

# Risk Aversion via Excess Probabilities in Stochastic Programs with Mixed-Integer Recourse

Rüdiger Schultz\* and Stephan Tiedemann†  
Institute of Mathematics, Gerhard-Mercator University Duisburg  
Lotharstr. 65, D-47048 Duisburg, Germany

July 8, 2002

Revised: February 11, 2003

## Abstract

We consider linear two-stage stochastic programs with mixed-integer recourse. Instead of basing the selection of an optimal first-stage solution on expected costs alone, we include into the objective a risk term reflecting the probability that a preselected cost threshold is exceeded. After we have put the resulting mean-risk model into perspective with stochastic dominance, we study further structural properties of the model and derive some basic stability results. In the algorithmic part of the paper, we propose a scenario decomposition method and report initial computational experience.

**Key Words.** Stochastic programming, mean-risk models, mixed-integer optimization.

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

## 1 Introduction

Stochastic programming with recourse deals with two- or multi-stage sequential decision processes under uncertainty and aims, in its traditional setting, at the optimization of the expected value of some random objective reflecting costs or revenues, for instance. In the present paper, we are heading for an extension towards risk aversion with a risk measure based on excess probabilities of random costs.

Criteria for the selection of risk measures are a topic of extensive discussion in the literature, cf. e.g. [2, 31, 32] and the references therein. The discussion covers a wide range of issues, such as compatibility with axiomatic settings or stochastic ordering principles, smoothness and convexity properties, and, last but not least, algorithmic possibilities for the resulting optimization problems. It goes without saying that, given the variety of criteria, there is no universally recommendable risk measure.

In recourse stochastic programming, the random variables whose risk shall be controlled are implicit entities closely related with value functions that become discontinuous and nonconvex in the presence of integer decision variables. When imposing a risk measure in this situation, care has to be taken to arrive at stochastic integer programs that are structurally sound and amenable to algorithmic treatment. In what follows we will confirm that excess probabilities lead to a risk measure that serves these purposes and is consistent with first-order stochastic dominance [31, 32].

Our paper is organized as follows. In Section 2 we extend the traditional modeling in two-stage stochastic integer programming towards risk aversion. We formulate a risk measure based on excess probabilities, put it into perspective with stochastic dominance, and discuss relations of the resulting mean-risk model with robust optimization. Section 3 analyzes structure and stability of the added model components. Algorithmic issues including remarks on the efficient frontier and some first numerical experiments are presented in Section 4.

---

\*schultz@math.uni-duisburg.de

†tiedemann@math.uni-duisburg.de

## 2 Two-Stage Stochastic Integer Programs with Excess Probabilities

Throughout the paper, we impose a cost minimization framework. We are given the following random mixed-integer linear program

$$\min_{x,y,y'} \{c^T x + q^T y + q'^T y' : Tx + Wy + W'y' = h(\omega), x \in X, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'}\}. \quad (1)$$

It is assumed that all ingredients in (1) have conformal dimensions, that  $W, W'$  are rational matrices, and that  $X \subseteq \mathbb{R}^m$  is a nonempty closed set, possibly involving integer requirements to components of  $x$ . The right-hand side  $h(\omega) \in \mathbb{R}^s$  is a random vector on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Along with (1) we have the constraint that the variables  $x$  are to be fixed before observing  $h(\omega)$ , and that the variables  $(y, y')$  may be fixed afterwards. Accordingly,  $x$  and  $(y, y')$  are called first- and second-stage variables, respectively. The mentioned information constraint is usually referred to as nonanticipativity.

The mixed-integer value function

$$\Phi(t) := \min\{q^T y + q'^T y' : Wy + W'y' = t, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'}\} \quad (2)$$

is an essential object in our subsequent stochastic programming models. According to integer programming theory([30]), this function is real-valued on  $\mathbb{R}^s$  provided that  $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$  and  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$  which, therefore, will be assumed throughout.

With

$$Q_{\mathbb{E}}(x) := \int_{\Omega} (c^T x + \Phi(h(\omega) - Tx)) \mathbb{P}(d\omega) \quad (3)$$

the traditional expectation-based stochastic program with recourse is the optimization problem

$$\min\{Q_{\mathbb{E}}(x) : x \in X\}. \quad (4)$$

Introducing the excess probability functional

$$Q_{\mathbb{P}}(x) := \mathbb{P}(\{\omega \in \Omega : c^T x + \Phi(h(\omega) - Tx) > \varphi_o\}) \quad (5)$$

problem (4) is extended into the mean-risk model

$$\min\{Q_{\mathbb{E}}(x) + \rho Q_{\mathbb{P}}(x) : x \in X\}. \quad (6)$$

Here  $\varphi_o \in \mathbb{R}$  denotes some preselected threshold, and  $\rho \in \mathbb{R}_+$  is a suitable weight factor. The proposal to include a probability term like (5) into the objective of a stochastic program with recourse seemingly dates back to Bercanu [6] and, hitherto, has not been elaborated in much detail.

The modeling background behind the above construction is the following: The central issue is optimizing the first-stage decisions  $x$  that have to be taken without anticipation of future realizations of  $h(\omega)$ . After having decided for  $x$  and observed  $h(\omega)$ , the remaining decisions  $(y, y')$ , of course, shall be taken in an optimal way. This results in the mixed-integer linear program defining the function  $\Phi$  in (2). The costs of the sequential process of decision and observation then are expressed by the random variable  $c^T x + \Phi(h(\omega) - Tx)$ . Finding an optimal first-stage decision  $x \in X$  can be understood as selecting a “best” random variable from the indexed family  $(c^T x + \Phi(h(\omega) - Tx))_{x \in X}$ . The models (4) and (6) are based on scalar criteria for making this selection.

