Probability Objectives in Stochastic Programs with Recourse

Rüdiger Schultz
Institute of Mathematics, Gerhard-Mercator University Duisburg
Lotharstr. 65, D-47048 Duisburg, Germany

Abstract

Traditional models in multistage stochastic programming are directed to minimizing the expected value of random optimal costs arising in a multistage, non-anticipative decision process under uncertainty. Motivated by risk aversion, we consider minimization of the probability that the random optimal costs exceed some preselected threshold value. For the two-stage case, we analyse structural properties and propose algorithms both for models with integer decisions and for those without. Extension of the modeling to the multistage situation concludes the paper.

Key Words. Stochastic programming, mixed-integer optimization.

AMS subject classifications. 90C15, 90C11, 90C06.

1 Introduction

Stochastic programs with recourse arise as deterministic equivalents to random optimization problems. In the present paper the main accent will be placed at the two-stage situation, and the most general random optimization problems to be considered are random mixed-integer linear programs. These are accompanied by a two-stage scheme of alternating decision and observation. After having decided on parts of the variables in a first stage, the random data infecting the problem are observed, and in turn the remaining (second-stage or recourse) variables are fixed. In our present analysis two basic assumptions underly this scheme. First, and naturally, the first-stage decision has to be taken on a "here-and-now" basis, i.e., it must not depend on (or anticipate) the outcome of the random data. Secondly, and providing some modeling restriction, the first-stage decision does not influence the probability distribution of the random data.

In multistage stochastic programs the above two-stage scheme is extended into a finite horizon sequential decision process under uncertainty. Again we have to maintain nonanticipativity of decisions, and, so far, almost all results concern problems where the decisions do not influence the probability distribution of the random data. In the final section of the present paper we will return to multistage stochastic programs.

After having sketched the rules for how to make decisions, let us now discuss criteria for how to select a "best" decision. In this respect, the existing literature on stochastic programs with recourse (cf. the textbooks [5, 15, 20] and the references therein) almost unanimously suggests to start out from expectations of objective function values of the random optimization problem. For two-stage models (in a cost minimization framework) this implies that the deterministic first-stage decision is selected such that the expectation of the sum of the deterministic first-stage costs and the random second-stage costs (induced by the random data and an optimal second-stage decision) becomes minimal. Such a criterion has proven useful in many applications. In case the random optimzation problem is a linear program without integer requirements, the resulting stochastic program with recourse enjoys convexity in the first-stage variables. This enabled application of powerful tools from convex analysis, both for structural investigations and algorithm design (cf. [4, 5, 15, 20, 32]).

In the present paper, we will discuss recourse stochastic programs where the optimization is based on minimizing the probability that the above sum of deterministic and random costs exceeds a given threshold value. Such models provide an opportunity to address risk aversion in the framework of recourse stochastic programming.

The proposal to replace the usual expectation-based objective function in recourse stochastic programming by a probability objective seemingly dates back to Bereanu [2] and, hitherto, has not been elaborated in much detail. Reformulating the stochastic program by adding another variable and including level sets of the objective into the constraints leads to a chance constrained stochastic program which is nonconvex in general. We will see that, along this line, some structural knowledge on chance constraints (cf. [5, 15, 16, 20, 29]) reappears in the structural analysis of our models. Algorithmically, we will view several well-established techniques from a fresh perspective. Among them there are cutting planes from convex subgradient optimization, Lagrangian relaxation of mixed-integer programs, and decomposition techniques for block-angular stochastic programs.
The paper is organized as follows. In Section 2 we formalize the modeling outlined above, collect some prerequisites, and compare with the usual expectation-based modeling in recourse stochastic programming. Section 3 is devoted to structural results. In Section 4 we present some first algorithmic approaches. Separate attention is paid to models without integer decisions since they allow for an algorithmic shortcut. As already announced, the final section will discuss the extension of our modeling to multistage stochastic programs.

2 Modeling

Consider the following random mixed-integer linear program

$$\min_{x,y,y'}\{c^Tx+q^Ty+q'^Ty' \;:\; Tx+Wy+W'y' \;=\; h(\omega),\; x\in X,\; y\in \mathbb{Z}_+^{\bar{m}},\; y'\in \mathbb{R}_+^{m'}\}. \tag{1}$$

We assume that all ingredients above have conformal dimensions, that W, W' are rational matrices, and that $X \subseteq \mathbb{R}^m$ is a nonempty closed polyhedron. Integer requirements to components of x are formally possible but will not be imposed for ease of exposition. For the same reason, randomness is kept as simple as possible by claiming that only the right-hand side $h(\omega) \in \mathbb{R}^s$ is random, i.e., a random vector on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Decision variables are divided into two groups: first-stage variables x to be fixed before and second-stage variables (y, y') to be fixed after observation of $h(\omega)$.
Let us denote

$$\Phi(t) := \min\{q^Ty+q'^Ty' \;:\; Wy+W'y'=t,\; y\in Z_+^{\bar{m}},\; y'\in \mathbb{R}_+^{m'}\}. \tag{2}$$

According to integer programming theory([19]), this function is real-valued on $\mathbb{R}^s$ provided that $W(\mathbb{Z}_+^{\bar{m}})+W'(\mathbb{R}_+^{m'})=\mathbb{R}^s$ and $\{u\in\mathbb{R}^s \;:\; W^Tu\le q,\; {W'}^Tu\le q'\}\neq\emptyset$ which, therefore, will be assumed throughout.
The classical expectation-based stochastic program with recourse now is the optimization problem

$$\min\Big\{\int_\Omega (c^Tx+\Phi(h(\omega)-Tx))\,\mathbb{P}(d\omega) \;:\; x\in X\Big\}. \tag{3}$$

The recourse stochastic program with probability objective reads

$$\min\Big\{\mathbb{P}\big(\{\omega\in\Omega \;:\; c^Tx+\Phi(h(\omega)-Tx)>\varphi_o\}\big) \;:\; x\in X\Big\} \tag{4}$$

where $\varphi_o \in \mathbb{R}$ denotes some preselected threshold (some ruin level in a cost framework, for instance). For convenience, we will call (3) the expectation-based and (4) the probability-based recourse model. In doing so, we are well aware of the fact that, of course, (4) is expectation-based too, if probabilities are understood as expectations of indicator functions.
We will see in a moment, that both (3) and (4) are well-defined nonlinear optimization problems. Their objective functions are denoted by $Q_{\mathbb{E}}(x)$ and $Q_{\mathbb{P}}(x)$, respectively. To detect their structure, the function Φ is crucial, which arises as a value function of a mixed-integer linear program. From parametric optimization ([1, 6]) the following is known

