SIAM J. OPTIM.
Vol. 17, No. 2, pp. 511–525

STABILITY OF MULTISTAGE STOCHASTIC PROGRAMS*

H. HEITSCH†, W. RÖMISCH†, AND C. STRUGAREK‡

Abstract. Quantitative stability of linear multistage stochastic programs is studied. It is shown that the infima of such programs behave (locally) Lipschitz continuous with respect to the sum of an L_r-distance and of a distance measure for the filtrations of the original and approximate stochastic (input) processes. Various issues of the result are discussed and an illustrative example is given. Consequences for the reduction of scenario trees are also discussed.

Key words. stochastic programming, multistage, nonanticipativity, stability, filtration, probability metrics

AMS subject classification. 90C15

DOI. 10.1137/050632865

1. Introduction.

We consider a finite horizon sequential decision process under uncertainty, in which a decision made at t is based only on information available at t $(1 \leq t \leq T)$. We assume that the information is given by a discrete time multivariate stochastic process $\{\xi_t\}_{t=1}^T$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with ξ_t taking values in $\mathbb{R}^d$. The information available at t consists of the random vector $\xi^t := (\xi_1, \dots, \xi_t)$, and the stochastic decision x_t at t varying in $\mathbb{R}^{m_t}$ is assumed to depend only on ξ^t. The latter property is called *nonanticipativity* and is equivalent to the measurability of x_t with respect to the σ-field $\mathcal{F}_t \subseteq \mathcal{F}$, which is generated by ξ^t. Hence, we have $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \dots, T-1$ and we assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$, i.e., ξ_1 and x_1 are deterministic and, with no loss of generality, that $\mathcal{F}_T = \mathcal{F}$.
More precisely, we consider the following *linear multistage stochastic program*:

$$(1.1)\ \min \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle\right] \;\middle|\; \begin{array}{l} x_t \in X_t, \\ x_t \text{ is } \mathcal{F}_t\text{-measurable}, t = 1, \dots, T, \\ A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t), t = 2, \dots, T \end{array} \right\},$$

where the subsets X_t of $\mathbb{R}^{m_t}$ are nonempty, closed, and polyhedral; the cost coefficients $b_t(\xi_t)$ belong to $\mathbb{R}^{m_t}$; the right-hand sides $h_t(\xi_t)$ are in $\mathbb{R}^{n_t}$; $A_{t,0}$ are fixed (n_t, m_t)-matrices; and $A_{t,1}(\xi_t)$ are (n_t, m_{t-1})-matrices, respectively. We assume that $b_t(\cdot)$, $h_t(\cdot)$, and $A_{t,1}(\cdot)$ depend affinely linearly on ξ_t covering the situation that some of the components of b_t and h_t, and of the elements of $A_{t,1}$ are random.

The challenge of multistage models consists in the presence of two groups of entirely different constraints, namely of measurability and of pointwise constraints for the decisions x_t. This fact does not lead to consequences in the two-stage situation $(T = 2)$. In general, however, it is the origin of both the theoretical and computa-

*Received by the editors June 1, 2005; accepted for publication (in revised form) February 9, 2006; published electronically August 16, 2006. This work was supported by the DFG Research Center MATHEON (Mathematics for key technologies) in Berlin and by a grant of EDF (Electricité de France).

http://www.siam.org/journals/siopt/17-2/63286.html

†Institute of Mathematics, Humboldt-University Berlin, D-10099 Berlin, Germany (heitsch@math.hu-berlin.de, romisch@math.hu-berlin.de).

‡EdF R&D, OSIRIS, 1 Avenue du Général de Gaulle F-92141 Clamart Cedex, France (cyrille.strugarek@edf.fr), Ecole Nationale des Ponts et Chaussées, Paris, France, and Ecole Nationale Supérieure de Techniques Avancées, Paris, France.

tional challenges of multistage models. In the present paper, it produces the essential difference of quantitative stability estimates compared to the two-stage case.

When solving multistage models computationally, the first step consists of approximating the stochastic process $\xi = \{\xi_t\}_{t=1}^T$ by a process having finitely many scenarios that exhibit tree structure and have its root at the fixed element ξ_1 of $\mathbb{R}^d$ (see the survey [4] for further information). In this way, both the random vectors ξ^t and the σ-fields $\mathcal{F}_t$ are approximated at each t. This process finally leads to linear programming models that are very large scale in most cases and may be solved by decomposition methods that exploit specific structures of the model (see [31] for additional background). In order to reduce the model dimension, it might be desirable to reduce the originally designed tree. The approaches to scenario reduction in [5, 11] and to scenario tree generation in [21, 14, 10] make use of probability metrics, i.e., of metric distances on spaces of probability measures, where the metrics are selected such that the optimal values of original and approximate stochastic program are close if the distance of the original probability distribution $P = \mathcal{L}(\xi)$ of ξ and its approximation Q is small.

Such quantitative stability results are well developed for two-stage models (cf. the survey [28]). It turned out that distances of probability measures are relevant which are given by certain Monge–Kantorovich mass transportation problems. Such problems are of the form

$$\inf\Big\{\int_{\Xi\times\Xi} c(\xi,\tilde{\xi})\eta(d\xi,d\tilde{\xi}) : \eta\in\mathcal{P}(\Xi\times\Xi),\ \pi_1\eta = P,\ \pi_2\eta = Q\Big\}, \tag{1.2}$$

where Ξ is a closed subset of some Euclidean space, π_1 and π_2 denote the projections onto the first and second components, respectively, c is a nonnegative, symmetric, and continuous cost function and P and Q belong to a set $\mathcal{P}_c(\Xi)$ of probability measures on Ξ, where all integrals are finite. Two types of cost functions have been used in stability analysis of stochastic programs [5, 29], namely,

$$c(\xi,\tilde{\xi}) := \|\xi-\tilde{\xi}\|^r \quad (\xi,\tilde{\xi}\in\Xi) \tag{1.3}$$

and

$$c(\xi,\tilde{\xi}) := \max\{1, \|\xi-\xi_0\|^{r-1}, \|\tilde{\xi}-\xi_0\|^{r-1}\}\|\xi-\tilde{\xi}\| \quad (\xi,\tilde{\xi}\in\Xi) \tag{1.4}$$

for some $r \geq 1$ and $\xi_0 \in \Xi$. In both cases, the set $\mathcal{P}_c(\Xi)$ may be chosen as the set $\mathcal{P}_r(\Xi)$ of all probability measures on Ξ having absolute moments of order r. The cost (1.3) leads to L_r*-minimal metrics* ℓ_r [25], which are defined by

$$\ell_r(P,Q) := \inf\left\{\int_{\Xi\times\Xi} \|\xi-\tilde{\xi}\|^r \eta(d\xi,d\tilde{\xi})\,|\,\eta\in\mathcal{P}(\Xi\times\Xi),\ \pi_1\eta = P,\ \pi_2\eta = Q\right\}^{\frac{1}{r}} \tag{1.5}$$

and sometimes also called Wasserstein metrics of order r [9]. The mass transportation problem (1.2) with cost (1.4) defines the Monge–Kantorovich functionals $\hat{\mu}_r$ [22, 24]. A variant of the functional $\hat{\mu}_r$ appears if, in its definition by (1.2), the conditions $\eta \in \mathcal{P}(\Xi\times\Xi), \pi_1\eta = P,\ \pi_2\eta = Q$ are replaced by η being a finite measure on $\Xi\times\Xi$ such that $\pi_1\eta - \pi_2\eta = P - Q$. The corresponding functionals $\mathring{\mu}_r$ are smaller than $\hat{\mu}_r$ and turn out to be metrics on $\mathcal{P}_r(\Xi)$. They are called Fortet–Mourier metrics of order r [8, 22]. The convergence of sequences of probability measures, with respect to both metrics ℓ_r and $\mathring{\mu}_r$, is equivalent to their weak convergence and the convergence of their rth order absolute moments. For $r = 1$ we have the identity $\mathring{\mu}_1 = \hat{\mu}_1 = \ell_1$ and the corresponding metric is also called Kantorovich distance. Two-stage models are known to behave stable with respect to Fortet–Mourier metrics [23].

