p_t} &= \inf_{y_t \in M_t(z^{t-1}, \xi_t)} \int_\Omega |x_t(\omega) - y_t(\omega)|_{n_t}^{p_t} \mu(d\omega) \\
&= \int_\Omega \inf_{v_t \in B_t(g_t(z^{t-1}(\omega), \xi_t(\omega)))} |x_t(\omega) - v_t|_{n_t}^{p_t} \mu(d\omega) \\
&\leq L_t^{p_t} \int_\Omega |g_t(x^{t-1}(\omega), \eta_t(\omega)) - g_t(z^{t-1}(\omega), \xi_t(\omega))|_{s_t}^{p_t} \mu(d\omega) \\
&= L_t^{p_t} \|g_t(x^{t-1}(\cdot), \eta_t(\cdot)) - g_t(z^{t-1}(\cdot), \xi_t(\cdot))\|_{(t)}^{p_t}
\end{aligned}$$

and, hence, by (A3)

$$\begin{aligned}\|x_t - z_t\|_t &\le L_t[\|g_t(x^{t-1}(\cdot),\eta_t(\cdot)) - g_t(z^{t-1}(\cdot),\eta_t(\cdot))\|_{(t)} \\ &\quad + \|g_t(z^{t-1}(\cdot),\eta_t(\cdot)) - g_t(z^{t-1}(\cdot),\xi_t(\cdot))\|_{(t)}] \\ &\le L_t K_t\left[(1+\|\eta_t\|_{*,t}) \max_{i=1,\dots,t-1} \|x_i - z_i\|_i \right. \\ &\quad \left. + \left(1 + \max_{i=1,\dots,t-1} \|z_i\|_i\right)\|\eta_t - \xi_t\|_{*,t}\right].\end{aligned}$$

Since $z \in M(\xi)$, we obtain analogously to that part of the proof leading to (3.5) that $\max_{t=1,\dots,T} \|z_t\|_t$ is bounded by an expression consisting of positive constants and $\|\xi\|_*$. If we now restrict η to vary only in a bounded neighbourhood U of ξ, we can continue our estimate to

$$\|x_t - z_t\|_t \le \tilde{K}_t\left[\max_{i=1,\dots,t-1} \|x_i - z_i\|_i + \|\eta_t - \xi_t\|_{*,t}\right]$$

with some positive constant $\tilde{K}_t$. A successive application of the last estimate now leads to

$$\|x - z\| = \max_{t=1,\dots,T} \|x_t - z_t\|_t \le \tilde{L} \max_{t=1,\dots,T} \|\eta_t - \xi_t\|_{*,t} = \tilde{L}\|\eta - \xi\|$$

with some constant $\tilde{L}$ which is independent of x. By taking the supremum over $x \in M(\eta)$ on the left-hand side of the last inequality, we obtain the desired Lipschitz estimate

$$\Delta(M(\eta), M(\xi)) \le \tilde{L}\|\eta - \xi\|_* \quad \text{for all } \eta \in U,$$

and the proof is complete. □

Remark 3.2

Assumption (A3) provides some flexibility for the choice of the spaces $X_t = L_{p_t}(\Omega, \Sigma_t, \mu; \mathbb{R}^{n_t})$, P_t and of the corresponding norm $\|\cdot\|_{*,t}$ at each stage t. To give an idea how this flexibility can be exploited, consider the function $g_t(x^{t-1}, \xi_t) = -\sum_{i=1}^{t-1} A_{ti}(\xi_t)x_i + b_t(\xi_t)$ in its general form, where the matrices $A_{ti}(\cdot)$, $i = 1, \dots, t-1$, and $b_t(\cdot)$ depend affine linearly on ξ_t. Then estimates of the form required in (A3) can always be attained if the spaces and the corresponding norms allow estimates of the following type:

$$\int_{\Omega} |A_{ti}(\xi_t(\omega))\, x_i(\omega)|_{s_t}\, \mu(d\omega) \le \hat{K}_{ti}(1 + \|\xi_t\|_{*,t})\, \|x_i\|_i, \quad i = 1, \ldots, t-1,$$

$$\int_{\Omega} |b_t(\xi_t(\omega))|_{s_t}\, \mu(d\omega) \le \hat{K}_{tt}(1 + \|\xi_t\|_{*,t}).$$

For instance, let $P_t := L_{q_t}(\Omega, \Sigma_t, \mu; \mathbb{R}^{r_t})$ with $q_t \in (1, +\infty]$. If all matrices A_{ti} are non-random, then $p_t := q_t$ is the natural choice. If, vice versa, all matrices are random, the following conditions have to be fulfilled: $(1/p_i) + (1/q_t) \le 1$, for $i = 1, \ldots, t$.