The mean-risk model (6) aims at controlling variability of second-stage solutions and, thus, is closely related with the robust optimization approach proposed by Mulvey, Vanderbei, and Zenios in [29]. To discuss similarities and differences between (6) and [29] we denote by  $(y_{opt}(x, \omega), y'_{opt}(x, \omega))$  an optimal solution to the optimization problem defining  $\Phi(h(\omega) - Tx)$ , cf. (2). The random variable  $c^T x + \Phi(h(\omega) - Tx)$  then coincides with

$$f(x, \omega) := c^T x + q^T y_{opt}(x, \omega) + q'^T y'_{opt}(x, \omega),$$

and the mean-risk model (6) can be written as

$$\min \{ \mathbb{E}_{\omega} [f(x, \omega)] + \rho Risk_{\omega} [f(x, \omega)] : x \in X \}. \quad (7)$$

Here  $\mathbb{E}_\omega$  denotes the expectation, and  $Risk_\omega$  is a symbol for an abstract risk measure, specified in our model by (5). The ROBUST model of [29] incorporates both variability of second-stage costs and penalization of second-stage infeasibility. In our terminology, the variability term of [29] is based on the random variable

$$g(x, \omega) := c^T x + q^T y(\omega) + q'^T y'(\omega),$$

and the counterpart model of [29] to our model (6) would read

$$\min \{ \mathbb{E}_\omega[g(x, \omega)] + \rho Risk_\omega[g(x, \omega)] \quad : \quad Tx + Wy(\omega) + W'y'(\omega) = h(\omega), x \in X, y(\omega) \in \mathbb{Z}_+^{\bar{m}}, y'(\omega) \in \mathbb{R}_+^{m'}, \forall \omega \in \Omega \} \quad (8)$$

where  $Risk_\omega$  is specified by the variance.

In (7) the statistical parameter  $\mathbb{E}_\omega[\cdot] + \rho Risk_\omega[\cdot]$  is optimized over all feasible first-stage solutions  $x \in X$  and all optimal second-stage decisions  $(y_{opt}(x, \omega), y'_{opt}(x, \omega))$ . In (8) the statistical parameter is optimized jointly over all feasible first- and second-stage decisions. The essential structural difference between (7) and (8) hence is in the order of integration and second-stage minimization. If the statistical parameter in (7), (8) is just  $\mathbb{E}_\omega[\cdot]$  then (7) and (8) are equivalent which is a basic fact of stochastic linear programming. For statistical parameters involving risk terms this equivalence is no longer valid in general. In particular, second-stage portions of optimal solutions to (8) need no longer be optimal for the second-stage, i.e., for the optimization problem behind  $\Phi(h(\omega) - Tx)$ . For problems without integer variables this issue is addressed in [21, 49].

In [49] it is shown that (7) and (8) are equivalent if the risk term depends on the second-stage costs only and fulfills a monotonicity condition. If, in addition, the risk term is convex, then numerical treatment of the problem is possible by a direct transfer of L-shaped decomposition techniques. Another issue discussed in [49] is ranking the random variables  $(c^T x + \Phi(h(\omega) - Tx))_{x \in X}$  by means of a convex disutility function. Again L-shaped techniques can readily be employed to solve the resulting optimization problem.

Apart from the integer decision variables in both stages of our model the major distinction between (6) and the ranking model of [49] is in the nonconvex discontinuous disutility function we employ, in fact, a sum of indicator functions of level sets. In Section 4 we will show how numerical treatment of (6) becomes possible by a reformulation using additional Boolean variables.

Among the fundamental concepts in decision theory the relation of stochastic dominance introduces a partial order in the space of real random variables. This provides some basis for selecting “best” members from families of random variables. Recently, Ogryczak and Ruszczyński have studied scalar criteria and their consistency with the multiobjective criteria induced by stochastic dominance, see [31, 32]. Let us quickly outline how the scalar criterion proposed in (6) can be put into the perspective of [31, 32].

Let  $f(x, \omega), x \in X \subseteq \mathbb{R}^m$  be real random variables on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Consider the distribution function

$$F_x(\eta) := \mathbb{P}(\{\omega \in \Omega : f(x, \omega) \leq \eta\}).$$

Since we prefer smaller outcomes to larger we define that  $x'$  dominates  $x''$  to first degree ( $x' \succeq x''$ ) if

$$F_{x'}(\eta) \geq F_{x''}(\eta) \quad \text{for all } \eta \in \mathbb{R}. \quad (9)$$

Let  $m_x$  be the expectation  $\mathbb{E}(f(x, \omega))$  and  $r_x$  denote some functional measuring the risk of the outcome  $f(x, \omega)$ . We adapt the setting of [31, 32] to preference of smaller outcomes and say that the mean-risk model  $(m_x, r_x)$  is  $\alpha$ -consistent with first degree stochastic dominance, where  $\alpha > 0$ , if  $x' \succeq x''$  implies  $m_{x'} + \alpha r_{x'} \leq m_{x''} + \alpha r_{x''}$ .

**Lemma 2.1** *The mean-risk model  $(m_x, r_x)$  with  $m_x := \mathbb{E}(f(x, \omega))$  and  $r_x := \mathbb{P}(\{\omega \in \Omega : f(x, \omega) > \varphi_o\})$ , with fixed  $\varphi_o \in \mathbb{R}$ , is  $\alpha$ -consistent with first degree stochastic dominance for all  $\alpha > 0$ .*

**Proof:** The fact that (9) implies  $m_{x'} \leq m_{x''}$  is well-known in probability theory. Moreover, it holds

$$r_{x'} = 1 - F_{x'}(\varphi_o) \leq 1 - F_{x''}(\varphi_o) = r_{x''},$$

and the proof is complete.  $\square$

In conclusion, the excess probability in (5) entails a risk measure fulfilling a weak requirement of consistency with stochastic dominance. For further risk measures fulfilling stronger consistencies with stochastic dominance we refer to [31, 32].