Much less is known, however, of the multistage case. The present paper may be regarded as an extension of the quantitative analysis in [7], which considers a less general probabilistic setup and assumes implicitly that the filtrations of the original and approximate stochastic processes coincide. The paper [19] and the recent work [20] provide (qualitative) convergence results of approximations and [16, 32] deal with empirical estimates in multistage models. In the recent paper [34] the role of probability metrics for studying stability of multistage models is questioned critically. An example is given showing that closeness of original and approximate probability distributions in terms of some probability metric is not sufficient for the infima to be close in general. The recent thesis [1] focuses precisely on the question of information in stochastic programs. The conclusions of this work do not address stability, but only discretization of multistage stochastic programs. They illuminate the role which should be played by σ-field distances in order to obtain a consistent discretization of such programs.

The main result of the present paper (Theorem 2.1) provides stability of infima of the multistage model (1.1) with respect to a sum of the L_r-norm and of a distance of the information structures, i.e., the filtrations of σ-fields, of the original and approximate stochastic (input) processes. Hence, it enlightens the corresponding arguments in [34]. Several comments are given on the stability result, its assumptions, the filtration distance, and on the choice of the underlying probability space if the original and approximate (input) probability distributions are given in practical models. Furthermore, we provide an illustrative example which shows that the filtration distance is indispensable for stability (Example 2.6). Finally, some consequences for designing scenario reduction schemes in multistage models are discussed.

2. Stability of multistage models. Under weak hypotheses, the program (1.1) can be equivalently reformulated as a minimization problem for the deterministic first stage decision x_1 (see [31, Chapter 1] or [6, 26] for example). It is of the form

$$\min\left\{\mathbb{E}[f(x_1,\xi)] = \int_\Xi f(x_1,\xi)P(d\xi) : x_1 \in X_1\right\}, \tag{2.1}$$

where Ξ is a closed subset of $\mathbb{R}^{Td}$ containing the support of the probability distribution P of ξ, and f is an integrand on $\mathbb{R}^{m_1}\times\Xi$ given by the dynamic programming recursion

$$\begin{aligned} f(x_1,\xi) &:= \Phi_1(x_1,\xi^1) = \langle b_1(\xi_1), x_1\rangle + \Phi_2(x_1,\xi^2), \qquad (2.2)\\ \Phi_t(x_1,\dots,x_{t-1},\xi^t) &:= \inf\Big\{\langle b_t(\xi_t), x_t\rangle + \mathbb{E}\left[\Phi_{t+1}(x_1,\dots,x_t,\xi^{t+1})|\mathcal{F}_t\right] : x_t \in X_t,\\ &\qquad x_t \text{ is } \mathcal{F}_t\text{-measurable}, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)\Big\}\\ &\qquad (t = 2,\dots,T),\\ \Phi_{T+1}(x_1,\dots,x_T,\xi^{T+1}) &:= 0. \end{aligned}$$

Using the representation (2.2) of the integrand f for $T = 2$ quantitative stability results are proved in [23, 28] with respect to Fortet–Mourier metrics of probability distributions and earlier in [29] with respect to L_r-minimal metrics. For $T > 2$, however, the integrand f depends on conditional expectations with respect to the σ-fields $\mathcal{F}_t$ and, hence, on the underlying probability measure $\mathbb{P}$ in a nonlinear way. Consequently, the methodology for studying quantitative stability properties of stochastic programs of the form (2.1) developed in [23, 28] does not apply to multistage models in general.

An alternative for studying stability of multistage models consists in considering them as optimization problems in functional spaces (see also [18, 26]), where the Banach spaces $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ with $m = \sum_{t=1}^T m_t$ and endowed with the norm

$$\|x\|_{r'} := \left(\sum_{t=1}^T \mathbb{E}[\|x_t\|^{r'}]\right)^{\frac{1}{r'}} \quad \text{for } r' \geq 1 \text{ and } \quad \|x\|_\infty := \max_{t=1,\dots,T} \operatorname{ess\,sup} \|x_t\|$$

are appropriate. Here, the stochastic input process ξ belongs to $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ for some $r \geq 1$ and $s := Td$, and r' is defined by

$$r' := \begin{cases} \frac{r}{r-1} & \text{if only costs are random,} \\ r & \text{if only right-hand sides are random,} \\ r = 2 & \text{if only costs and right-hand sides are random,} \\ \infty & \text{if all technology matrices are random and } r = T. \end{cases}$$

The number r corresponds to the order of (absolute) moments of ξ that are required to exist. The definition of the numbers r' implies that the objective function is well defined and finite. In the third case it may alternatively be required that the costs $b_t(\xi_t)$ have finite moments of order $\hat{r} \geq 1$. Then we choose $r' := \frac{\hat{r}}{\hat{r}-1}$ and require that $h_t(\xi_t)$ belongs to $L_{r'}$.

Let us introduce some notations. Let F denote the objective function defined on $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) \times L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) \to \mathbb{R}$ by $F(\xi, x) := \mathbb{E}[\sum_{t=1}^T \langle b_t(\xi_t), x_t \rangle]$, let

$$\mathcal{X}_t(x_{t-1}; \xi_t) := \{x_t \in X_t | A_{t,0} x_t + A_{t,1}(\xi_t) x_{t-1} = h_t(\xi_t)\}$$

denote the tth feasibility set for every $t = 2, \dots, T$, and

$$\mathcal{X}(\xi) := \{x \in \times_{t=1}^T L_{r'}(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{m_t}) | x_1 \in X_1, x_t \in \mathcal{X}_t(x_{t-1}; \xi_t), t = 2, \dots, T\}$$

denote the set of feasible elements of the stochastic program (1.1) with input ξ. Then the stochastic program (1.1) may be rewritten in the form

$$\min\{F(\xi, x) : x \in \mathcal{X}(\xi)\}. \tag{2.3}$$

Let $v(\xi)$ denote the optimal value of (2.3) and let, for any $\alpha \geq 0$,

$$l_\alpha(F(\xi, \cdot)) := \{x \in \mathcal{X}(\xi) : F(\xi, x) \leq v(\xi) + \alpha\}$$

denote its α-*level set*. The following conditions are imposed on (2.3).

(A1) There exists a $\delta > 0$ such that for any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$, any $t = 2, \dots, T$ and any $x_1 \in X_1$, $x_\tau \in L_{r'}(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{m_\tau})$ with $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$, $\tau = 2, \dots, t-1$, the tth feasibility set $\mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ is nonempty (*relatively complete recourse locally around* ξ).

(A2) The optimal values $v(\tilde{\xi})$ of (2.3) with input $\tilde{\xi}$ are finite for all $\tilde{\xi}$ in a neighborhood of ξ and the objective function F is *level-bounded locally uniformly at* ξ, i.e., for some $\alpha > 0$ there exists a $\delta > 0$ and a bounded subset B of $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ such that $l_\alpha(F(\tilde{\xi}, \cdot))$ is nonempty and contained in B for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

(A3) $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ for some $r \geq 1$.

To state our main result we introduce the distance $D_{\mathrm{f}}(\xi, \tilde{\xi})$ of the filtrations of ξ and its approximation (or perturbation) $\tilde{\xi}$, respectively. It is defined by

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) := \sup_{\varepsilon \in (0,\alpha]} D_{\mathrm{f},\varepsilon}(\xi, \tilde{\xi}) \tag{2.4}$$

and $D_{\mathrm{f},\varepsilon}(\xi, \tilde{\xi})$ denotes the ε-filtration distance given by

$$D_{\mathrm{f},\varepsilon}(\xi, \tilde{\xi}) := \inf \sum_{t=2}^{T-1} \max\{\|x_t - \mathbb{E}[x_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\}, \tag{2.5}$$

where the infimum is taken with respect to all $x \in l_\varepsilon(F(\xi, \cdot))$ and $\tilde{x} \in l_\varepsilon(F(\tilde{\xi}, \cdot))$, respectively, i.e., with respect to all feasible decisions belonging to the ε-level sets of the original and perturbed programs. Furthermore, $\mathcal{F}_t$ and $\tilde{\mathcal{F}}_t$, $t = 1, \ldots, T$, denote the filtrations of ξ and $\tilde{\xi}$, respectively.