Proposition 2.1 *Assume that $W(\mathbb{Z}_+^{\bar{m}})+W'(\mathbb{R}_+^{m'})=\mathbb{R}^s$ and $\{u\in\mathbb{R}^s \;:\; W^Tu\le q,\; {W'}^Tu\le q'\}\neq\emptyset$. Then it holds*

(i) Φ is real-valued and lower semicontinuous on $\mathbb{R}^s$,

(ii) there exists a countable partition $\mathbb{R}^s = \cup_{i=1}^\infty \mathcal{T}_i$ such that the restrictions of Φ to $\mathcal{T}_i$ are piecewise linear and Lipschitz continuous with a uniform constant $L>0$ not depending on i,

(iii) *each of the sets* $\mathcal{T}_i$ *has a representation* $\mathcal{T}_i = \{t_i + \mathcal{K}\} \setminus \cup_{j=1}^{N} \{t_{ij} + \mathcal{K}\}$ *where* $\mathcal{K}$ *denotes the polyhedral cone* $W'(I\!R_+^{m'})$ *and* t_i, t_{ij} *are suitable points from* $I\!R^s$*, moreover,* N *does not depend on* i*,*

(iv) *there exist positive constants* β, γ *such that* $|\Phi(t_1) - \Phi(t_2)| \leq \beta \|t_1 - t_2\| + \gamma$ *whenever* $t_1, t_2 \in I\!R^s$*.*

In case $\bar{m} = 0$, i.e., if there are no integer requirements in the second stage, Φ becomes the value function of a linear program. Under the assumptions of Proposition 2.1, Φ is real-valued on $I\!R^s$. By linear programming duality it is convex, piecewise linear, and adopts a representation

$$\Phi(t) = \max_{j=1,\ldots,J} d_j^T t$$

where $d_1, \ldots, d_J$ are the vertices of $\{u \in I\!R^s \ : \ W'^T u \leq q'\}$, which is a compact set in this case.
As an immediate conclusion we obtain, that, without integer requirements in the second stage, $1 - Q_{I\!P}(x)$ coincides with the probability of a closed polyhedron, providing a direct link to chance constrained stochastic programming ([5, 15, 20]).
Before we will turn our attention to $Q_{I\!P}(x)$, we review some properties of $Q_{I\!E}(x)$. For convenience we denote by μ the image measure $I\!P \circ h^{-1}$ on $I\!R^s$. Without integer requirements ($\bar{m} = 0$), convexity of Φ extends to $Q_{I\!E}$ under mild conditions. A standard result of stochastic linear programming reads

Proposition 2.2 *Assume* $\bar{m} = 0$, $W'(I\!R_+^{m'}) = I\!R^s$, $\{u \in I\!R^s \ : \ W'^T u \leq q'\} \neq \emptyset$, *and* $\int_{I\!R^s} \|h\| \, \mu(dh) < \infty$. *Then* $Q_{I\!E} \, : \, I\!R^m \longrightarrow I\!R$ *is a real-valued convex function.*

As already mentioned in the introduction, convexity has been exploited extensively in stochastic linear programming. For further reading we refer to the textbooks [5, 15, 20]. The remaining models, both expectation- and probability-based, to be discussed in the present paper enjoy convexity merely in exceptional situations. Straightforward examples (cf. e.g. [35]) confirm that convexity in (3) is lost already for very simple models as soon as integer requirements enter the second stage. In [33] the following is shown.

Proposition 2.3 *Assume that* $W(Z\!\!Z_+^{\bar{m}}) + W'(I\!R_+^{m'}) = I\!R^s$, $\{u \in I\!R^s \ : \ W^T u \leq q, \ W'^T u \leq q'\} \neq \emptyset$, *and* $\int_{I\!R^s} \|h\| \, \mu(dh) < \infty$. *Then it holds*

(i) $Q_{I\!E} \, : \, I\!R^m \longrightarrow I\!R$ *is a real-valued lower semicontinuous function,*

(ii) *if* μ *has a density, then* $Q_{I\!E}$ *is continuous on* $I\!R^m$*.*

3 Structure

To analyse the structure of $Q_{I\!P}$ we introduce the notation

$$M(x) \ := \ \{h \in I\!R^s \, : \, c^T x + \Phi(h - Tx) > \varphi_o\}, \ \ x \in I\!R^m.$$

By $\liminf_{x_n \to x} M(x_n)$ and $\limsup_{x_n \to x} M(x_n)$ we denote the (set theoretic) limes inferior and limes superior, i.e., the set of all points belonging to all but a finite number of the sets $M(x_n)$, $n \in I\!N$, and to infinitely many of the sets $M(x_n)$, respectively. Moreover, we denote

$$\begin{array}{rcl} M_e(x) & := & \{h \in I\!R^s \ : \ c^T x + \Phi(h - Tx) = \varphi_o\}, \\ M_d(x) & := & \{h \in I\!R^s \ : \ \Phi \text{ is discontinuous at } h - Tx\}. \end{array}$$

Note that, by Proposition 2.1, both $M_e(x)$ and $M_d(x)$ are measurable sets for all $x \in I\!R^m$.

Lemma 3.1 *For all* $x \in I\!R^m$ *there holds*

$$M(x) \ \subseteq \ \liminf_{x_n \to x} M(x_n) \ \subseteq \ \limsup_{x_n \to x} M(x_n) \ \subseteq \ M(x) \, \cup \, M_e(x) \, \cup \, M_d(x).$$

Proof: Let $h \in M(x)$. The lower semicontinuity of Φ (Proposition 2.1) yields

$$\liminf_{x_n \to x} (c^T x_n + \Phi(h - Tx_n)) \;\geq\; c^T x + \Phi(h - Tx) \;>\; \varphi_o.$$

Therefore, there exists an $n_o \in I\!N$ such that $c^T x_n + \Phi(h - Tx_n) > \varphi_o$ for all $n \geq n_o$, implying $h \in M(x_n)$ for all $n \geq n_o$. Hence, $M(x) \subseteq \liminf_{x_n \to x} M(x_n)$.
Let $h \in \limsup_{x_n \to x} M(x_n) \setminus M(x)$. Then there exists an infinite subset $\tilde{I\!N}$ of $I\!N$ such that

$$c^T x_n + \Phi(h - Tx_n) > \varphi_o \;\forall n \in \tilde{I\!N} \quad \text{and} \quad c^T x + \Phi(h - Tx) \leq \varphi_o.$$

Now two cases are posssible. First, Φ is continuous at $h - Tx$. Passing to the limit in the first inequality then yields that $c^T x + \Phi(h - Tx) \geq \varphi_o$, and $h \in M_e(x)$. Secondly, Φ is discontinuous at $h - Tx$. In other words, $h \in M_d(x)$. $\square$

Proposition 3.2 *Assume that* $W(Z\!\!Z_+^{\bar{m}}) + W'(I\!R_+^{m'}) = I\!R^s$ *and* $\{u \in I\!R^s \;:\; W^T u \leq q,\; W'^T u \leq q'\} \neq \emptyset$. *Then* $Q_{I\!P} : I\!R^m \longrightarrow I\!R$ *is a real-valued lower semicontinuous function.*
If in addition $\mu\big(M_e(x) \cup M_d(x)\big) = 0$, *then* $Q_{I\!P}$ *is continuous at* x.