In the subsequent sections we will show that (6) is well-posed from formal viewpoint. We will establish structural properties of the functional  $Q_{\mathcal{P}}$ , and we will demonstrate that solution methodology from mixed-integer linear programming (the class our initial random optimization problem (1) belongs to) can be employed for solving (6).

The expectation-based optimization problem (4) meanwhile belongs to the well-studied objects in stochastic programming. Therefore, the main accent in our further investigations will be on the functional  $Q_{\mathcal{P}}$ . Without going into details, we mention that, under mild conditions,  $Q_{\mathcal{E}}$  is real-valued and lower semicontinuous and that  $Q_{\mathcal{E}}$  is continuous if the distribution of  $h(\omega)$  has a density. Optimal values and optimal solutions to (4) behave stable under perturbations of the probability distribution of  $h(\omega)$ . This gives rise to discrete approximations of the probability distribution for which (4) can be rewritten equivalently as a block-structured mixed-integer linear program. The latter is amenable to decomposition methods splitting (4) into smaller mixed-integer linear programs that are often tractable by standard solvers like CPLEX, [14]. Detailed expositions of the mentioned results can be found in [1, 13, 18, 46, 47]. Without integer requirements, (4) becomes a convex optimization problem allowing for application of various analytical and algorithmic techniques from convex analysis, see [10, 19, 35] and the references therein.

### 3 Structure and Stability

The mixed-integer value function  $\Phi$  from (2) is crucial for the structural understanding of  $Q_{\mathcal{P}}$ . From parametric optimization ([4, 11]) the following is known about  $\Phi$ .

**Proposition 3.1** *Assume that  $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$  and  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ . Then it holds*

- (i)  $\Phi$  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  $\Phi$  to  $\mathcal{T}_i$  are piecewise linear and Lipschitz continuous with a uniform constant not depending on  $i$ , more specifically, on each  $\mathcal{T}_i$ , the function  $\Phi$  admits a representation

$$\Phi(t) = \min_{y \in Y(t)} \{q^T y + \max_{k=1, \dots, K} d_k^T (t - Wy)\}$$

where  $Y(t) := \{y \in \mathbb{Z}_+^{\bar{m}} : t \in Wy + W'(\mathbb{R}_+^{m'})\}$  and  $d_k, k = 1, \dots, K$ , are the vertices of the polyhedron  $\{u \in \mathbb{R}^s : W'^T u \leq q'\}$ ;

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

- (iv) there exist positive constants  $\beta, \gamma$  such that  $|\Phi(t_1) - \Phi(t_2)| \leq \beta \|t_1 - t_2\| + \gamma$  whenever  $t_1, t_2 \in \mathbb{R}^s$ .

To facilitate notation we introduce for all  $x \in \mathbb{R}^m$

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

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

**Lemma 3.2** *Assume that  $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$  and  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ . Then it holds for all  $x \in \mathbb{R}^m$*

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

**Proof:** To show the first inclusion let  $h \in M(x)$ . By the lower semicontinuity of  $\Phi$  (Proposition 3.1(i)) we have

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

Therefore, there exists an  $n_o \in \mathbb{N}$  such that  $c^T x_n + \Phi(h - T x_n) > \varphi_o$  for all  $n \geq n_o$ , implying  $h \in M(x_n)$  for all  $n \geq n_o$ , and we obtain that  $M(x) \subseteq \liminf_{x_n \rightarrow x} M(x_n)$ .

The second inclusion being valid by definition we turn to the third. Let  $h \in \limsup_{x_n \rightarrow x} M(x_n) \setminus M(x)$ . Then there exists an infinite subset  $\tilde{N}$  of  $\mathbb{N}$  such that

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

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

For convenience, we denote by  $\mu$  the image measure  $\mathbb{P} \circ h^{-1}$  on  $\mathbb{R}^s$ . By the lower semicontinuity of  $\Phi$ , the sets  $M(x)$ ,  $M_e(x)$ , and  $M_d(x)$  are  $\mu$ -measurable for all  $x \in \mathbb{R}^m$ .

**Proposition 3.3** *Assume that  $W(\mathbb{Z}_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$  and  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ . Then  $Q_{\mathbb{P}} : \mathbb{R}^m \rightarrow \mathbb{R}$  is a real-valued lower semicontinuous function. If, in addition, it holds that  $\mu(M_e(x) \cup M_d(x)) = 0$ , then  $Q_{\mathbb{P}}$  is continuous at  $x$ . If  $\mu$  has a density, then  $Q_{\mathbb{P}}$  is continuous on  $\mathbb{R}^m$ .*

**Proof:** The function  $Q_{\mathbb{P}}$  is real-valued on  $\mathbb{R}^m$  due to the  $\mu$ -measurability of  $M(x)$ . By Lemma 3.2 and the (semi-) continuity of the probability measure on sequences of sets we have for all  $x \in \mathbb{R}^m$

$$Q_{\mathbb{P}}(x) = \mu(M(x)) \leq \mu(\liminf_{x_n \rightarrow x} M(x_n)) \leq \liminf_{x_n \rightarrow x} \mu(M(x_n)) = \liminf_{x_n \rightarrow x} Q_{\mathbb{P}}(x_n),$$