Now, we are ready to state our main stability result for multistage stochastic programs.

THEOREM 2.1. *Let* (A1), (A2), *and* (A3) *be satisfied and* X_1 *be bounded. Then there exists positive constants* L, α, *and* δ *such that the estimate*

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}(\xi, \tilde{\xi})) \tag{2.6}$$

holds for all random elements $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ *with* $\|\tilde{\xi} - \xi\|_r \leq \delta$.

Proof. Let M_t denote the set-valued mappings $u \mapsto \{x \in \mathbb{R}^{m_t} | A_{t,0}x = u, x \in X_t\}$ from $\mathbb{R}^{n_t}$ to $\mathbb{R}^{m_t}$ for $t = 2, \ldots, T$. The mappings have polyhedral graph and, hence, are Lipschitz continuous with respect to the Hausdorff distance on their domain $\operatorname{dom} M_t \subseteq \mathbb{R}^{n_t}$ [27, Example 9.35]. Hence, there exist positive constants l_t such that we have

$$\sup_{x \in M_t(\bar{u})} d(x, M_t(\tilde{u})) \leq l_t \|\bar{u} - \tilde{u}\| \tag{2.7}$$

for all $\bar{u}, \tilde{u} \in \operatorname{dom} M_t$, where $d(x, C)$ denotes the distance of x to a nonempty set C in $\mathbb{R}^{m_t}$.

Now, let $\alpha > 0$ and $\delta > 0$ be selected as in (A1) and (A2). Let $\varepsilon \in (0, \alpha]$, $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ be such that $\|\tilde{\xi} - \xi\|_r < \delta$ and $v(\tilde{\xi}) \in \mathbb{R}$, and let $\bar{x} \in l_\varepsilon(F(\xi, \cdot))$. By $\tilde{\mathcal{F}}_t$ we denote the σ-field generated by $\tilde{\xi}^t := (\tilde{\xi}_1, \ldots, \tilde{\xi}_t)$ for $t = 1, \ldots, T$. Now, we show recursively the existence of constants $\hat{L}_t > 0$ and of elements $\tilde{x}_t$ belonging to the appropriate spaces $L_{r'}(\Omega, \tilde{\mathcal{F}}_t, \mathbb{P}; \mathbb{R}^{m_t})$ for each $t = 1, \ldots, T$ such that $\tilde{x}_t \in X_t$, $t = 1, \ldots, T$, $A_{t,0}\tilde{x}_t + A_{t,1}(\tilde{\xi}_t)\tilde{x}_{t-1} = h_t(\tilde{\xi}_t)$, $t = 2, \ldots, T$, and that

$$\|\mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t] - \tilde{x}_t\|$$

can be estimated recursively with respect to t. Let $t = 1$, we then set $\tilde{x}_1 := \bar{x}_1$ and $\hat{L}_1 := 1$. For $t > 1$, we assume that $\hat{L}_{t-1}$ and $\tilde{x}_{t-1}$ have already been constructed, set $\bar{u}_t := h_t(\xi_t) - A_{t,1}(\xi_t)\bar{x}_{t-1}$, $\tilde{u}_t := h_t(\tilde{\xi}_t) - A_{t,1}(\tilde{\xi}_t)\tilde{x}_{t-1}$ and consider the following set-valued mappings from Ω to $\mathbb{R}^{m_t}$ given by

$$\omega \to M_t(\tilde{u}_t(\omega)) \qquad \text{and} \qquad \omega \to \arg \min_{x \in M_t(\tilde{u}_t(\omega))} \|\mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t](\omega) - x\|.$$

Both are measurable with respect to the σ-field $\tilde{\mathcal{F}}_t$ due to the measurability of $\tilde{x}_{t-1}$ with respect to $\tilde{\mathcal{F}}_{t-1}$ and well-known measurability results for set-valued mappings (e.g., [27, Theorem 14.36]). In addition, the set-valued mapping $\omega \to M_t(\tilde{u}_t(\omega))$ is nonempty-valued due to (A1). Hence, by appealing to [27, Theorem 14.37] there exists a $\tilde{\mathcal{F}}_t$-measurable selection $\tilde{x}_t$ of the second mapping. Since $\mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]$ belongs to $M_t(\mathbb{E}[\bar{u}_t|\tilde{\mathcal{F}}_t])$, (2.7) provides the estimate

$$\begin{aligned}
\|\mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t] - \tilde{x}_t\| &\leq l_t\|\mathbb{E}[\bar{u}_t|\tilde{\mathcal{F}}_t] - \tilde{u}_t\| \\
&\leq l_t(\|\mathbb{E}[h_t(\xi_t)|\tilde{\mathcal{F}}_t] - h_t(\tilde{\xi}_t)\| + \|\mathbb{E}[A_{t,1}(\xi_t)\bar{x}_{t-1}|\tilde{\mathcal{F}}_t] - A_{t,1}(\tilde{\xi}_t)\tilde{x}_{t-1}\|) \\
&\leq l_t(K_t\|\mathbb{E}[\xi_t|\tilde{\mathcal{F}}_t] - \tilde{\xi}_t\| + \|\mathbb{E}[A_{t,1}(\xi_t)\bar{x}_{t-1} - A_{t,1}(\tilde{\xi}_t)\bar{x}_{t-1}|\tilde{\mathcal{F}}_t]\| \\
&\qquad + \|A_{t,1}(\tilde{\xi}_t)\|\|\mathbb{E}[\bar{x}_{t-1}|\tilde{\mathcal{F}}_t] - \tilde{x}_{t-1}\|) \\
&\leq l_t\bar{K}_t(\|\mathbb{E}[\xi_t - \tilde{\xi}_t|\tilde{\mathcal{F}}_t]\| + \mathbb{E}[\|\xi_t - \tilde{\xi}_t\|\|\bar{x}_{t-1}\||\tilde{\mathcal{F}}_t] \\
&\qquad + \max\{1, \|\tilde{\xi}_t\|\}(\|\mathbb{E}[\bar{x}_{t-1} - \mathbb{E}[\bar{x}_{t-1}|\tilde{\mathcal{F}}_{t-1}]|\tilde{\mathcal{F}}_t]\| \\
&\qquad + \|\mathbb{E}[\bar{x}_{t-1}|\tilde{\mathcal{F}}_{t-1}] - \tilde{x}_{t-1}\|)),
\end{aligned}$$

where K_t and $\bar{K}_t$ are certain positive constants, the affine linearity of $h_t(\cdot)$ and $A_{t,1}(\cdot)$ and Jensen's inequality is used for the second summand. Clearly, we have $\|\tilde{\xi}_\tau\| \leq C\|\tilde{\xi}^t\|$ with some constant C for all $\tau = 2, \ldots, t$, $t = 2, \ldots, T$, and the corresponding norms in $\mathbb{R}^d$ and $\mathbb{R}^{td}$. Using Jensen's inequality also in the first and third summand of the latter estimate we obtain recursively

$$\begin{aligned}
\|\mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t] - \tilde{x}_t\| \leq \hat{L}_t\Big(&\sum_{\tau=2}^{t} \max\{1, \|\tilde{\xi}^t\|^{t-\tau}\}\mathbb{E}[(1 + \|\bar{x}_{\tau-1}\|)\|\xi_\tau - \tilde{\xi}_\tau\|\,|\tilde{\mathcal{F}}_\tau] \\
&+ \sum_{\tau=2}^{t-1} \max\{1, \|\tilde{\xi}^t\|^{t-\tau}\}\mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau|\tilde{\mathcal{F}}_\tau]\|\,|\tilde{\mathcal{F}}_{\tau+1}]\Big)
\end{aligned} \tag{2.8}$$

with some positive constant $\hat{L}_t$ for $t = 2, \ldots, T$. Note that the sum on the right-hand side of (2.8) disappears if only costs are random. The max-terms in (2.8) and the norms $\|x_{\tau-1}\|$ in (2.8) vanish if the technology matrices are not random. Inserting $\bar{x}$ and $\tilde{x}$ into the objective function we obtain

$$v(\tilde{\xi}) - v(\xi) \leq F(\tilde{\xi}, \tilde{x}) - F(\xi, \bar{x}) + \varepsilon. \tag{2.9}$$