Proof: The lower semicontinuity of Φ ensures that $M(x)$ is measurable for all $x \in I\!R^m$, and hence $Q_{I\!P}$ is real-valued on $I\!R^m$. By Lemma 3.1 and the (semi-) continuity of the probability measure on sequences of sets we have for all $x \in I\!R^m$

$$Q_{I\!P}(x) \;=\; \mu\big(M(x)\big) \;\leq\; \mu\big(\liminf_{x_n \to x} M(x_n)\big) \;\leq\; \liminf_{x_n \to x} \mu\big(M(x_n)\big) \;=\; \liminf_{x_n \to x} Q_{I\!P}(x_n),$$

establishing the asserted lower semicontinuity. In case $\mu\big(M_e(x) \cup M_d(x)\big) = 0$ this argument extends as follows

$$\begin{aligned} Q_{I\!P}(x) \;&=\; \mu\big(M(x)\big) \;=\; \mu\big(M(x) \cup M_e(x) \cup M_d(x)\big) \;\geq\; \mu\big(\limsup_{x_n \to x} M(x_n)\big) \\ &\geq\; \limsup_{x_n \to x} \mu\big(M(x_n)\big) \;=\; \limsup_{x_n \to x} Q_{I\!P}(x_n), \end{aligned}$$

and $Q_{I\!P}$ is continuous at x. $\square$

Proposition 2.1 now reveals that, for given $x \in I\!R^m$, both $M_e(x)$ and $M_d(x)$ are contained in a countable union of hyperplanes. The latter being of Lebesgue measure zero we obtain that $\mu\big(M_e(x) \cup M_d(x)\big) = 0$ is valid for all $x \in I\!R^m$ provided that μ has a density. This proves

Conclusion 3.3 *Assume that* $W(Z\!\!Z_+^{\bar{m}}) + W'(I\!R_+^{m'}) = I\!R^s$, $\{u \in I\!R^s \;:\; W^T u \leq q,\; W'^T u \leq q'\} \neq \emptyset$, *and that* μ *has a density. Then* $Q_{I\!P}$ *is continuous on* $I\!R^m$.

This analysis can be extended towards Lipschitz continuity of $Q_{I\!P}$. In [36], Tiedemann has shown

Proposition 3.4 *Assume that* q, q' *are rational vectors,* $W(Z\!\!Z_+^{\bar{m}}) + W'(I\!R_+^{m'}) = I\!R^s$, $\{u \in I\!R^s \;:\; W^T u \leq q,\; W'^T u \leq q'\} \neq \emptyset$, *and that for any nonsingular linear transformation* $B \in L(I\!R^s, I\!R^s)$ *all one-dimensional marginal distributions of* $\mu \circ B$ *have bounded densities which, outside some bounded interval, are monotonically decreasing with growing absolute value of the argument. Then* $Q_{I\!P}$ *is Lipschitz continuous on any bounded subset of* $I\!R^m$.

From numerical viewpoint, the optimization problems (3) and (4) pose the major difficulty that their objective functions are given by multidimensional integrals with implicit integrands. If $h(\omega)$ follows a continuous probability distribution the computation of $Q_{I\!E}$ and $Q_{I\!P}$ has to rely on approximations. Here, it is quite common to approximate the probability distribution of $h(\omega)$ by discrete distributions, turning the integrals in (3), (4) into sums this way. In the next section we will see that discrete distributions, despite the poor analytical properties they imply for $Q_{I\!E}$ and $Q_{I\!P}$, are quite attractive algorithmically, since they allow for integer programming techniques.
Approximating the underlying probability measures in (3) and (4) raises the question whether "small" perturbations in the measures result in only "small" perturbations of optimal values and optimal solutions. Subjective assumptions and incomplete knowledge on $\mu = I\!P \circ h^{-1}$ in many practical modeling situations

provide further motivation for asking this question. Therefore, stability analysis has gained some interest in stochastic programming (for surveys see [9, 35]).

For the models (3) and (4) qualitative and quantitative continuity of $Q_{\mathbb{E}}$, $Q_{\mathbb{P}}$ jointly in the decision variable x and the probability measure μ becomes a key issue then. Once established, the continuity, together with well-known techniques from parametric optimization, lead to stability in the spirit sketched above. In the present paper, we will not pursue stability analysis, but show how to arrive at qualitative joint continuity of $Q_{\mathbb{P}}$. For continuity results on $Q_{\mathbb{E}}$ we refer to [14, 24, 30, 33, 34], for extensions towards stability to [35] and the references therein.

For the rest of this section, we consider $Q_{\mathbb{P}}$ as a function mapping from $\mathbb{R}^m \times \mathcal{P}(\mathbb{R}^s)$ to $\mathbb{R}$. By $\mathcal{P}(\mathbb{R}^s)$ we denote the set of all Borel probability measures on $\mathbb{R}^s$. While $\mathbb{R}^s$ is equipped with the usual topology, the set $\mathcal{P}(\mathbb{R}^s)$ is endowed with weak convergence of probability measures. This has proven both sufficiently general to cover relevant applications and sufficiently specific to enable substantial statements. A sequence $\{\mu_n\}$ in $\mathcal{P}(\mathbb{R}^s)$ is said to converge weakly to $\mu \in \mathcal{P}(\mathbb{R}^s)$, written $\mu_n \xrightarrow{w} \mu$, if for any bounded continuous function $g : \mathbb{R}^s \to \mathbb{R}$ we have

$$\int_{\mathbb{R}^s} g(\xi)\mu_n(d\xi) \to \int_{\mathbb{R}^s} g(\xi)\mu(d\xi) \quad \text{as} \quad n \to \infty. \tag{5}$$

A basic reference for weak convergence of probability measures is Billingsley's book [3].

Proposition 3.5 *Assume that* $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ *and* $\{u \in \mathbb{R}^s \ : \ W^T u \le q, \ W'^T u \le q'\} \neq \emptyset$. *Let* $\mu \in \mathcal{P}(\mathbb{R}^s)$ *be such that* $\mu\big(M_e(x) \cup M_d(x)\big) = 0$. *Then* $Q_{\mathbb{P}} \ : \ \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^s) \longrightarrow \mathbb{R}$ *is continuous at* (x, μ).