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

$$\begin{aligned} Q_{\mathbb{P}}(x) &= \mu(M(x)) = \mu(M(x) \cup M_e(x) \cup M_d(x)) \geq \mu(\limsup_{x_n \rightarrow x} M(x_n)) \\ &\geq \limsup_{x_n \rightarrow x} \mu(M(x_n)) = \limsup_{x_n \rightarrow x} Q_{\mathbb{P}}(x_n), \end{aligned}$$

and  $Q_{\mathbb{P}}$  is continuous at  $x$ . In view of Proposition 3.1 (ii), (iii), for given  $x \in \mathbb{R}^m$ , both  $M_e(x)$  and  $M_d(x)$  are contained in a countable union of hyperplanes, i.e., in a set of Lebesgue measure zero. Since  $\mu$  has a density, it is absolutely continuous with respect to the Lebesgue measure, hence  $\mu(M_e(x) \cup M_d(x)) = 0$ , and the proof is complete.  $\square$

**Remark 3.4** *The above lower semicontinuity of  $Q_{\mathbb{P}}$  implies in particular that problem (6) is well-posed in the sense that, provided  $X$  is compact, the infimum in (6) is finite and is attained. Given the discontinuity of  $\Phi$ , the well-posedness of (6) may become critical with other risk measures, cf. [27]. To see this let us consider the variance, leading to the functional*

$$Q_{\mathbb{V}}(x) := \int_{\Omega} \left[ c^T x + \Phi(h(\omega) - T x) - \int_{\Omega} (c^T x + \Phi(h(\omega) - T x)) \mathbb{P}(d\omega) \right]^2 \mathbb{P}(d\omega).$$

We consider the counterpart

$$\min\{Q_{\mathbb{E}}(x) + \rho Q_{\mathbb{V}}(x) : x \in X\} \tag{10}$$

to problem (6) with the specifications  $m = s = 1, c = 1, T = -1, \rho = 4, X = \{x \in \mathbb{R} : x \geq 0\}$ ,  $\Phi(t) = \min\{y : y \geq t, y \in \mathbb{Z}\}$ , and  $h(\omega)$  attaining the values 0 and  $\frac{1}{2}$  each with probability  $\frac{1}{2}$ . One computes that  $Q_{\mathbb{E}}(x) = x + \frac{1}{2}[x] + \frac{1}{2}[x + \frac{1}{2}]$  and  $Q_{\mathbb{V}}(x) = \frac{1}{4}([\![x] - [x + \frac{1}{2}]\!]^2)$ . Then (10) has the infimum 1, and any sequence  $(x_n)_{n \in \mathbb{N}}$  with  $x_n \downarrow 0, x_n \neq 0$  is a minimizing sequence. However, the infimum is not attained since the objective value for  $x = 0$  is  $\frac{3}{2}$ .

Before we turn to the Lipschitz continuity of  $Q_{\mathbb{P}}$  we study the boundary  $bd M(x)$  of the set  $M(x) = \{h \in \mathbb{R}^s : c^T x + \Phi(h - T x) > \varphi_o\}$ .

**Lemma 3.5** *Adopt the setting of Proposition 3.1 and assume in addition that  $q, q'$  are rational vectors. Then there exist a  $K_o \in \mathbb{N}$ , affine hyperplanes  $H_{\kappa y} \subseteq \mathbb{R}^s, \kappa = 1, \dots, K_o, y \in \mathbb{Z}_+^{\bar{m}}$ , and matrices  $T_\kappa, \kappa = 1, \dots, K_o$ , such that*

(i)

$$bd M(x) \subseteq \bigcup_{\kappa=1}^{K_o} \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \{T_\kappa x + H_{\kappa y}\} \quad \text{for all } x \in \mathbb{R}^m, \quad (11)$$

(ii) *there exists  $\mathbf{r} > 0$  such that for any  $\kappa \in \{1, \dots, K_o\}$  and any  $y', y'' \in \mathbb{Z}_+^{\bar{m}}$ , the hyperplanes  $H_{\kappa y'}, H_{\kappa y''}$  are either identical or have a Hausdorff distance of at least  $\mathbf{r}$ .*

**Proof:** Let us assume that  $W'(\mathbb{R}_+^{m'})$  has a representation  $\{t \in \mathbb{R}^s : \tilde{d}_l^T t \leq 0, l = 1, \dots, L\}$  with suitable vectors  $\tilde{d}_l, l = 1, \dots, L$ . According to Proposition 3.1(ii) we obtain that  $h$  belongs to the complement of  $M(x)$  if and only if there exists a  $y \in \mathbb{Z}_+^{\bar{m}}$  such that the following system of inequalities in  $h$  is fulfilled

$$q^T y + d_k^T (h - Tx - Wy) \leq \varphi_o - c^T x, \quad k = 1, \dots, K, \quad (12)$$

$$\tilde{d}_l^T (h - Tx - Wy) \leq 0, \quad l = 1, \dots, L. \quad (13)$$

If  $d_k = 0$  for some  $k \in \{1, \dots, K\}$  then the corresponding inequality in (12) turns into

$$q^T y \leq \varphi_o - c^T x$$

leading to a restriction on  $y$ , but not on  $h$ . It then would actually suffice to form the union in (11) over  $\mathbb{Z}_+^{\bar{m}} \cap \{y : q^T y \leq \varphi_o - c^T x\}$  instead of  $\mathbb{Z}_+^{\bar{m}}$ . So let us assume that  $d_k \neq 0$  for all  $k \in \{1, \dots, K\}$ . For some fixed  $k$  we assume that the first component  $d_{k(1)}$  of  $d_k$  is nonzero. We put

$$\tilde{T}_k := T - (d_{k(1)})^{-1} \begin{pmatrix} c^T \\ 0 \end{pmatrix}$$

and obtain that

$$d_k^T \tilde{T}_k = d_k^T T - (d_{k(1)})^{-1} d_k^T \begin{pmatrix} c^T \\ 0 \end{pmatrix} = d_k^T T - c^T.$$