Assumption (A4) means that all data processes belonging to P fulfil the property that at each stage $t \in \{1, \ldots, T\}$, we have "relatively complete recourse". Of course, (A4) is satisfied with $P := P_d$ (i.e. without restrictions on the perturbations), if we have $A_t(C_t) = \mathbb{R}^{s_t}$ ("complete recourse") for all stages t.

Now, we consider the important case of linear multistage models where the first-stage decisions are deterministic, i.e., $\Sigma_1 := \{\emptyset, \Omega\}$, $X_1 = \mathbb{R}^{n_1}$, $P_1 = \mathbb{R}^{r_1}$ and $f_0(x) := \sum_{t=1}^{T} q_t x_t$, with $q_t \in \mathbb{R}^{n_t}$, $t = 1, \ldots, T$. We put $A_1 := 0$ and $g_1(\xi_1) := 0$. Then (3.3) reads as follows:

$$\text{Minimize} \quad q_1 x_1 + E\left[\sum_{t=2}^{T} q_t x_t(\omega)\right] \tag{3.6}$$

$$\text{subject to} \quad \left.\begin{array}{l} A_t x_t(\omega) = g_t(x_1, x_2(\omega), \ldots, x_{t-1}(\omega), \xi_t(\omega)) \\ x_1 \in C_1, x_t(\omega) \in C_t \\ x_t \in L_{p_t}(\Omega, \Sigma_t, \mu; \mathbb{R}^{n_t}) \end{array}\right\} \mu\text{-a.s.}, t = 2, \ldots, T. \tag{3.7}$$

Together with problem (3.6)–(3.7), we consider the associated stochastic programming model with recourse:

$$\text{Minimize } q_1 x_1 + Q(x_1, \xi) \quad \text{subject to } x_1 \in C_1, \tag{3.8}$$

where

$$Q(x_1, \xi) := E\left[\inf\left\{\sum_{t=2}^{T} q_t x_t(\omega) \colon A_t x_t(\omega) = g_t(x_1, x_2(\omega), \ldots, x_{t-1}(\omega), \xi_t(\omega)),\right.\right.$$
$$\left.\left. x_t(\omega) \in C_t, \mu\text{-a.s.}, x_t \in L_{p_t}(\Omega, \Sigma_t, \mu; \mathbb{R}^{n_t}), t = 1, \ldots, T\right\}\right].$$

In our next result we conclude stability for the model (3.6)–(3.7) from theorem 3.1 by relating the boundedness condition for some level set (in theorem 3.1) to the boundedness of the optimal solution set of (3.8).

COROLLARY 3.3

Let C_1 be a nonempty convex polyhedron and suppose for $t = 2, \dots, T$, $C_t := \{x_t \in \mathbb{R}^{n_t}: x_t \geq 0\}, A_t(C_t) = \mathbb{R}^{s_t}$ ("complete recourse") and $\{u_t \in \mathbb{R}^{s_t}: A_t' u_t \leq q_t\} \neq \emptyset$ ("dual feasibility"). Let $\xi^0 \in P_d$ and assume that the optimal set $\psi_1(\xi^0)$ of (3.8) is nonempty and bounded. Moreover, suppose (A3). Then the assertions (i)–(iii) of theorem 3.1 hold with $P := P_d$ for (3.6)–(3.7), too.

Proof

Since the conditions (A1)–(A4) are fulfilled with $P = P_d$, it remains to verify that the level set $l_c(\xi^0) = \{x \in M(\xi^0): q_1 x_1 + E[\sum_{t=2}^T q_t x_t(\omega)] \leq c\}$ is bounded for some $c > \varphi(\xi^0)$. Here, $M(\xi^0)$ is given by the constraints (3.7) for $\xi := \xi^0$. By definition of Q, the level set $l_c(\xi^0)$ is contained in $\{x \in M(\xi^0): q_1 x_1 + Q(x_1, \xi^0) \leq c\}$. Our assumptions imply by standard arguments (see e.g. [32]) that $Q(\cdot, \xi^0)$ is convex on $\mathbb{R}^{n_1}$. Since $\psi_1(\xi^0)$ is nonempty and bounded, it follows from corollary 8.7.1 in [25] that the level set $\{x_1 \in C_1: q_1 x_1 + Q(x_1, \xi^0) \leq c\}$ is bounded, too. Furthermore, by repeating the argument in the proof of theorem 3.1 leading to (3.5), we obtain for each $(x_1, x_2(\cdot), \dots, x_T(\cdot)) \in M(\xi^0)$ an estimate for $\|x_t\|_t$ $(t = 2, \dots, T)$ in terms of $|x_1|_{n_1}$ and $\|\xi^0\|_*$. This estimate, together with the boundedness of the level set for x_1 above, implies that $\{x \in M(\xi^0): q_1 x_1 + Q(x_1, \xi^0) \leq c\}$ and, hence, $l_c(\xi^0)$ are bounded in X. □