Proof: Let $x_n \longrightarrow x$ and $\mu_n \xrightarrow{w} \mu$ be arbitrary sequences. By $\chi_n, \chi \ : \ \mathbb{R}^s \longrightarrow \{0,1\}$ we denote the indicator functions of the sets $M(x_n), M(x), n \in \mathbb{N}$. In addition, we introduce the exceptional set

$$E \ := \ \{h \in \mathbb{R}^s \ : \ \exists h_n \to h \text{ such that } \chi_n(h_n) \not\to \chi(h)\}.$$

Now we have $E \ \subseteq \ M_e(x) \cup M_d(x)$. To see this, assume that $h \in \big(M_e(x) \cup M_d(x)\big)^c = \big(M_e(x)\big)^c \cap \big(M_d(x)\big)^c$ where the superscript c denotes the set-theoretic complement. Then Φ is continuous at $h - Tx$, and either $c^T x + \Phi(h - Tx) > \varphi_o$ or $c^T x + \Phi(h - Tx) < \varphi_o$. Thus, for any sequence $h_n \to h$ there exists an $n_o \in \mathbb{N}$ such that for all $n \ge n_o$ either $c^T x_n + \Phi(h_n - Tx_n) > \varphi_o$ or $c^T x_n + \Phi(h_n - Tx_n) < \varphi_o$. Hence, $\chi_n(h_n) \to \chi(h)$ as $h_n \to h$, implying $h \in E^c$.

In view of $E \ \subseteq \ M_e(x) \cup M_d(x)$ and $\mu\big(M_e(x) \cup M_d(x)\big) = 0$ we obtain that $\mu(E) = 0$. A theorem on weak convergence of image measures attributed to Rubin in [3], p. 34, now yields that the weak convergence $\mu_n \xrightarrow{w} \mu$ implies the weak convergence $\mu_n \circ \chi_n^{-1} \xrightarrow{w} \mu \circ \chi^{-1}$.

Note that $\mu_n \circ \chi_n^{-1}, \mu \circ \chi^{-1}, n \in \mathbb{N}$ are probability measures on $\{0,1\}$. Their weak convergence then particularly implies that

$$\mu_n \circ \chi_n^{-1}\big(\{1\}\big) \ \longrightarrow \ \mu \circ \chi^{-1}\big(\{1\}\big).$$

In other words, $\mu_n\big(M(x_n)\big) \longrightarrow \mu\big(M(x)\big)$ or $Q_{\mathbb{P}}(x_n, \mu_n) \longrightarrow Q_{\mathbb{P}}(x, \mu)$. □

As done for the expectation-based model (3) in [33], continuity of optimal values and upper semicontinuity of optimal solution sets of the probability-based model (4) can be derived from Proposition 3.5.

Remark 3.6 *(probability-based model without integer decisions)*
Without integer second-stage variables the set $M_d(x)$ *is always empty, and Propositions 3.2 and 3.5 readily specify. A direct approach to these models including stability analysis and algorithmic techniques has been carried out in [23]. Lower semicontinuity of* $Q_{\mathbb{P}}$ *in the absence of integer variables can already be derived from Proposition 3.1 in [29], a statement concerning chance constrained stochastic programs. Some early work on continuity properties of general probability functionals has been done by Raik ([21, 22], see also [16, 20]).*

4 Algorithms

In the present section we will review two algorithms for solving the probability-based recourse problem (4) provided the underlying measure μ is discrete, say with realizations h_j and probabilities $\pi_j, j = 1, \dots, J$. The algorithms were first proposed in [23] and [36], respectively, where further details can be found.

4.1 Linear Recourse

We assume that there are no integer requirements to second-stage variables which is usually referred to as linear recourse in the literature. Suppose that μ is the above discrete measure and consider problem (4) with

$$\Phi(t) := \min\{q^T y \;:\; Wy \geq t,\; y \in I\!R^{m'}_+\}. \tag{6}$$

For ease of exposition let $X \subseteq I\!R^m$ be a nonempty compact polyhedron. Let $e \in I\!R^s$ denote the vector of all ones and consider the set

$$D \;:=\; \{(u, u_o) \in I\!R^{s+1} \;:\; 0 \leq u \leq e,\; 0 \leq u_o \leq 1,\; W^T u - u_o q \leq 0\}$$

together with its extreme points (d_k, d_{ko}), $k = 1, \ldots, K$. Furthermore, consider the indicator function

$$\chi(x, h) := \begin{cases} 1 & , \quad h \in M(x) \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{7}$$

The key idea of the subsequent algorithm is to represent χ by a binary variable and a number of optimality cuts which enables exploitation of cutting plane techniques from convex subgradient optimization. The latter have proven very useful in classical two-stage linear stochastic programming, see e.g. [4, 32].

Lemma 4.1 *There exists a sufficiently large constant $M_o > 0$ such that problem (4) can be equivalently restated as*

$$\min_{x,\theta}\{\sum_{j=1}^{J} \pi_j \theta_j \quad : \quad (h_j - Tx)^T d_k \;+\; (c^T x - \varphi_o) d_{ko} \;\leq\; M_o \theta_j, \quad x \in X,\; \theta_j \in \{0, 1\},\; k = 1, \ldots, K,\; j = 1, \ldots, J\}. \tag{8}$$

Proof: For any $x \in X$ and any $j \in \{1, \ldots, J\}$ consider the feasibility problem

$$\min\{e^T t + t_o \;:\; Wy + t \geq h_j - Tx,\; q^T y - t_o \leq \varphi_o - c^T x,\; y \in I\!R^{m'}_+,\; (t, t_o) \in I\!R^{s+1}_+\} \tag{9}$$

and its linear programming dual

$$\max\{(h_j - Tx)^T u + (c^T x - \varphi_o) u_o \;:\; 0 \leq u \leq e,\; 0 \leq u_o \leq 1,\; W^T u - u_o q \leq 0\}.$$

Clearly, both programs are always solvable. Their optimal value is equal to zero, if and only if $\chi(x, h_j) = 0$. In addition, D coincides with the feasible set of the dual. If M_o is selected as

$$M_o \;:=\; \max\{(h_j - Tx)^T d_k + (c^T x - \varphi_o) d_{ko} \;:\; x \in X,\; k \in \{1, \ldots, K\}, j \in \{1, \ldots, J\}\},$$

then, for any $x \in X$, the vector (x, θ) with $\theta_j = 1, j = 1, \ldots, J$ is feasible for (8).
If $\chi(x, h_j) = 1$ for some $x \in X$ and $j \in \{1, \ldots, J\}$, then there has to exist some $k \in \{1, \ldots, K\}$ such that

$$(h_j - Tx)^T d_k + (c^T x - \varphi_o) d_{ko} > 0.$$

Hence, given $x \in X$, $\theta_j = 0$ is feasible in (8) if and only if $\chi(x, h_j) = 0$. Therefore, (8) is equivalent to $\min\{\sum_{j=1}^{J} \pi_j \chi(x, h_j) \;:\; x \in X\}$. □