This allows us to rewrite (12) as

$$q^T y + d_k^T (h - \tilde{T}_k x - Wy) \leq \varphi_o, \quad k = 1, \dots, K. \quad (14)$$

If  $h$  belongs to the boundary of  $M(x)$  then at least one of the inequalities in (13) and (14) has to be fulfilled as an equation. Defining the affine hyperplanes

$$H_{ky} := \{t \in \mathbb{R}^s : d_k^T t = \varphi_o + d_k^T W y - q^T y\}, \quad y \in \mathbb{Z}_+^{\bar{m}}, k \in \{1, \dots, K\} \quad (15)$$

and

$$H_{ly} := \{t \in \mathbb{R}^s : \tilde{d}_l^T t = \tilde{d}_l^T W y\}, \quad y \in \mathbb{Z}_+^{\bar{m}}, l \in \{1, \dots, L\} \quad (16)$$

we obtain that  $h \in bd M(x)$  implies

$$h \in \bigcup_{k=1}^K \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \{\tilde{T}_k x + H_{ky}\} \cup \bigcup_{l=1}^L \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \{Tx + H_{ly}\} =: \bigcup_{\kappa=1}^{K_o} \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \{T_\kappa x + H_{\kappa y}\},$$

proving (i).

Since  $q', W'$  are rational,  $d_k, k = 1, \dots, K$ , and  $\tilde{d}_l, l = 1, \dots, L$ , can be selected as rational vectors. According to the definitions in (15), (16) and the rationality of  $q, W$  this yields that there exists  $\mathbf{r} > 0$  such that, for arbitrary  $\kappa \in \{1, \dots, K_o\}$  and  $y', y'' \in \mathbb{Z}_+^{\bar{m}}$ , the hyperplanes  $H_{\kappa y'}, H_{\kappa y''}$  are either identical or have a Hausdorff distance that is bounded below by  $\mathbf{r}$ . The latter is seen as follows. For  $H_{\kappa y}$  we have a representation

$$H_{\kappa y} = \{t \in \mathbb{R}^s : \delta_\kappa^T t = b_{\kappa o} + b_\kappa^T y\} \quad (17)$$

with rational vectors  $\delta_\kappa, b_\kappa$ . So if  $H_{\kappa y'} \neq H_{\kappa y''}$ , then  $b_\kappa^T (y' - y'') \neq 0$ . Let  $b_* > 0$  denote the least common multiple of the (absolute values of) the denominators of the components of  $b_\kappa$ . By  $y' - y'' \in \mathbb{Z}$  the number  $b_\kappa^T (y' - y'') \neq 0$  then has to be a multiple of  $\frac{1}{b_*}$ . In the representations for  $H_{\kappa y'}$  and  $H_{\kappa y''}$  the right-hand sides therefore differ at least by  $\frac{1}{b_*}$ , proving the claim.  $\square$

**Proposition 3.6** *Assume that  $q, q'$  are rational vectors,  $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ ,  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ , and that for any nonsingular linear transformation  $B \in L(\mathbb{R}^s, \mathbb{R}^s)$  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  $Q_{\mathbb{P}}$  is Lipschitz continuous on any bounded subset of  $\mathbb{R}^m$ .*

**Proof:** Let  $S$  be a bounded subset of  $\mathbb{R}^m$  and  $x', x'' \in S$ . It holds

$$|Q_{\mathbb{P}}(x') - Q_{\mathbb{P}}(x'')| = |\mu(M(x')) - \mu(M(x''))| \leq \mu(M(x') \setminus M(x'')) + \mu(M(x'') \setminus M(x')).$$

For symmetry reasons it is sufficient to establish the assertion for the first member of the sum on the right.

Recall the representation (17) for the hyperplanes  $H_{\kappa y}$  arising in Lemma 3.5 and put  $\bar{\delta}_{\kappa y} := b_{\kappa o} + b_{\kappa}^T y$  such that we have  $H_{\kappa y} = \{t \in \mathbb{R}^s : \delta_{\kappa}^T t = \bar{\delta}_{\kappa y}\}$  for  $\kappa = 1, \dots, K_o, y \in \mathbb{Z}_+^{\bar{m}}$ . Consider the halfspaces

$$H_{\kappa y}^- := \{t \in \mathbb{R}^s : \delta_{\kappa}^T t \leq \bar{\delta}_{\kappa y}\}. \quad (18)$$

Then it holds

$$M(x') \setminus M(x'') \subseteq \bigcup_{\kappa=1}^{K_o} \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \left\{ \text{cl} \left[ \{T_{\kappa} x' + H_{\kappa y}^-\} \setminus \{T_{\kappa} x'' + H_{\kappa y}^-\} \right] \cup \text{cl} \left[ \{T_{\kappa} x'' + H_{\kappa y}^-\} \setminus \{T_{\kappa} x' + H_{\kappa y}^-\} \right] \right\}.$$

As usual, the symbol “cl” denotes the closure. To estimate the  $\mu$ -measure of the set on the right we fix some  $\kappa \in \{1, \dots, K_o\}$ . Without loss of generality we may assume that  $\delta_{\kappa}^T T_{\kappa} x'' \leq \delta_{\kappa}^T T_{\kappa} x'$  such that  $\{T_{\kappa} x'' + H_{\kappa y}^-\} \setminus \{T_{\kappa} x' + H_{\kappa y}^-\} = \emptyset$  for all  $y \in \mathbb{Z}_+^{\bar{m}}$ . It remains to estimate

$$\begin{aligned} & \mu \left( \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \text{cl} \left[ \{T_{\kappa} x' + H_{\kappa y}^-\} \setminus \{T_{\kappa} x'' + H_{\kappa y}^-\} \right] \right) \\ &= \mu \left( \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \{h \in \mathbb{R}^s : \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'' \leq \delta_{\kappa}^T h \leq \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'\} \right). \end{aligned}$$