In case of only right-hand sides being random we continue (2.9) using (2.8) and obtain

$$\begin{aligned}
v(\tilde{\xi}) - v(\xi) &\leq \sum_{t=2}^{T} \mathbb{E}[\langle b_t, \mathbb{E}[\tilde{x}_t - \bar{x}_t|\tilde{\mathcal{F}}_t]\rangle] + \varepsilon \leq \sum_{t=2}^{T} \|b_t\|\mathbb{E}[\|\tilde{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|] + \varepsilon \\
&\leq \hat{L}\sum_{t=2}^{T} \mathbb{E}\Big[\sum_{\tau=2}^{t} \mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\||\tilde{\mathcal{F}}_\tau] + \sum_{\tau=2}^{t-1} \mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau|\tilde{\mathcal{F}}_\tau]\||\tilde{\mathcal{F}}_{\tau+1}]\Big] + \varepsilon \\
&\leq \hat{L}T\mathbb{E}\Big[\sum_{t=2}^{T} \|\xi_t - \tilde{\xi}_t\| + \sum_{\tau=2}^{T-1} \|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau|\tilde{\mathcal{F}}_\tau]\|\Big] + \varepsilon \\
&\leq \hat{L}T\Big(\|\xi - \tilde{\xi}\|_r + \sum_{\tau=2}^{T-1} \|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau|\tilde{\mathcal{F}}_\tau]\|_r\Big) + \varepsilon,
\end{aligned}$$

where $\hat{L} := \max_{t=1,\ldots,T} \hat{L}_t\|b_t\|$. If costs are random, we obtain the estimate

$$v(\tilde{\xi}) - v(\xi) \leq F(\tilde{\xi}, \tilde{x}) - F(\tilde{\xi}, \bar{x}) + F(\tilde{\xi}, \bar{x}) - F(\xi, \bar{x}) + \varepsilon$$

$$
\begin{aligned}
&\leq \mathbb{E}\Big[\sum_{t=2}^{T}\langle b_t(\tilde{\xi}_t), \mathbb{E}[\tilde{x}_t - \bar{x}_t|\tilde{\mathcal{F}}_t]\rangle\Big] + \mathbb{E}\Big[\sum_{t=1}^{T}\langle b_t(\tilde{\xi}_t) - b_t(\xi_t), \bar{x}_t\rangle\Big] + \varepsilon \\
&\leq \hat{K}\mathbb{E}\Big[\sum_{t=2}^{T}\max\{1, \|\tilde{\xi}_t\|\}\|\tilde{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\| + \sum_{t=1}^{T}\|\tilde{\xi}_t - \xi_t\|\|\bar{x}_t\|\Big] + \varepsilon \qquad (2.10)
\end{aligned}
$$

with some positive constant $\hat{K}$. In case of only costs being random, i.e., $r' = \frac{r}{r-1}$, we continue with

$$
\begin{aligned}
v(\tilde{\xi}) - v(\xi) &\leq \hat{K}\mathbb{E}\Big[\sum_{t=2}^{T}\max\{1, \|\tilde{\xi}_t\|\}\|\tilde{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|\Big] + \hat{K}\|\tilde{\xi} - \xi\|_r\|\bar{x}\|_{r'} + \varepsilon \\
&\leq \hat{K}\mathbb{E}\Big[\sum_{t=2}^{T}\max\{1, \|\tilde{\xi}_t\|\}\|\tilde{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|\Big] + K\|\tilde{\xi} - \xi\|_r + \varepsilon,
\end{aligned}
$$

where Hölder's inequality and the boundedness of $\|\bar{x}\|_{r'}$ according to (A2) were used leading to some constant $K > 0$. Using the estimate (2.8), we conclude that

$$
v(\tilde{\xi}) - v(\xi) \leq L\Big(\|\tilde{\xi} - \xi\|_r + \sum_{t=2}^{T-1}\|\bar{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|_{r'}\Big) + \varepsilon,
$$

where Hölder's inequality and the fact that $\tilde{\xi}$ varies in a bounded set in L_r were used leading to some constant $L > 0$ (depending on ξ).

Next, we consider the case $r = r' = 2$. Starting from (2.10) we use the Cauchy–Schwarz inequality and obtain

$$
\begin{aligned}
v(\tilde{\xi}) - v(\xi) &\leq \hat{K}\Big[\Big(\sum_{t=2}^{T}\mathbb{E}[\max\{1, \|\tilde{\xi}_t\|^2\}]\Big)^{\frac{1}{2}}\Big(\sum_{t=2}^{T}\mathbb{E}[\|\tilde{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|^2]\Big)^{\frac{1}{2}} \\
&\qquad + \|\tilde{\xi} - \xi\|_2\|\bar{x}\|_2\Big] + \varepsilon \\
&\leq \Big(\|\tilde{\xi} - \xi\|_2 + \sum_{t=2}^{T-1}\|\bar{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|_2\Big) + \varepsilon
\end{aligned}
$$

with some constant $L > 0$ (depending on ξ) due to (2.8), (A2), and the fact that $\tilde{\xi}$ varies in some bounded set in L_2.

Finally, we consider the situation that costs, right-hand sides, and technology matrices are random, i.e., $r = T$ and $r' = \infty$. In this case, the estimate (2.8) attains the form

$$
\begin{aligned}
\|\mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t] - \tilde{x}_t\| \leq \hat{L}_t\Big(&\sum_{\tau=2}^{t}\max\{1, \|\tilde{\xi}^t\|^{t-\tau}\}\mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\|\,|\tilde{\mathcal{F}}_\tau] \\
&+ \sum_{\tau=2}^{t-1}\max\{1, \|\tilde{\xi}^t\|^{t-\tau}\}\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau|\tilde{\mathcal{F}}_\tau]\|_\infty\Big).
\end{aligned}
$$

Now, we start again from (2.10) and use the latter estimate and obtain

$$
v(\tilde{\xi}) - v(\xi) \leq \hat{L}\mathbb{E}\Big[\sum_{t=2}^{T}\Big(\sum_{\tau=2}^{t}\max\{1, \|\tilde{\xi}^t\|^{t+1-\tau}\}\mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\|\,|\tilde{\mathcal{F}}_\tau]
$$

$$
\begin{aligned}
&\qquad + \sum_{\tau=2}^{t-1} \max\{1, \|\tilde{\xi}^t\|^{t+1-\tau}\} \|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau|\tilde{\mathcal{F}}_\tau]\|_\infty \Big) + \sum_{t=1}^{T} \|\tilde{\xi}_t - \xi_t\| \Big] + \varepsilon \\
&\le \tilde{L}\mathbb{E}\Big[\sum_{t=2}^{T} \max\{1, \|\tilde{\xi}^t\|^{t-1}\} \mathbb{E}[\|\xi_t - \tilde{\xi}_t\| \,|\tilde{\mathcal{F}}_t] \Big] \qquad (2.11) \\
&\qquad + \sum_{t=2}^{T-1} \mathbb{E}[\max\{1, \|\tilde{\xi}^t\|^{t-1}\}] \|\bar{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|_\infty + \|\tilde{\xi} - \xi\|_1 + \varepsilon \\
&\le \bar{L}\mathbb{E}[\max\{1, \|\tilde{\xi}\|^T\}] \Big(\|\xi - \tilde{\xi}\|_T + \sum_{t=2}^{T-1} \|\bar{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|_\infty \Big) + \varepsilon,
\end{aligned}
$$

where $\hat{L}, \tilde{L}, \bar{L}$ are certain positive constants and Hölder's inequality was used. Since $\tilde{\xi}$ varies in a bounded subset of L_T, there exists a constant $L > 0$ (depending on ξ) such that

$$
v(\tilde{\xi}) - v(\xi) \le L\Big(\|\xi - \tilde{\xi}\|_r + \sum_{t=2}^{T-1} \|\bar{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|_{r'} \Big) + \varepsilon, \tag{2.12}
$$

where $r = T$ and $r' = \infty$. Hence, an estimate of the form (2.12) is obtained in all cases. Changing the role of ξ and $\tilde{\xi}$ leads to an estimate of the form

$$
v(\xi) - v(\tilde{\xi}) \le L\Big(\|\xi - \tilde{\xi}\|_r + \sum_{t=2}^{T-1} \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|_{r'} \Big) + \varepsilon. \tag{2.13}
$$