Remark 3.4

Theorem 3.1 and corollary 3.3 apply immediately to linear stochastic two-stage problems, i.e., to (3.6)–(3.7) with $T := 2$:

$$\min\{q_1 x_1 + E[q_2 x_2(\omega)]: x_1 \in C_1, A_2 x_2(\omega) = b(\xi(\omega)) - A_{21}(\xi(\omega)) x_1,$$
$$x_2(\omega) \in C_2, \mu\text{-a.s}, x_2 \in L_p(\Omega, \Sigma, \mu; \mathbb{R}^{n_2})\}. \tag{3.9}$$

As distinct from the stability results for (3.9), which can be derived from the general theory in [15, 23, 29], our results are formulated in terms of perturbations of ξ^0 in some space $L_q = L_q(\Omega, \Sigma, \mu; \mathbb{R}^r)$ of random variables rather than in terms of spaces of probability distributions (endowed with a suitable topology or metric). Our results yield an upper semicontinuity property for the first-stage optimal set ψ_1 in $\mathbb{R}^{n_1}$ and, simultaneously, weak upper semicontinuity for the second-stage optimal solutions in L_p. The additional price that has to be paid is the assumption $\xi^0 \in L_q$ for $q > 1$ (cf. remark 3.2).

The following interpretation of our results is closer to the theory developed in [29, 30]. Let us consider the following distance on the set of all probability measures (defined on $\mathbb{R}^r$):

$$W_q(P, Q) := \inf\{\|\xi - \eta\|_{L_q} : D(\xi) = P, D(\eta) = Q\},$$

where the infimum is taken over all random variables $\xi, \eta \in L_q$ having the probability distribution P and Q, respectively. W_q is the so-called L_q-Wasserstein distance and is a minimal metric [21, 35]. Since the optimal value $\varphi(\xi)$ and the solution set $\psi_1(\xi)$ of (3.9) only depend on the probability distribution $D(\xi)$ of the random variable ξ, our results imply the local Lipschitz continuity of φ and the upper semicontinuity of ψ_1 with respect to perturbations of $D(\xi^0)$ in terms of the metric W_q.

Remark 3.5

The results in the present paper apply, in particular, to perturbations of (1.1)–(1.2) arising from discrete approximations of the original data process ξ^0. For this case, our stability results are close to those obtained in [17] (for linear two-stage problems with random right-hand sides) and in [19] (for linear multistage problems). Without going into detail, we just mention that our results also apply to the case of random technology matrices and work without restrictive assumptions on the distributions of ξ^0 and its perturbations. A barycentric approximation scheme for multistage stochastic models including error bounds was developed in [12]. Our results are also relevant for studying the convergence of this scheme if the cells become arbitrarily small.

Acknowledgements

This paper has benefited from discussions with M.A.H. Dempster (University of Essex, Colchester), B. Kummer and R. Schultz (both Humboldt-Universität Berlin). The comments and suggestions of two anonymous referees are gratefully acknowledged.

References

[1] V.I. Arkin and I.V. Evstigneev, *Stochastic Models of Control and Economic Dynamics* (Academic Press, London, 1987).

[2] B. Bank, J. Guddat, D. Klatte, B. Kummer and K. Tammer, *Non-Linear Parametric Optimization* (Akademie-Verlag, Berlin, 1982).

[3] J.R. Birge and M.A.H. Dempster, Optimality conditions for match-up strategies in stochastic scheduling, The University of Michigan, Department of Industrial and Operations Engineering, Technical Report 92-58 (1992).

[4] M.A.H. Dempster, On stochastic programming II: Dynamic problems under risk, Stochastics 25 (1988) 15–42.

[5] J. Dupačová, Stability and sensitivity analysis for stochastic programming, Ann. Oper. Res. 27 (1990) 115–142.

[6] J. Dupačová, Postoptimality for multistage stochastic linear programs, Ann. Oper. Res. 56 (1995), this volume.

[7] J. Dupačová, Multistage stochastic programs: The state-of-the-art and selected bibliography, Kybernetika 31 (1995), to appear.

[8] M.J. Eisner and P. Olsen, Duality for stochastic programming interpreted as LP in L_p-space, SIAM J. Appl. Math. 28 (1975) 779–792.

[9] I.V. Evstigneev, Measurable selection and dynamic programming, Math. Oper. Res. 1 (1976) 267–272.

[10] S.D. Flåm, Resource management under uncertainty, The Chr. Michelsen Institute, Report CMI No. 842555-5 (1984).