Let  $B_{\kappa}$  be a nonsingular matrix whose first row coincides with  $\delta_{\kappa}^T$ . Let  $\zeta := B_{\kappa} h$  be the corresponding linear transformation and  $\zeta_{(1)}$  denote the first component of  $\zeta$ . Then it holds

$$\begin{aligned} & \mu(\{h \in \mathbb{R}^s : \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'' \leq \delta_{\kappa}^T h \leq \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'\}) \\ &= (\mu \circ B_{\kappa}^{-1})(B_{\kappa}(\{h \in \mathbb{R}^s : \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'' \leq \delta_{\kappa}^T h \leq \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'\})) \\ &= (\mu \circ B_{\kappa}^{-1})(\{\zeta \in \mathbb{R}^s : \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'' \leq \zeta_{(1)} \leq \bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'\}) \\ &= \int_{\bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x''}^{\bar{\delta}_{\kappa y} + \delta_{\kappa}^T T_{\kappa} x'} \theta_{\kappa}(\tau) d\tau. \end{aligned}$$

In the last row above,  $\theta_{\kappa}$  denotes a marginal density of the first component with respect to the image measure  $\mu \circ B_{\kappa}^{-1}$ . The density is selected such that it fulfills the requirements made in the assumptions of our proposition.

Let  $(\bar{\delta}_{\kappa i})_{i \in \mathbb{N}}$  be an enumeration of the distinct values attained by the numbers  $\bar{\delta}_{\kappa y}, y \in \mathbb{Z}_+^{\bar{m}}$ . By the argument from the proof of Lemma 3.5(ii), the sequence  $(\bar{\delta}_{\kappa i})_{i \in \mathbb{N}}$  has no accumulation points.

Since  $x', x''$  belong to the bounded set  $S$  and the  $\bar{\delta}_{\kappa i}$  do not accumulate, there exists an index  $\bar{i} = \bar{i}(S)$ , independent of  $x', x''$ , such that the intervals  $[\bar{\delta}_{\kappa i} + \delta_{\kappa}^T T_{\kappa} x'', \bar{\delta}_{\kappa i} + \delta_{\kappa}^T T_{\kappa} x']$ , up to renumbering, meet the bounded interval arising in the assumptions at most for  $i \leq \bar{i}$ . By assumption, we have an upper bound  $\bar{\theta}_{\kappa}$  for  $\theta_{\kappa}(\cdot)$ . For  $i > \bar{i}$ , we denote by  $\tilde{\tau}_{\kappa i}$  the left or right endpoint of  $[\bar{\delta}_{\kappa i} + \delta_{\kappa}^T T_{\kappa} x'', \bar{\delta}_{\kappa i} + \delta_{\kappa}^T T_{\kappa} x']$  depending on whether  $\theta_{\kappa}$  is decreasing or increasing on that interval. This allows the estimate

$$\begin{aligned} \sum_{i \in \mathbb{N}} \int_{\bar{\delta}_{\kappa i} + \delta_{\kappa}^T T_{\kappa} x''}^{\bar{\delta}_{\kappa i} + \delta_{\kappa}^T T_{\kappa} x'} \theta_{\kappa}(\tau) d\tau &\leq \sum_{i \leq \bar{i}} \bar{\theta}_{\kappa} \cdot \|\delta_{\kappa}^T T_{\kappa}\| \cdot \|x' - x''\| \\ &\quad + \sum_{i > \bar{i}} \theta_{\kappa}(\tilde{\tau}_{\kappa i}) \cdot \|\delta_{\kappa}^T T_{\kappa}\| \cdot \|x' - x''\|. \end{aligned}$$

Next we show that  $\sum_{i > \bar{i}} \theta_\kappa(\tilde{\tau}_{\kappa i})$  is finite. It is sufficient to do that for the sum over all  $i > \bar{i}$  belonging to those  $\tilde{\tau}_{\kappa i}$  around which  $\theta_\kappa$  is decreasing. For the remaining  $i > \bar{i}$  a similar argument applies. Since the  $\bar{\delta}_{\kappa i}$  do not accumulate, there exists an  $\varepsilon > 0$  such that

$$1 \geq \sum_i \int_{\tilde{\tau}_{\kappa i} - \varepsilon}^{\tilde{\tau}_{\kappa i}} \theta_\kappa(\tau) d\tau \geq \sum_i \int_{\tilde{\tau}_{\kappa i} - \varepsilon}^{\tilde{\tau}_{\kappa i}} \theta_\kappa(\tilde{\tau}_{\kappa i}) d\tau = \varepsilon \cdot \sum_i \theta_\kappa(\tilde{\tau}_{\kappa i}).$$

This provides the desired finiteness. Repeating the above arguments for all  $\kappa = 1, \dots, K_o$  we obtain a constant  $c_o > 0$ , not depending on  $x', x''$ , such that

$$\mu(M(x') \setminus M(x'')) \leq c_o \cdot \|x' - x''\|,$$

and the proof is complete.  $\square$

**Remark 3.7** Among the probability measures fulfilling the requirements of Proposition 3.6 there are the so called  $r$ -convex measures, in particular the (nondegenerate) multivariate normal distribution and the  $t$ -distribution. For details see [45].