We note that the second summands in the estimates (2.12) and (2.13) are bounded by

$$
\sum_{t=2}^{T-1} \max\{\|\bar{x}_t - \mathbb{E}[\bar{x}_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\}. \tag{2.14}
$$

Since the estimates (2.12) and (2.13) are valid for all $\bar{x} \in l_\varepsilon(F(\xi,\cdot))$ and $\tilde{x} \in l_\varepsilon(F(\tilde{\xi},\cdot))$, we arrive at the estimate

$$
|v(\xi) - v(\tilde{\xi})| \le L\Big(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f},\varepsilon}(\xi,\tilde{\xi}) \Big) + \varepsilon \le L\Big(\|\xi - \tilde{\xi}\|_r + \sup_{\varepsilon \in (0,\alpha]} D_{\mathrm{f},\varepsilon}(\xi,\tilde{\xi}) \Big) + \varepsilon.
$$

Finally, it remains to take the infimum of the right-hand side with respect to $\varepsilon > 0$ and the proof is complete. ∎

Remark 2.2. A sufficient condition for (A1) to hold is the *complete fixed recourse* condition on all matrices $A_{t,0}$, i.e., the sets X_t are polyhedral cones and $A_{t,0}X_t = \mathbb{R}^{n_t}$ holds for $t = 2, \ldots, T$. Assumption (A2) on the locally uniform level-boundedness of the objective function F is quite standard in perturbation results for optimization problems (see, e.g., [27, Theorem 1.17]). The finiteness condition for the optimal values is needed because it is not implied by the level-boundedness of F for all relevant pairs (r, r'). In the case that Ω is finite or $1 < r' < \infty$, the existence of solutions of (2.3) (and, thus, the finiteness of $v(\xi)$) is a simple consequence of the compactness or the weak sequential compactness of $l_\alpha(F(\xi,\cdot))$ in the reflexive Banach space $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ and of the linearity of the objective. Then the filtration distance is of the form

$$
D_{\mathrm{f}}(\xi,\tilde{\xi}) = \inf\Big\{ \sum_{t=2}^{T-1} D_t(\xi,\tilde{\xi}) : x \in l_0(F(\xi,\cdot)), \tilde{x} \in l_0(F(\tilde{\xi},\cdot)) \Big\}, \tag{2.15}
$$

where $D_t(\xi,\tilde{\xi})$ is defined by

$$D_t(\xi,\tilde{\xi}) := \max\{\|x_t - \mathbb{E}[x_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\} \\ = \max\{\|x_t - \mathbb{E}[x_t|\tilde{\xi}_1,\dots,\tilde{\xi}_t]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\xi_1,\dots,\xi_t]\|_{r'}\}. \tag{2.16}$$

Remark 2.3. In practical situations, the available knowledge on the stochastic input consists in (partial or complete) information on its probability distribution. Which probability space should be selected? A natural answer certainly is: Take a probability space where the L_r-distance $\|\xi-\tilde{\xi}\|_r$ and the $L_{r'}$-distances $\|x_t-\mathbb{E}[x_t|\tilde{\xi}_1,\dots,\tilde{\xi}_t]\|_{r'}$ and $\|\tilde{x}_t-\mathbb{E}[\tilde{x}_t|\xi_1,\dots,\xi_t]\|_{r'}$, $t=2,\dots,T-1$, are minimal. Let us explain this minimality condition in case of the L_r-distance $\|\xi-\tilde{\xi}\|_r$. Let P and Q in $\mathcal{P}_r(\Xi)$ be the probability distributions of ξ and $\tilde{\xi}$. Then there exists an optimal solution $\eta^* \in \mathcal{P}(\Xi\times\Xi)$ of the mass transportation problem (1.5) [22, Theorem 8.1.1], i.e.,

$$\ell_r^r(P,Q) = \int_{\Xi\times\Xi} \|\xi-\tilde{\xi}\|^r \eta^*(d\xi,d\tilde{\xi}),$$

where $\pi_1\eta^* = P$ and $\pi_2\eta^* = Q$. Furthermore, there exists a probability space $(\Omega',\mathcal{F}',\mathbb{P}')$ and an optimal coupling, i.e., a pair $(\xi'(\cdot),\tilde{\xi}'(\cdot))$ of Ξ-valued random elements defined on it, such that the probability distribution of $(\xi'(\cdot),\tilde{\xi}'(\cdot))$ is just η^* [22, Theorem 2.5.1]. In particular, we have that the distance in $L_r(\Omega',\mathcal{F}',\mathbb{P}';\mathbb{R}^s)$ is just the L_r-minimal distance of the probability distributions, i.e.,

$$\ell_r(P,Q) = \|\xi'(\cdot)-\tilde{\xi}'(\cdot)\|_r.$$

In the same way, the relevant minimal $L_{r'}$-distances $\|x_t - \mathbb{E}[x_t|\tilde{\xi}_1,\dots,\tilde{\xi}_t]\|_{r'}$ and $\|\tilde{x}_t-\mathbb{E}[\tilde{x}_t|\xi_1,\dots,\xi_t]\|_{r'}$ correspond to the $\ell_{r'}$-distance of the probability distributions of $x(t)$ and $\mathbb{E}[x_t|\tilde{\xi}_1,\dots,\tilde{\xi}_t]$, and of $\tilde{x}(t)$ and $\mathbb{E}[\tilde{x}_t|\xi_1,\dots,\xi_t]$, respectively.

Remark 2.4 (stability of first-stage solutions). Using the same technique as for proving [28, Theorem 9], the continuity property of infima in Theorem 2.1 can be supplemented by a quantitative stability property of the solution set $S(\xi)$ of (2.1), i.e., of the set of first stage solutions. Namely, there exists a constant $\hat{L}>0$ such that

$$\sup_{x\in S(\tilde{\xi})} d(x,S(\xi)) \le \Psi_\xi^{-1}(\hat{L}(\|\xi-\tilde{\xi}\|_r + D_{\mathrm{f}}(\xi,\tilde{\xi}))), \tag{2.17}$$

where $\Psi_\xi(\tau) := \inf\left\{\mathbb{E}[f(x_1,\xi)] - v(\xi) : d(x_1,S(\xi)) \ge \tau,\, x_1\in X_1\right\}$ with $\Psi_\xi^{-1}(\alpha) := \sup\{\tau\in\mathbb{R}_+ : \Psi_\xi(\tau)\le\alpha\}$ $(\alpha\in\mathbb{R}_+)$ is the growth function of the original problem (2.1) near its solution set $S(\xi)$. The boundedness condition for X_1 in Theorem 2.1 can be relaxed to the assumption that the set $S(\xi)$ is bounded. In the latter case a version of (2.6) is derived that contains localized optimal values. Then the estimate (2.6) is valid whenever its right-hand side is sufficiently small.

Remark 2.5 (convergence of filtrations). This remark aims at precising the link between the filtration distance (2.4) and previous work on *convergence of information.* A distance between σ-fields was introduced in [2]. It metrizes a topology called uniform topology on the set of σ-fields. Due to the work of [30] and [17], this distance reads, for all $\mathcal{B},\mathcal{B}'$ sub-σ-fields of $\mathcal{F}$

$$d_B(\mathcal{B},\mathcal{B}') := \sup_{f\in\Phi} \|\mathbb{E}[f|\mathcal{B}] - \mathbb{E}[f|\mathcal{B}']\|_1, \tag{2.18}$$

with Φ the set of all $\mathcal{F}$-measurable functions f such that for all $\omega\in\Omega$, $\|f(\omega)\|\le 1$. Thanks to [15], a filtration can be said to converge to another one if and only if each σ-field at each time step converges according to the distance d_B. Hence, a distance between filtrations can be introduced, based on the sum of the distances between σ-fields. The second summand in our stability result can be seen as such a distance between the filtrations generated by the two stochastic processes ξ and $\tilde{\xi}$. This summand is not exactly the same as the sum of distances d_B, but it has the same sense: If the feasible set of the stochastic program is bounded, the filtration distance (2.4) is bounded by a sum of distances d_B. Other distances between filtrations and σ-fields have been introduced (see, e.g., [3]) to fit with stochastic optimization problems. The thesis [1] provides a good survey and a few new results on the application of such information distances.