The algorithm progresses by sequentially solving a master problem and adding violated optimality cuts generated through the solution of subproblems (9). These cuts correspond to constraints in (8). Assuming that the cuts generated before iteration ν correspond to subsets $\mathcal{K}_\nu \subseteq \{1, \ldots, K\}$ the current master problem reads

$$\min_{x,\theta}\{\sum_{j=1}^{J} \pi_j \theta_j \quad : \quad (h_j - Tx)^T d_k \;+\; (c^T x - \varphi_o) d_{ko} \;\leq\; M_o \theta_j, \quad x \in X,\; \theta_j \in \{0, 1\},\; k \in \mathcal{K}_\nu,\; j = 1, \ldots, J\}. \tag{10}$$

The full algorithm proceeds as follows.

Algorithm 4.2
Step 1 *(Initialization): Set* $\nu = 0$ *and* $\mathcal{K}_o = \emptyset$.
Step 2 *(Solving the master problem): Solve the current master problem (10) and let* (x^ν, θ^ν) *be an optimal solution.*
Step 3 *(Solving subproblems): Solve the feasibility problem (9) for* $x = x^\nu$ *and all* $j \in \{1, \ldots, J\}$ *such that* $\theta_j^\nu = 0$. *Consider the following situations:*

1. *If all these problems have optimal value equal to zero, then the current* x^ν *is optimal for (8).*

2. *If some of these problems have optimal value strictly greater than zero, then, via the dual solutions, a subset* $(d_k, d_{ko}), k \in \tilde{\mathcal{K}} \subseteq \{1, \ldots, K\}$ *of extreme points of* D *is identified. The corresponding cuts are added to the master.*
   Set $\mathcal{K}_{\nu+1} := \mathcal{K}_\nu \cup \tilde{\mathcal{K}}$ *and* $\nu := \nu + 1$*; go to Step 2.*

The algorithm terminates since D has a finite number of extreme points. For further details on correctness of the algorithm and first computational experiments we refer to [23].

4.2 Linear Mixed-Integer Recourse

In the present subsection we allow for integer requirements to second-stage variables. Again we assume that $X \subseteq \mathbb{R}^m$ is a nonempty compact polyhedron and that μ is the discrete measure introduced at the beginning of the present section. We consider problem (4) with

$$\Phi(t) := \min\{q^T y \ : \ Wy \geq t, \ y \in Y\}. \tag{11}$$

For notational convenience we have integrated the former vector (y, y') into one vector y now varying in $Y := \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$. Accordingly, the former (q, q') and (W, W') are integrated into q and W. To be consistent with Subsection 4.1 we have inequality constraints in (11).

Lemma 4.3 *There exists a sufficiently large constant* $M_1 > 0$ *such that problem (4) can be equivalently restated as*

$$\min_{x,y,\theta} \{\sum_{j=1}^J \pi_j \theta_j \quad : \quad Wy_j \ \geq \ h_j - Tx, \quad q^T y_j + c^T x - \varphi_o \ \leq \ M_1 \theta_j, \quad x \in X, \ \ y_j \in Y, \ \ \theta_j \in \{0,1\}, \ \ j = 1, \ldots, J\}. \tag{12}$$

Proof: We choose M_1 by

$$M_1 \ := \ \sup\{c^T x + \Phi(h_j - Tx) \ : \ x \in X, \ j = 1, \ldots, J\}.$$

To see that this supremum is finite, recall the compactness of X and the general assumptions on Φ in the paragraph following formula (2). Part (iv) of Proposition 2.1 then confirms that $\Phi(h_j - Tx)$ remains bounded if x and j vary over X and $\{1, \ldots, J\}$, respectively.
The selection of M_1 guarantees that for any $x \in X$ and $y_j \in Y$ such that $Wy_j \geq h_j - Tx$ the selection $\theta_j = 1$ is feasible.
Given x, the selection $\theta_j = 0$ is feasible if and only if there exists a $y_j \in Y$ fulfilling $Wy_j \geq h_j - Tx$ and $c^T x + q^T y_j \leq \varphi_o$. The latter holds if and only if $c^T x + \Phi(h_j - Tx) \leq \varphi_o$ which is equvalent to $\chi(x, h_j) = 0$. This proves that (12) is equivalent to $\min\{\sum_{j=1}^J \pi_j \chi(x, h_j) \ : \ x \in X\}$. □

Compared with problem (8), problem (12) again arises by representing the indicator function χ from (7) by a binary variable. Lacking duality, however, prevents the usage of optimality cuts such that minimization with respect to y has to be carried out explicitly in (12). Hence, (8) is a variant of (12) where the linear programming nature of the second stage enables an algorithmic shortcut.

Problem (12) is a mixed-integer linear program that quickly becomes large-scale in practical applications. General purpose mixed-integer linear programming algorithms and software fail in such situations. As an alternative, we present a decomposition method based on Lagrangian relaxation of nonanticipativity.

This decomposition method for block-angular stochastic integer programs has been elaborated for the first time in [7] for the expectation-based model (3).

Introduce in (12) copies $x_j, j = 1, \ldots, J$, according to the number of scenarios, and add the nonanticipativity constraints $x_1 = \ldots = x_J$ (or an equivalent system), for which we use the notation $\sum_{j=1}^{J} H_j x_j = 0$ with proper $(l, n)-$matrices $H_j, j = 1, \ldots, J$. Problem (12) then becomes

$$\min_{x,y,\theta} \{\sum_{j=1}^{J} \pi_j \theta_j \quad : \quad Tx_j + Wy_j \ \geq \ h_j, \quad c^T x_j + q^T y_j - M_1 \theta_j \ \leq \ \varphi_o, \qquad x_j \in X, \ \ y_j \in Y, \ \ \theta_j \in \{0,1\}, \ \ j = 1, \ldots, J, \ \ \sum_{j=1}^{J} H_j x_j = 0\}. \tag{13}$$

This formulation suggests Lagrangian relaxation of the interlinking constraints $\sum_{j=1}^{J} H_j x_j = 0$. For $\lambda \in I\!R^l$ we consider the functions

$$L_j(x_j, y_j, \theta_j, \lambda) \ := \ \pi_j \theta_j \ + \ \lambda^T H_j x_j, \ \ j = 1, \ldots, J,$$

and form the Lagrangian

$$L(x, y, \theta, \lambda) \ := \ \sum_{j=1}^{J} L_j(x_j, y_j, \theta_j, \lambda).$$