**Remark 3.8** Without integer requirements ( $\bar{m} = 0$ ) the function  $\Phi$  is less complicated. Imposing the assumptions  $W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ ,  $\{u \in \mathbb{R}^s : W'^T u \leq q'\} \neq \emptyset$  one obtains due to linear programming duality

$$\Phi(t) = \max\{t^T u : W'^T u \leq q'\} = \max_{k=1, \dots, K} d_k^T t$$

where  $d_k, k = 1, \dots, K$  are the vertices of  $\{u \in \mathbb{R}^s : W'^T u \leq q'\}$ . This implies that, for all  $x \in \mathbb{R}^m$ , the set  $M_d(x)$  is empty and the complement  $M(x)^c$  of  $M(x)$  is a single polyhedron. This provides a link with linear chance constraints which belong to the well-studied objects in stochastic programming [10, 19, 35]. Lower semicontinuity of  $Q_{\mathbb{P}}$ , for instance, then already follows from Proposition 3.1 in [42]. Further material about  $Q_{\mathbb{P}}$  in the absence of integer requirements can be found in [39].

**Remark 3.9** For some early work on continuity properties of general probability functionals we refer to Raik ([37, 38], see also [20, 35]).

**Remark 3.10** With the additional assumption that  $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$  the statements of Propositions 3.3 and 3.6 are valid for  $Q_{\mathbb{E}}$  as well. The set  $M_e(x)$  turns out irrelevant for the continuity of  $Q_{\mathbb{E}}$  such that the corresponding assumption turns into  $\mu(M_d(x)) = 0$ . For details we refer to [45, 46].

In many practical modeling situations knowledge about the underlying probability measure  $\mu = \mathbb{P} \circ h^{-1}$  in (6) is subjective. Furthermore, the multivariate integration required in (3) and (5) often has to rely on approximations, in particular if  $h(\omega)$  is multidimensional and follows a continuous probability distribution. These issues motivate the stability analysis of (6) under perturbations of  $\mu$ . The aim is to identify sufficient conditions such that “small” perturbations in  $\mu$  result in only “small” perturbations of optimal values and optimal solutions to (6). Qualitative and quantitative continuity of  $Q_{\mathbb{E}}$  and  $Q_{\mathbb{P}}$  jointly in  $x$  and  $\mu$  become a key issue then. For  $Q_{\mathbb{E}}$  this has been settled in [1, 46, 47] such that we will now focus on  $Q_{\mathbb{P}}$ .

Let  $\mathcal{P}(\mathbb{R}^s)$  denote the set of all Borel probability measures on  $\mathbb{R}^s$ . We consider  $Q_{\mathbb{P}}$  as a function mapping from  $\mathbb{R}^m \times \mathcal{P}(\mathbb{R}^s)$  to  $\mathbb{R}$  where  $\mathbb{R}^m$  is equipped with the usual topology. On  $\mathcal{P}(\mathbb{R}^s)$  a notion of convergence is desirable that is both sufficiently general to cover relevant applications and sufficiently specific to enable substantial statements. This is met by weak convergence of probability measures for which [9] is a basic reference. We say that a sequence  $\{\mu_n\}$  in  $\mathcal{P}(\mathbb{R}^s)$  converges weakly to  $\mu \in \mathcal{P}(\mathbb{R}^s)$ , written  $\mu_n \xrightarrow{w} \mu$ , if for any bounded continuous function  $g : \mathbb{R}^s \rightarrow \mathbb{R}$  it holds

$$\int_{\mathbb{R}^s} g(h) \mu_n(dh) \rightarrow \int_{\mathbb{R}^s} g(h) \mu(dh) \quad \text{as } n \rightarrow \infty. \quad (19)$$

**Proposition 3.11** Assume that  $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$  and  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ . Let  $\mu \in \mathcal{P}(\mathbb{R}^s)$  be such that  $\mu(M_e(x) \cup M_d(x)) = 0$ . Then  $Q_{\mathbb{P}} : \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^s) \rightarrow \mathbb{R}$  is continuous at  $(x, \mu)$ .



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

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

Now we have  $E \subseteq M_e(x) \cup M_d(x)$ . To see this, consider  $h \in (M_e(x) \cup M_d(x))^c = M_e(x)^c \cap M_d(x)^c$  where the superscript  $c$  denotes the set-theoretic complement. Then  $\Phi$  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  $h_n \rightarrow h$  there exists an  $n_o \in \mathbb{N}$  such that for all  $n \geq n_o$  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,  $\chi_n(h_n) \rightarrow \chi(h)$  as  $h_n \rightarrow h$ , implying  $h \in E^c$ .

In view of  $E \subseteq M_e(x) \cup M_d(x)$  and  $\mu(M_e(x) \cup M_d(x)) = 0$  we obtain that  $\mu(E) = 0$ . A theorem on weak convergence of image measures attributed to Rubin in [9], 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

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

In other words,  $\mu_n(M(x_n)) \rightarrow \mu(M(x))$  or  $Q_{\mathcal{P}}(x_n, \mu_n) \rightarrow Q_{\mathcal{P}}(x, \mu)$ .  $\square$

To analyze the quantitative continuity of  $Q_{\mathcal{P}}$  as a function of the underlying probability measure let us again consider the hyperplane arrangement

$$\bigcup_{\kappa=1}^{K_o} \bigcup_{y \in \mathbb{Z}_+^{\bar{m}}} \{T_\kappa x + H_{\kappa y}\} \quad (20)$$

arising in Lemma 3.5. Associated with that arrangement there are the affine halfspaces

$$T_\kappa x + H_{\kappa y}^- \quad \text{and} \quad T_\kappa x + H_{\kappa y}^+, \quad \kappa = 1, \dots, K_o, \quad y \in \mathbb{Z}_+^{\bar{m}} \quad (21)$$

where  $H_{\kappa y}^-$  is defined as in (18) and, accordingly,  $H_{\kappa y}^+ := \{t \in \mathbb{R}^s : \delta_\kappa^T t \geq \bar{\delta}_{\kappa y}\}$ .