The following example shows that filtration distances are indispensable for the stability of multistage models.

Example 2.6. We consider a multistage stochastic program that models the optimal purchase over time under cost uncertainty. Its decisions x_t correspond to the amounts to be purchased at each time period. The uncertain prices are ξ_t, $t = 1, \ldots, T$, and the objective consists in minimizing the expected costs such that a prescribed amount $a > 0$ is achieved at the end of a given time horizon. The problem is of the form

$$\min \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \xi_t x_t\right] \;\middle|\; \begin{array}{l} (x_t, s_t) \in X_t = \mathbb{R}^2_+, \\ (x_t, s_t) \text{ is } \mathcal{F}_t\text{-measurable}, \\ s_t - s_{t-1} = x_t,\ t = 2, \ldots, T, \\ s_1 = 0, \quad s_T = a \end{array} \right\},$$

where the state variable s_t corresponds to the amount at time t and $\mathcal{F}_t := \sigma\{\xi_1, \ldots, \xi_t\}$. Let $T := 3$ and P_ε denote the probability distribution of the stochastic price process. P_ε is given by the two scenarios $\xi^1_\varepsilon = (3, 2+\varepsilon, 3)$ ($\varepsilon \in [0,1)$) and $\xi^2_\varepsilon = (3, 2, 1)$ each endowed with probability $\frac{1}{2}$. Let $Q := P_0$ denote the approximation of P_ε given by the two scenarios $\tilde{\xi}^1 = (3, 2, 3)$ and $\tilde{\xi}^2 = (3, 2, 1)$ with the same probabilities $\frac{1}{2}$. We assume that the scenario trees of the processes ξ_ε and $\tilde{\xi}$ are of the form displayed in Figure 2.1, i.e., the filtrations of σ-fields generated by ξ_ε and $\tilde{\xi}$ do not coincide.

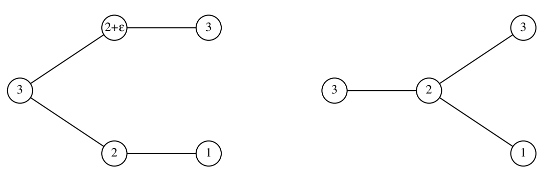

FIG. 2.1. *Scenario trees for P_ε (left) and Q.*

We obtain

$$v(\xi_\varepsilon) = \frac{3+\varepsilon}{2} a \quad \text{and} \quad v(\tilde{\xi}) = 2a\,, \text{ but} \quad \ell_1(P_\varepsilon, Q) = \|\xi_\varepsilon - \tilde{\xi}\|_1 = \frac{\varepsilon}{2}.$$

Hence, the multistage stochastic purchasing model is *not stable* with respect to the L_1-distance $\|\cdot\|_1$. However, the estimate for $|v(\xi) - v(\tilde{\xi})|$ in Theorem 2.1 is valid with $L = 1$ since $D_{\mathrm{f}}(\xi, \tilde{\xi}) = \frac{a}{2}$ holds for the filtration distance (with $r' = \infty$).

Finally, let us consider the case of discrete probability measures P and Q. Let P have scenarios ξ^i with probabilities $p_i > 0$, $i = 1, \ldots, N$, and Q scenarios $\tilde{\xi}^j$ and probabilities $q_j > 0$, $j = 1, \ldots, M$. Clearly, $\sum_{i=1}^N p_i = 1$ and $\sum_{j=1}^M q_j = 1$. Then $\ell_r^r(P, Q)$ is the optimal value of a finite-dimensional linear transportation problem (e.g., [24]) and there exist optimal weights $\eta_{ij} \geq 0$ of the scenario pair $(\xi^i, \tilde{\xi}^j)$, $i = 1, \ldots, N$, $j = 1, \ldots, M$. Hence, there exists a pair $(\xi, \tilde{\xi})$ of random vectors on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = \{\omega_{ij} : i = 1, \ldots, N, j = 1, \ldots, M\}$ and $\mathbb{P}(\omega_{ij}) = \eta_{ij}$, $i = 1, \ldots, N$, $j = 1, \ldots, M$. We define $\xi(\omega_{ij}) = \xi^i$ for every $j = 1, \ldots, M$ and $\tilde{\xi}(\omega_{ij}) = \tilde{\xi}^j$ for every $i = 1, \ldots, N$.

Now, our aim is to study the second term in the stability estimate in Theorem 2.1, namely, the distance of filtrations. Let $\mathcal{F}_t$ and $\tilde{\mathcal{F}}_t$ denote the σ-fields generated by $(\xi_1, \ldots, \xi_t)$ and $(\tilde{\xi}_1, \ldots, \tilde{\xi}_t)$, respectively. Let I_t and $\tilde{I}_t$ denote the index set of realizations of ξ_t and $\tilde{\xi}_t$, respectively. Furthermore, let $\mathcal{E}_t$ and $\tilde{\mathcal{E}}_t$ denote families of nonempty elements of $\mathcal{F}_t$ and $\tilde{\mathcal{F}}_t$, respectively, that form partitions of Ω and generate the corresponding σ-fields. We set $E_{ts} := \{\omega \in \Omega : (\xi_1(\omega), \ldots, \xi_t(\omega)) = (\xi_1^s, \ldots, \xi_t^s)\}$, $s \in I_t$, and $\tilde{E}_{ts} := \{\omega \in \Omega : (\tilde{\xi}_1(\omega), \ldots, \tilde{\xi}_t(\omega)) = (\tilde{\xi}_1^s, \ldots, \tilde{\xi}_t^s)\}$, $s \in \tilde{I}_t$.

We set $r = r' = 1$ and require conditions (A1) and (A2) to hold. Since (2.3) is finite-dimensional in this case, optimal solutions x and $\tilde{x}$ exist and we obtain according to Remark 2.2 that

$$
\begin{aligned}
D_t(\xi, \tilde{\xi}) &= \max\Big\{ \sum_{i,j} \eta_{ij} \|x_t(\omega_{ij}) - \mathbb{E}[x_t|\tilde{\mathcal{F}}_t](\omega_{ij})\|, \\
&\qquad\quad \sum_{i,j} \eta_{ij} \|\tilde{x}_t(\omega_{ij}) - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t](\omega_{ij})\| \Big\} \\
&= \max\Big\{ \sum_{s\in\tilde{I}_t} \sum_{\omega_{ij}\in\tilde{E}_{ts}} \eta_{ij} \Big\| x_t(\omega_{ij}) - \frac{\sum_{\omega_{kl}\in\tilde{E}_{ts}} \eta_{kl} x_t(\omega_{kl})}{\sum_{\omega_{kl}\in\tilde{E}_{ts}} \eta_{kl}} \Big\|, \\
&\qquad\quad \sum_{s\in I_t} \sum_{\omega_{ij}\in E_{ts}} \eta_{ij} \Big\| \tilde{x}_t(\omega_{ij}) - \frac{\sum_{\omega_{kl}\in E_{ts}} \eta_{kl} \tilde{x}_t(\omega_{kl})}{\sum_{\omega_{kl}\in E_{ts}} \eta_{kl}} \Big\| \Big\}.
\end{aligned} \tag{2.19}
$$

The latter representation of D_t has potential to be further estimated in specific cases. In particular, it simplifies considerably for the situation of scenario reduction.