[11] S.D. Flåm, Lagrange multipliers in stochastic programming, SIAM J. Contr. Optim. 30 (1992) 1–10.

[12] K. Frauendorfer, Multistage stochastic programming: Error analysis for the convex case, ZOR-Math. Meth. Oper. Res. 39 (1994) 93–122.

[13] F. Hiai and H. Umegaki, Integrals, conditional expectations, and martingales of multivalued functions, J. Multivariate Anal. 7 (1977) 149–182.

[14] J.-B. Hiriart-Urruty, Extension of Lipschitz integrands and minimization of nonconvex integral functionals: Applications to the optimal recourse problem in discrete time, Prob. Math. Statist. 3 (1982) 19–36.

[15] P. Kall, On approximations and stability in stochastic programming, in: *Parametric Optimization and Related Topics*, eds. J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička (Akademie-Verlag, Berlin, 1987) pp. 387–407.

[16] A. Kirsch, Continuous perturbations of infinite optimization problems, J. Optim. Theory Appl. 32 (1980) 171–182.

[17] R. Lepp, Discrete approximation of linear two-stage stochastic programming problem, Numer. Functional Anal. Optim. 9 (1987) 19–33.

[18] P. Olsen, Multistage stochastic programming with recourse as mathematical programming in an L_p-space, SIAM J. Contr. Optim. 14 (1976) 528–537.

[19] P. Olsen, Discretizations of multistage stochastic programming problems, Math. Progr. Study 6 (1976) 111–124.

[20] M.V.F. Pereira and L.M.V.G. Pinto, Multi-stage stochastic optimization applied to energy planning, Math. Progr. 52 (1991) 359–375.

[21] S.T. Rachev, *Probability Metrics and the Stability of Stochastic Models* (Wiley, Chichester, 1991).

[22] S.M. Robinson, Stability theory for systems of inequalities. Part I: Linear systems, SIAM J. Numer. Anal. 12 (1975) 754–769.

[23] S.M. Robinson and R.J.-B. Wets, Stability in two-stage stochastic programming, SIAM J. Contr. Optim. 25 (1987) 1409–1416.

[24] R.T. Rockafellar, Level sets and continuity of conjugate convex functions, Trans. Amer. Math. Soc. 123 (1966) 46–63.

[25] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970).

[26] R.T. Rockafellar, Integral functionals, normal integrands and measurable selections, in: *Nonlinear Operators and the Calculus of Variations*, eds. G.P. Gossez et al., Lecture Notes in Mathematics Vol. 543 (Springer, New York, 1976) pp. 157–207.

[27] R.T. Rockafellar and R.J.-B. Wets, Nonanticipativity and $\mathscr{L}^1$-martingales in stochastic optimization problems, Math. Progr. Study 6 (1976) 170–187.

[28] R.T. Rockafellar and R.J.-B. Wets, The optimal recourse problem in discrete time: L^1 multipliers for inequality constraints, SIAM J. Contr. Optim. 16 (1978) 16–36.
[29] W. Römisch and R. Schultz, Stability analysis for stochastic programs, Ann. Oper. Res. 30 (1991) 241–266.
[30] W. Römisch and R. Schultz, Stability of solutions for stochastic programs with complete recourse, Math. Oper. Res. 18 (1993) 590–609.
[31] W. Römisch and A. Wakolbinger, Obtaining convergence rates for approximations in stochastic programming, in: *Parametric Optimization and Related Topics*, eds. J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička (Akademie-Verlag, Berlin, 1987) pp. 327–343.
[32] R.J.-B. Wets, Stochastic programs with recourse: A basic theorem for multistage problems, Zeits. Wahrscheinlichkeitstheorie und verwandte Gebiete 21 (1972) 201–206.
[33] R.J.-B. Wets, On the relation between stochastic and deterministic optimization, in: *Control Theory, Numerical Methods and Computer Systems Modelling*, eds. A. Bensoussan and J.L. Lions, Lecture Notes in Economics and Mathematical Systems Vol. 107 (Springer, Berlin, 1975) pp. 350–361.
[34] R.J.-B. Wets, Stochastic programming, in: *Handbooks in Operations Research and Management Science*, Vol. 1: *Optimization*, eds. G.L. Nemhauser, A.H.G. Rinnooy Kan and M.J. Todd (North-Holland, Amsterdam, 1989) pp. 573–629.
[35] V.M. Zolotarev, Metric distances in spaces of random variables and their distributions, Math. USSR Sbornik 30 (1976) 373–401.
[36] J. Zowe and S. Kurcyusz, Regularity and stability of the mathematical programming problem in Banach spaces, Appl. Math. Optim. 5 (1979) 49–62.