The Lagrangian dual of (13) then is the optimization problem

$$\max\{D(\lambda) \ : \ \lambda \in I\!R^l\} \tag{14}$$

where

$$D(\lambda) \ = \ \min\{\sum_{j=1}^{J} L_j(x_j, y_j, \theta_j, \lambda) \quad : \quad Tx_j + Wy_j \ \geq \ h_j, \quad c^T x_j + q^T y_j - M_1 \theta_j \ \leq \ \varphi_o, \qquad x_j \in X, \ \ y_j \in Y, \ \ \theta_j \in \{0,1\}, \ \ j = 1, \ldots, J\}.$$

For separability reasons we have

$$D(\lambda) \ = \ \sum_{j=1}^{J} D_j(\lambda) \tag{15}$$

where

$$D_j(\lambda) \ = \ \min\{L_j(x_j, y_j, \theta_j, \lambda) \quad : \quad Tx_j + Wy_j \ \geq \ h_j, \quad c^T x_j + q^T y_j - M_1 \theta_j \ \leq \ \varphi_o, \qquad x_j \in X, \ \ y_j \in Y, \ \ \theta_j \in \{0,1\}\}. \tag{16}$$

$D(\lambda)$ being the pointwise minimum of affine functions in λ, it is piecewise affine and concave. Hence, (14) is a non-smooth concave maximization (or convex minimization) problem. Such problems can be tackled with advanced bundle methods, for instance with Kiwiel's proximal bundle method NOA 3.0, [17, 18]. At each iteration, these methods require the objective value and one subgradient of D. The structure of D, cf. (15), enables substantial decomposition, since the single-scenario problems (16) can be tackled separately. Their moderate size often allows application of general purpose mixed-integer linear programming codes.

Altogether, the optimal value z_{LD} of (14) provides a lower bound to the optimal value z of problem (12). From integer programming ([19]) it is well-known, that in general one has to live with a positive duality gap. On the other hand, it holds that $z_{LD} \geq z_{LP}$ where z_{LP} denotes the optimal value to the LP relaxation of (12). The lower bound obtained by the above procedure, hence, is never worse the bound obtained by eliminating the integer requirements.

In Lagrangian relaxation, the results of the dual optimization often provide starting points for heuristics to find promising feasible points. Our relaxed constraints being very simple ($x_1 = \ldots = x_N$), ideas for such heuristics come up straightforwardly. For example, examine the x_j-components, $j = 1, \ldots, J$, of solutions to (16) for optimal or nearly optimal λ, and decide for the most frequent value arising or average

and round if necessary.
If the heuristic yields a feasible solution to (12), then the objective value of the latter provides an upper bound $\bar{z}$ for z. Together with the lower bound z_{LD} this gives the quality certificate (gap) $\bar{z} - z_{LD}$.
The full algorithm improves this certificate by embedding the procedure described so far into a branch-and-bound scheme in the spirit of global optimization. Let $\mathcal{P}$ denote the list of current problems and $z_{LD} = z_{LD}(P)$ the Lagrangian lower bound for $P \in \mathcal{P}$. The algorithm then proceeds as follows.

Algorithm 4.4
Step 1 *(Initialization): Set $\bar{z} = +\infty$ and let $\mathcal{P}$ consist of problem (13).*
Step 2 *(Termination): If $\mathcal{P} = \emptyset$ then the solution $\hat{x}$ that yielded $\bar{z} = Q_{I\!P}(\hat{x})$, cf. (4), is optimal.*
Step 3 *(Node selection): Select and delete a problem P from $\mathcal{P}$ and solve its Lagrangian dual. If the optimal value $z_{LD}(P)$ hereof equals $+\infty$ (infeasibility of a subproblem) then go to Step 2.*
Step 4 *(Bounding): If $z_{LD}(P) \geq \bar{z}$ go to Step 2 (this step can be carried out as soon as the value of the Lagrangian dual rises above $\bar{z}$). Consider the following situations:*

1. *The scenario solutions x_j, $j = 1, \ldots, J$, are identical: If $Q_{I\!P}(x_j) < \bar{z}$ then let $\bar{z} = Q_{I\!P}(x_j)$ and delete from $\mathcal{P}$ all problems P' with $z_{LD}(P') \geq \bar{z}$. Go to Step 2.*

2. *The scenario solutions x_j, $j = 1, \ldots, J$ differ: Compute the average $\bar{x} = \sum_{j=1}^{J} \pi_j x_j$ and round it by some heuristic to obtain $\bar{x}^R$. If $Q_{I\!P}(\bar{x}^R) < \bar{z}$ then let $\bar{z} = Q_{I\!P}(\bar{x}^R)$ and delete from $\mathcal{P}$ all problems P' with $z_{LD}(P') \geq \bar{z}$. Go to Step 5.*

Step 5 *(Branching): Select a component $x_{(k)}$ of x and add two new problems to $\mathcal{P}$ obtained from P by adding the constraints $x_{(k)} \leq \lfloor \bar{x}_{(k)} \rfloor$ and $x_{(k)} \geq \lfloor \bar{x}_{(k)} \rfloor + 1$, respectively (if $x_{(k)}$ is an integer component), or $x_{(k)} \leq \bar{x}_{(k)} - \varepsilon$ and $x_{(k)} \geq \bar{x}_{(k)} + \varepsilon$, respectively, where $\varepsilon > 0$ is a tolerance parameter to have disjoint subdomains. Go to Step 3.*

The algorithm works both with and without integer requirements in the first stage. It is obviously finite in case X is bounded and all $x-$components have to be integers. If x is mixed-integer (or continuous, as in the former presentation) some stopping criterion to avoid endless branching on the continuous components has to be employed. Some first computational experiments with Algorithm 4.4 are reported in [36].