Example 2.7 (scenario reduction). Let us consider the case of deleting scenario $l \in \{1, \ldots, N\}$ of ξ according to the methodology in [5, 11] for the distance ℓ_1 and $r = r' = 1$. Then $\tilde{\xi}$ has the scenarios $\xi^1, \ldots, \xi^{l-1}, \xi^{l+1}, \ldots, \xi^N$ and the probabilities of ξ^j are $q_j = p_j$ for every $j \notin \{j(l), l\}$ and $q_{j(l)} = p_{j(l)} + p_l$, where $j(l) \in \arg\min_{j\neq l} \|\xi^j - \xi^l\|$ (see [5, Theorem 2]). This corresponds to $\tilde{\xi}(\omega_{ij}) = \xi^j$ for every $i = 1, \ldots, N$, $j = 1, \ldots, N$, $j \neq l$. We also infer from [5, Theorem 2] that the optimal weights of the transportation problem defining $\ell_1(P, Q)$ are

$$
\eta_{ij} = \begin{cases} p_l, & i = l,\ j = j(l), \\ p_j, & i = j \neq l, \\ 0 & \text{otherwise.} \end{cases}
$$

We set $\hat{\omega}_j := \omega_{jj}$ for every $j = 1, \ldots, N$, $j \neq l$, $\hat{\omega}_l = \omega_{lj(l)}$ and introduce the notation E_{ts_j} and $\tilde{E}_{ts_j}$ for the sets in $\mathcal{E}_t$ and $\tilde{\mathcal{E}}_t$, respectively, that contain $\hat{\omega}_j$. From (2.19) we conclude the following representations of D_t:

$$
D_t(\xi, \tilde{\xi}) = \max\left\{ \sum_{s\in\tilde{I}_t} \sum_{\hat{\omega}_j\in\tilde{E}_{ts}} p_j \Big\| x_t(\hat{\omega}_j) - \frac{\sum_{\hat{\omega}_k\in\tilde{E}_{ts}} p_k x_t(\hat{\omega}_k)}{\sum_{\hat{\omega}_k\in\tilde{E}_{ts}} p_k} \Big\|, \right.
$$

$$\sum_{s\in I_t}\sum_{\hat\omega_j\in E_{ts}} p_j\Big\|\tilde x_t(\hat\omega_j)-\frac{\sum\limits_{\omega_k\in E_{ts}} p_k\tilde x_t(\hat\omega_k)}{\sum\limits_{\hat\omega_k\in E_{ts}} p_k}\Big\|\Bigg\}$$

$$= \max\Bigg\{\sum_{s\in\tilde I_t}\frac{1}{\sum\limits_{\hat\omega_k\in\tilde E_{ts}} p_k}\sum_{\hat\omega_j\in\tilde E_{ts}}\Big\|\sum_{\hat\omega_k\in\tilde E_{ts}} p_kp_j[x_t(\hat\omega_j)-x_t(\hat\omega_k)]\Big\|,$$

$$\sum_{s\in I_t}\frac{1}{\sum\limits_{\hat\omega_k\in E_{ts}} p_k}\sum_{\hat\omega_j\in E_{ts}}\Big\|\sum_{\hat\omega_k\in E_{ts}} p_kp_j[\tilde x_t(\hat\omega_j)-\tilde x_t(\hat\omega_k)]\Big\|\Bigg\}$$

$$= \max\Bigg\{\sum_{s\in\tilde I_t}\frac{1}{\sum\limits_{\hat\omega_k\in\tilde E_{ts}} p_k}\sum_{\hat\omega_j\in\tilde E_{ts}}\Big\|\sum_{\hat\omega_k\in\tilde E_{ts}\setminus E_{ts_j}} p_kp_j[x_t(\hat\omega_j)-x_t(\hat\omega_k)]\Big\|,$$

$$\sum_{s\in I_t}\frac{1}{\sum\limits_{\hat\omega_k\in E_{ts}} p_k}\sum_{\hat\omega_j\in E_{ts}}\Big\|\sum_{\hat\omega_k\in E_{ts}\setminus\tilde E_{ts_j}} p_kp_j[\tilde x_t(\hat\omega_j)-\tilde x_t(\hat\omega_k)]\Big\|\Bigg\},$$

where the final equality is a consequence of the corresponding measurability properties of x_t, which imply $x_t(\hat\omega_j)=x_t(\hat\omega_k)$ if $\hat\omega_k\in E_{ts}\cap\tilde E_{ts_j}$ and $\hat\omega_k\in\tilde E_{ts}\cap E_{ts_j}$, respectively. Since $E_{ts_j}=\tilde E_{ts_j}$ for $j\notin\{l,j(l)\}$ and $\tilde E_{ts_l}=E_{tj(l)}\cup\{\hat\omega_l\}$, we may continue with

$$D_t(\xi,\tilde\xi)=\max\Bigg\{\frac{1}{\sum\limits_{\hat\omega_k\in\tilde E_{ts_l}} p_k}\sum_{\hat\omega_j\in\tilde E_{ts_l}}\Big\|\sum_{\hat\omega_k\in\tilde E_{ts_l}\setminus E_{ts_j}} p_kp_j[x_t(\hat\omega_j)-x_t(\hat\omega_k)]\Big\|,$$

$$\frac{1}{\sum\limits_{\hat\omega_k\in E_{ts_l}} p_k}\sum_{\hat\omega_j\in E_{ts_l}}\Big\|\sum_{\hat\omega_k\in E_{ts_l}\setminus\tilde E_{ts_j}} p_kp_j[\tilde x_t(\hat\omega_j)-\tilde x_t(\hat\omega_k)]\Big\|\Bigg\}$$

$$=\max\Bigg\{\frac{1}{\sum\limits_{\hat\omega_k\in\tilde E_{ts_l}} p_k}\Bigg\{\sum_{\hat\omega_k\in E_{ts_{j(l)}}}\Big\|p_lp_k[x_t(\hat\omega_k)-x_t(\hat\omega_l)]\Big\|$$

$$+\Big\|\sum_{\hat\omega_k\in E_{ts_{j(l)}}} p_kp_l[\tilde x_t(\hat\omega_l)-\tilde x_t(\hat\omega_k)]\Big\|\Bigg\},$$

$$\frac{1}{\sum\limits_{\hat\omega_k\in E_{ts_l}} p_k}\Bigg\{\sum_{\hat\omega_k\in E_{ts_l}\setminus\{\hat\omega_l\}}\Big\|p_lp_k[x_t(\hat\omega_k)-x_t(\hat\omega_l)]\Big\|$$

$$+\Big\|\sum_{\hat\omega_k\in E_{ts_l}\setminus\{\hat\omega_l\}} p_kp_l[\tilde x_t(\hat\omega_l)-\tilde x_t(\hat\omega_k)]\Big\|\Bigg\}\Bigg\}$$

$$\le\max\Bigg\{\frac{\sum\limits_{\hat\omega_k\in E_{ts_{j(l)}}} 2p_lp_k\|x_t(\hat\omega_k)-x_t(\hat\omega_l)\|}{p_l+\sum\limits_{\hat\omega_k\in E_{ts_{j(l)}}} p_k},\frac{\sum\limits_{\hat\omega_k\in E_{ts_l}\setminus\{\hat\omega_l\}} 2p_lp_k\|\tilde x_t(\hat\omega_k)-\tilde x_t(\hat\omega_l)\|}{p_l+\sum\limits_{\hat\omega_k\in E_{ts_l}\setminus\{\hat\omega_l\}} p_k}\Bigg\}$$

$$\le 2p_l\max\Big\{\|x_t(\hat\omega_{j(l)})-x_t(\hat\omega_l)\|,\min_{\hat\omega_k\in E_{ts_l}\setminus\{\hat\omega_l\}}\|\tilde x_t(\hat\omega_k)-\tilde x_t(\hat\omega_l)\|\Big\}, \tag{2.20}$$

where the convention is used that $\min_{\hat\omega_k\in E_{ts_l}\setminus\{\hat\omega_l\}}=0$ if $E_{ts_l}\setminus\{\hat\omega_l\}=\emptyset$. The final estimate makes use of the fact that all $x_t(\hat{\omega}_k)$ with $\hat{\omega}_k \in E_{ts_{j(l)}}$ and $\hat{\omega}_k \in E_{ts_l} \setminus \{\hat{\omega}_l\}$, respectively, coincide.