5 Multistage Extension

The two-stage stochastic programs introduced in Section 2 are based on the assumptions that uncertainty is unveiled at once and that decisions subdivide into those before and those after unveiling uncertainty. Often, a more complex view is appropriate at this place. Multistage stochastic programs address the situation where uncertainty is unveiled stepwise with intermediate decisions.
The modeling starts with a finite horizon sequential decision process under uncertainty where the decision $x_t \in I\!R^{m_t}$ at stage $t \in \{1, \ldots, T\}$ is based on information available up to time t only. Information is modeled as a discrete time stochastic process $\{\xi_t\}_{t=1}^{T}$ on some probability space $(\Omega, \mathcal{A}, I\!P)$ with ξ_t taking values in $I\!R^{s_t}$. The random vector $\xi^t := (\xi_1, \ldots, \xi_t)$ then reflects the information available up to time t. Nonanticipativity, i.e., the requirement that x_t must not depend on future information, is formalized by saying that x_t is measurable with respect to the $\sigma-$algebra $\mathcal{A}_t \subseteq \mathcal{A}$ which is generated by $\xi^t, t = 1, \ldots, T$. Clearly, $\mathcal{A}_t \subseteq \mathcal{A}_{t+1}$ for all $t = 1, \ldots, T-1$. As in the two-stage case, the first-stage decision x_1 usually is deterministic. Therefore, $\mathcal{A}_1 = \{\emptyset, \Omega\}$. Moreover, we assume that $\mathcal{A}_T = \mathcal{A}$.

The constraints of our multistage extensions can be subdivided into three groups. The first group comprises conditions on x_t arising from the individual time stages:

$$x_t(\omega) \in X_t, \quad B_t(\xi_t(\omega))x_t(\omega) \;\geq\; d_t(\xi_t(\omega)) \;\; I\!P\text{−almost surely}, \;\; t = 1, \ldots, T. \tag{17}$$

Here, $X_t \subseteq I\!R^{m_t}$ is a set whose convex hull is a polyhedron. In this way, integer requirements to components of x_t are allowed for. For simplicity we assume that X_t is compact. The next group of constraints models linkage between different time stages:

$$\sum_{\tau=1}^{t} A_{t\tau}(\xi_t(\omega))x_\tau(\omega) \;\geq\; g_t(\xi_t(\omega)) \;\; I\!P\text{−almost surely}, \;\; t = 2, \ldots, T. \tag{18}$$

Finally, there is the nonanticipativity of x_t, i. e.,

$$x_t \text{ is measurable with respect to } \mathcal{A}_t, \;\; t = 1, \ldots, T. \tag{19}$$

In addition to the constraints we have a linear objective function

$$\sum_{t=1}^{T} c_t(\xi_t(\omega))x_t(\omega).$$

The matrices $A_{t\tau}(.), B_t(.)$ as well as the right-hand sides $d_t(.), g_t(.)$ and the cost coefficients $c_t(.)$ all have conformal dimensions and depend affinely linearly on the relevant components of ξ.
The decisions x_t are understood as members of the function spaces $L_\infty(\Omega, \mathcal{A}, I\!P; I\!R^{m_t}), t = 1, \ldots, T$. The constraints (17), (18) then impose pointwise conditions on the x_t, whereas (19) imposes functional constraints, in fact, membership in a linear subspace of $\times_{t=1}^{T} L_\infty(\Omega, \mathcal{A}, I\!P; I\!R^{m_t})$, see e.g. [31] and the references therein.

Now we are in the position to formulate the multistage extensions to the expectation- and probability-based stochastic programs (3) and (4), respectively.
The multistage extension of (3) is the minimization of expected minimal costs subject to nonanticipativity of decisions:

$$\min \left\{ \int_\Omega \min_{x(\omega)} \Big\{ \sum_{t=1}^{T} c_t(\xi_t(\omega))x_t(\omega) \;:\; (17),\, (18) \Big\}\, I\!P(d\omega) \;\;:\;\; x \text{ fulfilling } (19) \right\} \tag{20}$$

To have the integral in the objective well-defined, the additional assumption $\xi_t \in L_1(\Omega, \mathcal{A}, I\!P; I\!R^{s_t})$, $t = 1, \ldots, T$, is imposed in model (20), see [31] for further details.

The multistage extension of (4) is the minimization of the probability that minimal costs do not exceed a preselected threshold $\varphi_o \in I\!R$. Again this minimization takes place over nonanticipative decisions only:

$$\min \left\{ I\!P\left(\Big\{ \omega \in \Omega \;:\; \min_{x(\omega)} \Big\{ \sum_{t=1}^{T} c_t(\xi_t(\omega))x_t(\omega) \;:\; (17),\, (18) \Big\} > \varphi_o \Big\} \right) \;\;:\;\; x \text{ fulfilling } (19) \right\} \tag{21}$$

The minimization in the integrand of (20) being separable with respect to $\omega \in \Omega$, it is possible to interchange integration and minimization. Then the problem can be restated as follows:

$$\min \Big\{ \int_\Omega \sum_{t=1}^{T} c_t(\xi_t(\omega))x_t(\omega)\, I\!P(d\omega) \;:\; x \text{ fulfilling } (17),\, (18),\, (19) \Big\}. \tag{22}$$

Extending the argument from Lemma 4.3 we introduce an additional variable $\theta \in L_\infty(\Omega, \mathcal{A}, I\!P; \{0,1\})$ as well as a sufficiently big constant $M > 0$. Then problem (21) can be equivalently rewritten as:

$$\min \Big\{ \int_\Omega \theta(\omega)\, I\!P(d\omega) \;\;:\;\; \sum_{t=1}^{T} c_t(\xi_t(\omega))x_t(\omega) - \varphi_o \le M \cdot \theta(\omega),\; \theta(\omega) \in \{0,1\}\; I\!P\text{-a.s.},$$
$$x \text{ fulfilling } (17),\, (18),\, (19) \Big\}. \tag{23}$$

Problem (22) is the well-known multistage stochastic (mixed-integer) linear program. Without integer requirements, the problem has been studied intensively, both from structural and from algorithmic viewpoints. The reader may wish to sample from [4, 5, 8, 10, 13, 15, 20, 25, 26, 27, 32] to obtain insights into these developments. With integer requirements, problem (22) is less well-understood. Existing results are reviewed in [31].

To the best of our knowledge, the multistage extension (21) has not been addressed in the literature so far. Some basic properties of (21), (23) regarding existence and structure of optimal solutions can be derived by following arguments that were employed for the expectation-based model (22) in [31]. Their mathematical foundations are laid out in [11, 12, 28]. The arguments can be outlined as follows:

Problem (23) concerns the minimization of an abstract expectation over a function space, subject to measurability with respect to a filtered sequence of $\sigma-$algebras. Theorems 1 and 2 in [12] (whose assumptions can be verified for (23) using statements from [11, 28]) provide sufficient conditions for the solvability of such minimization problems and for the solutions to be obtainable recursively by dynamic programming. The stage-wise recursion rests on minimizing in the $t-$th stage the regular conditional expectation (with respect to $\mathcal{A}_t$) of the optimal value from stage $t+1$. When arriving at the first-stage, a deterministic optimization problem in x_1 remains (recall that $\mathcal{A}_1 = \{\emptyset, \Omega\}$). Its objective function $Q^m_{I\!P}(x_1)$ can be regarded the multistage counterpart to the function $Q_{I\!P}(x)$ that we have studied in Section 3.

Given that (23) is a well-defined and solvable optimization problem, Sections 3 and 4 provide several points of departure for future research. For instance, unveiling the structure of $Q^m_{I\!P}(x_1)$ may be possible by analysing the interplay of conditional expectations and mixed-integer value functions. Regarding solution techniques, the extension of Algorithm 4.4 to the multistage situation may be fruitful. Indeed, it is well-known that the nonanticipativity in (19) is a linear constraint. With a discrete distribution of ξ this leads to a system of linear equations. Lagrangian relaxation of these constraints produces single-scenario subproblems, and the scheme of Algorithm 4.4 readily extends. However, compared with the two-stage situation, the relaxed constraints are more complicated such that primal heuristics are not that obvious, and the dimension l of the Lagrangian dual (14) may require approximative instead of exact solution of (14). Further algorithmic ideas for (23) may arise from Lagrangian relaxation of either (17) or (18). In [31] this is discussed for the expectation-based model (22).

Acknowledgement. I am grateful to Morten Riis (University of Aarhus), Werner Römisch (Humboldt-University Berlin), and Stephan Tiedemann (Gerhard-Mercator University Duisburg) for stimulating discussions and fruitful cooperation.

References

[1] Bank, B.; Mandel, R.: Parametric Integer Optimization, Akademie-Verlag, Berlin 1988.

[2] Bereanu, B.: Minimum risk criterion in stochastic optimization, Economic Computation and Economic Cybernetics Studies and Research 2 (1981), 31-39.

[3] Billingsley, P.: Convergence of Probability Measures, Wiley, New York, 1968.

[4] Birge, J.R.: Stochastic programming computation and applications. INFORMS Journal on Computing 9 (1997), 111-133.

[5] Birge, J.R.; Louveaux, F.: Introduction to Stochastic Programming, Springer, New York, 1997.

[6] Blair, C.E.; Jeroslow, R.G.: The value function of a mixed integer program: I, Discrete Mathematics 19 (1977), 121-138.

[7] Carøe, C.C.; Schultz, R.: Dual decomposition in stochastic integer programming, Operations Research Letters 24 (1999), 37-45.

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

[9] Dupačová, J.: Stochastic programming with incomplete information: a survey of results on postoptimization and sensitivity analysis, Optimization 18 (1987), 507-532.

[10] Dupačová, J.: Multistage stochastic programs: The state-of-the-art and selected bibliography, Kybernetika 31 (1995), 151-174.

[11] Dynkin, E.B., Evstigneev, I.V.: Regular conditional expectation of correspondences, Theory of Probability and Applications 21 (1976), 325-338.

[12] Evstigneev, I.: Measurable selection and dynamic programming, Mathematics of Operations Research 1 (1976), 267-272.

[13] Higle, J.L., Sen, S.: Duality in multistage stochastic programs, In: Prague Stochastics '98 (M. Hušková, P. Lachout, J.Á. Višek, Eds.), JČMF, Prague 1998, 233-236.

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

[15] Kall, P.; Wallace, S.W.: Stochastic Programming, Wiley, Chichester, 1994.

[16] Kibzun, A.I.; Kan, Y.S.: Stochastic Programming Problems with Probability and Quantile Functions, Wiley, Chichester, 1996.

[17] Kiwiel, K. C.: Proximity control in bundle methods for convex nondifferentiable optimization, Mathematical Programming 46 (1990), 105-122.

[18] Kiwiel, K. C.: User's Guide for NOA 2.0/3.0: A Fortran Package for Convex Nondifferentiable Optimization, Systems Research Institute, Polish Academy of Sciences, Warsaw, 1994.

[19] Nemhauser, G.L.; Wolsey, L.A.: Integer and Combinatorial Optimization, Wiley, New York 1988.

[20] Prékopa, A.: Stochastic Programming, Kluwer, Dordrecht, 1995.

[21] Raik, E.: Qualitative research into the stochastic nonlinear programming problems, Eesti NSV Teaduste Akademia Toimetised / Füüsika, Matemaatica (News of the Estonian Academy of Sciences / Physics, Mathematics) 20 (1971), 8-14. In Russian.

[22] Raik, E.: On the stochastic programming problem with the probability and quantile functionals, Eesti NSV Teaduste Akademia Toimetised / Füüsika, Matemaatica (News of the Estonian Academy of Sciences / Physics, Mathematics) 21 (1971), 142-148. In Russian.

[23] Riis, M.; Schultz, R.: Applying the minimum risk criterion in stochastic recourse programs, Preprint SM-DU-499, Department of Mathematics, Gerhard-Mercator University Duisburg, 2001.

[24] Robinson, S.M.; Wets, R.J-B: Stability in two-stage stochastic programming, SIAM Journal on Control and Optimization 25 (1987), 1409-1416.

[25] Rockafellar, R.T.: Duality and optimality in multistage stochastic programming, Annals of Operations Research 85 (1999), 1-19.

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

[27] Rockafellar, R.T., Wets, R.J-B: The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints, SIAM Journal on Control and Optimization 16 (1978), 16-36.

[28] Rockafellar, R.T., Wets, R.J-B: Variational Analysis, Springer-Verlag, Berlin, 1997.

[29] Römisch, W.; Schultz, R.: Stability analysis for stochastic programs, Annals of Operations Research 30 (1991), 241-266.

[30] Römisch, W.; Wakolbinger, A.: Obtaining convergence rates for approximations in stochastic programming, Parametric Optimization and Related Topics (J. Guddat, H.Th. Jongen, B. Kummer, F. Nožička, Eds.), Akademie Verlag, Berlin 1987, 327-343.

[31] Römisch, W.; Schultz, R.: Multistage stochastic integer programs: An introduction, Online Optimization of Large Scale Systems (M. Grötschel, S.O. Krumke, J. Rambau, Eds.), Springer-Verlag Berlin, 2001, 581-600.

[32] Ruszczyński, A.: Decomposition methods in stochastic programming. Mathematical Programming 79 (1997), 333-353.

[33] Schultz, R.: On structure and stability in stochastic programs with random technology matrix and complete integer recourse, Mathematical Programming 70 (1995), 73-89.

[34] Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse, SIAM Journal on Optimization 6 (1996), 1138-1152.

[35] Schultz, R.: Some aspects of stability in stochastic programming, Annals of Operations Research 100 (2000), 55-84.

[36] Tiedemann, S.: Probability Functionals and Risk Aversion in Stochastic Integer Programming, Diploma Thesis, Department of Mathematics, Gerhard-Mercator University Duisburg, 2001.