In the following two cases, the above estimate simplifies to

$$D_t(\xi, \tilde{\xi}) \leq \begin{cases} 0 & \text{if } \hat{\omega}_l \in E_{ts_{j(l)}}, \\ 2p_l \|x_t(\hat{\omega}_{j(l)}) - x_t(\hat{\omega}_l)\| & \text{if } E_{ts_l} = \{\hat{\omega}_l\}. \end{cases}$$

As the sets $l_0(F(\xi, \cdot))$ and $l_0(F(\tilde{\xi}, \cdot))$ of solutions of the original and perturbed multistage models are bounded in $L_{r'}$ due to (A2), there exists a constant $K > 0$ such that

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) \leq K p_l.$$

Hence, if the probability p_l of the deleted scenario is small, the filtration distance is also small. Then there is no need to modify the deletion procedure based on best approximations with respect to the metric ℓ_1. This is mostly the case if the tree is bushy, i.e., contains many scenarios.

A more reliable estimate for the filtration distance may be obtained by solving the stochastic program for an approximation $\hat{\xi}$ of ξ (on $\{\hat{\omega}_1, \ldots, \hat{\omega}_N\}$), which contains much less scenarios than ξ. Then an estimate for the filtration distance may be obtained by computing

$$2p_l \sum_{t=2}^{T-1} \max \left\{ \|\hat{x}_t(\hat{\omega}_{j(l)}) - \hat{x}_t(\hat{\omega}_l)\|, \min_{\hat{\omega}_k \in E_{ts_l} \setminus \{\hat{\omega}_l\}} \|\hat{x}_t(\hat{\omega}_k) - \hat{x}_t(\hat{\omega}_l)\| \right\},$$

where $\hat{x} \in l_0(F(\hat{\xi}, \cdot))$ is the corresponding solution. Altogether, some scenario deletion suggested by the strategy in [5, 11] can either be carried out if the bound (2.20) on the filtration distance remains small or is rejected.

3. Conclusions. While quantitative stability results for two-stage stochastic programs have to take into account only a suitable distance of probability distributions, this is no longer the case for multistage models, where the filtration distance enters stability estimates. This fact demonstrates the importance of the conditional structure of multistage stochastic programs. This is in line with the observations and results of [32]. In a sense, it also seems to illustrate the complexity results obtained in the recent paper [33]. It is shown there that multistage stochastic programs have higher complexity than two-stage models. Techniques for generating and reducing *scenario trees* in multistage stochastic programs, which are based on stability arguments, have to respect *both* probability *and* filtration distances as both contribute to changes of optimal values. Example 2.7 provides upper bounds for the filtration distance if some scenario is deleted. Bounding the filtration distance is also possible for the forward and backward scenario tree generation algorithms developed in [10] and [12]. Such bounds are derived and discussed in the companion paper [13].

Acknowledgments. The first two authors wish to thank the members of the OSIRIS Division at R&D of EDF for several stimulating discussions on scenario trees and stability. We extend our gratitude to René Henrion (WIAS Berlin) for his comments on an earlier version of this paper and to two anonymous referees for their insightful comments.

REFERENCES

[1] K. Barty, *Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique*, Thèse de Doctorat, École Nationale des Ponts et Chaussées, Paris, 2004.

[2] E. S. Boylan, *Equiconvergence of martingales*, Ann. Math. Statist., 42 (1971), pp. 552–559.

[3] K. D. Cotter, *Convergence of information, random variables and noise*, J. Math. Econom., 16 (1987), pp. 39–51.

[4] J. Dupačová, G. Consigli, and S. W. Wallace, *Scenarios for multistage stochastic programs*, Ann. Oper. Res., 100 (2000), pp. 25–53.

[5] J. Dupačová, N. Gröwe-Kuska, and W. Römisch, *Scenario reduction in stochastic programming: An approach using probability metrics*, Math. Program., 95 (2003), pp. 493–511.

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

[7] O. Fiedler and W. Römisch, *Stability in multistage stochastic programming*, Ann. Oper. Res., 56 (1995), pp. 79–93.

[8] R. Fortet and E. Mourier, *Convergence de la répartition empirique vers la répartition théorique*, Ann. Sci. Ecole Norm. Sup. (3), 70 (1953), pp. 267–285.

[9] C. R. Givens and R. M. Shortt, *A class of Wasserstein metrics for probability distributions*, Michigan Math. J., 31 (1984), pp. 231–240.

[10] N. Gröwe-Kuska, H. Heitsch, and W. Römisch, *Scenario reduction and scenario tree construction for power management problems*, A. Borghetti, C.A. Nucci, and M. Paolone eds., IEEE Bologna Power Tech Proceedings, Bologna, Italy, 2003.

[11] H. Heitsch and W. Römisch, *Scenario reduction algorithms in stochastic programming*, Comput. Optim. Appl., 24 (2003), pp. 187–206.

[12] H. Heitsch and W. Römisch, *Generation of multivariate scenario trees to model stochasticity in power management*, IEEE St. Petersburg Power Tech Proceedings, St. Petersburg, Russia, 2005.

[13] H. Heitsch and W. Römisch, *Scenario tree modelling for multistage stochastic programs*, preprint 296, DFG Research Center Matheon (Mathematics for key technologies), 2005, Berlin, Germany, (www.matheon.de).

[14] R. Hochreiter and G. Ch. Pflug, *Financial scenario generation for stochastic multi-stage decision processes as facility location problem*, Ann. Oper. Res., to appear.

[15] D. N. Hoover, *Convergence in distribution and Skorokhod convergence for the general theory of processes*, Probab. Theory Related Fields, 89 (1991), pp. 239–259.

[16] V. Kaňková, *Empirical estimates in multistage stochastic programs*, Report No. 1930, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, 1998.

[17] J. Neveu, *Note on the tightness of the metric on the set of complete sub σ-algebras of a probability space*, Ann. Math. Statist., 43 (1972), pp. 1369–1371.

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

[19] P. Olsen, *Discretizations of multistage stochastic programming problems*, Math. Programming Stud., 6 (1976), pp. 111–124.

[20] T. Pennanen, *Epi-convergent discretizations of multistage stochastic programs via integration quadratures*, Stochastic Programming E-Print Series 19–2004 (www.speps.org) and Math. Program., Ser. B, to appear.

[21] G. Ch. Pflug, *Scenario tree generation for multiperiod financial optimization by optimal discretization*, Math. Program., 89 (2001), pp. 251–271.

[22] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, Wiley and Sons, Chichester, UK, 1991.

[23] S. T. Rachev and W. Römisch, *Quantitative stability in stochastic programming: The method of probability metrics*, Math. Oper. Res., 27 (2002), pp. 792–818.

[24] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems*, Vol. I and II, Springer, Berlin, 1998.

[25] S. T. Rachev and A. Schief, *On L_p-minimal metrics*, Probab. Math. Statist., 13 (1992), pp. 311–320.

[26] R. T. Rockafellar and R. J-B Wets, *Nonanticipativity and $\mathcal{L}^1$-martingales in stochastic optimization problems*, Math. Programming Stud., 6 (1976), pp. 170–187.

[27] R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[28] W. Römisch, *Stability of stochastic programming problems*, in Stochastic Programming (A. Ruszczyński and A. Shapiro Eds.), Handbooks in Operations Research and Management

Science, 10, Elsevier, Amsterdam, 2003, pp. 483–554.

[29] W. Römisch and R. Schultz, *Stability analysis for stochastic programs*, Ann. Oper. Res., 30 (1991), pp. 241–266.

[30] L. Rogge, *Uniform inequalities for conditional expectations*, Ann. Probab., 2 (1974), pp. 486–489.

[31] A. Ruszczyński and A. Shapiro, eds., *Stochastic Programming*, Handbooks in Operations Research and Management Science, 10, Elsevier, Amsterdam, 2003.

[32] A. Shapiro, *Inference of statistical bounds for multistage stochastic programming problems*, Math. Methods Oper. Res., 58 (2003), pp. 57–68.

[33] A. Shapiro and A. Nemirovski, *On complexity of stochastic programming problems*, in Continuous Optimization: Current Trends and Applications, V. Jeyakumar and A. M. Rubinov eds., Springer, New York, 2005, pp. 111–144.

[34] C. Strugarek, *On the Fortet-Mourier metric for the stability of stochastic programming problems*, Stochastic Programming E-Print Series 25-2004 (www.speps.org).