Let  $\Pi(x)$  denote the family of all, not necessarily full-dimensional polyhedra in  $\mathbb{R}^s$  arising as intersections of halfspaces from (21). In the proof of Lemma 3.5 we have seen that the complement of  $M(x)$  is a countable union of polyhedra each arising as an intersection of halfspaces from (21). Thus, the set  $M(x)$  admits a representation

$$M(x) = \bigcup_{\iota=1}^{\infty} P_\iota(x) \quad (22)$$

such that  $P_{\iota_1}(x) \cap P_{\iota_2}(x) = \emptyset$  whenever  $\iota_1 \neq \iota_2$ , and, for all  $\iota \in \mathbb{N}$ , the closure  $cl P_\iota(x)$  belongs to  $\Pi(x)$ . Consider the outer normals  $\delta_\kappa$  and  $-\delta_\kappa$  of the affine halfspaces  $H_{\kappa y}^-$  and  $H_{\kappa y}^+$ , respectively. By  $\mathcal{B}_o$  we denote the family of all subsets in  $\mathbb{R}^s$  which are given as intersections of affine halfspaces with outer normals in  $\{\pm \delta_\kappa : \kappa = 1, \dots, K_o\}$ . Clearly,  $\Pi(x) \subseteq \mathcal{B}_o$  for all  $x \in \mathbb{R}^m$ , provided the setting of Lemma 3.5 is adopted.

The representation (22) now gives rise to the following variational distance of probability measures in  $\mathcal{P}(\mathbb{R}^s)$

$$\alpha_{\mathcal{B}_o}(\mu, \nu) := \sup \{|\mu(B) - \nu(B)| : B \in \mathcal{B}_o\}. \quad (23)$$

We further introduce

$$\Delta_{a,C}(\mathbb{R}^s) := \left\{ \nu \in \mathcal{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \|h\|^a \nu(dh) \leq C \right\}$$

where  $a > 0$  and  $C > 0$  are fixed constants.

**Proposition 3.12** *Assume that  $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$  and  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ , and that  $q, q'$  are rational vectors. Then there exists a constant  $L_o > 0$  such that*

$$|Q_{\mathcal{P}}(x, \mu) - Q_{\mathcal{P}}(x, \nu)| \leq L_o \cdot \alpha_{\mathcal{B}_o}(\mu, \nu)^{\frac{a}{s+a}}$$

for all  $x \in \mathbb{R}^m$  and all  $\mu, \nu \in \Delta_{a,C}(\mathbb{R}^s)$  with  $\alpha_{\mathcal{B}_o}(\mu, \nu) \neq 0$ .

**Proof:** Let  $x \in \mathbb{R}^m$  and  $\mu, \nu \in \Delta_{a,C}(\mathbb{R}^s)$  such that  $\alpha_{\mathcal{B}_o}(\mu, \nu) \neq 0$ . With  $R := \alpha_{\mathcal{B}_o}(\mu, \nu)^{-\frac{1}{s+a}}$  we consider the ball  $B_R := \{h \in \mathbb{R}^s : \|h\| \leq R\}$ . Recall the representation (22) for  $M(x)$  and denote  $\mathcal{N}_o := \{\iota \in \mathcal{N} : P_\iota(x) \cap B_R \neq \emptyset\}$ . Then it holds

$$\begin{aligned}
|Q_{\mathbb{P}}(x, \mu) - Q_{\mathbb{P}}(x, \nu)| &= |\mu(M(x)) - \nu(M(x))| \\
&\leq \sum_{\iota \in \mathcal{N}} |\mu(P_\iota(x)) - \nu(P_\iota(x))| \\
&\leq \sum_{\iota \in \mathcal{N}_o} |\mu(P_\iota(x)) - \nu(P_\iota(x))| + (\mu + \nu)(\{h \in \mathbb{R}^s : \|h\| \geq R\}) \\
&\leq \sum_{\iota \in \mathcal{N}_o} |\mu(P_\iota(x)) - \nu(P_\iota(x))| + \frac{2C}{R^a}
\end{aligned} \tag{24}$$

where Markov's Inequality has been used in the last estimate. We continue by estimating the cardinality of  $\mathcal{N}_o$ . Due to Lemma 3.5(ii) there exists a constant  $c_1 > 0$ , which does not depend on  $x$ , such that at most  $c_1 \cdot R$  hyperplanes from the arrangement in (20) intersect the ball  $B_R$ . From the theory of hyperplane arrangements it is known that the complement of an arrangement of  $N$  hyperplanes in  $\mathbb{R}^s$  consists of at most  $\sum_{i=0}^s \binom{N}{i} = O(N^s)$  connected cells of dimension  $s$ , see [12, 33]. Hence there exists a constant  $c_2 > 0$  such that at most  $c_2 \cdot R^s$  full-dimensional sets  $P_\iota(x)$  intersect the ball  $B_R$ . Since there are only finitely many normals  $\delta_\kappa, \kappa = 1, \dots, K_o$ , in the arrangement (20), there exists a constant  $c_3 > 0$ , again not depending on  $x$ , such that the number of all (not necessarily full-dimensional) sets  $P_\iota(x)$  intersecting the ball  $B_R$ , in other words the cardinality of  $\mathcal{N}_o$ , is bounded above by  $c_3 \cdot R^s$ .

For any  $\iota \in \mathcal{N}_o$  we have the estimate

$$|\mu(P_\iota(x)) - \nu(P_\iota(x))| \leq \alpha_{\mathcal{B}_o}(\mu, \nu)$$

where, in case  $P_\iota(x)$  is not closed,  $P_\iota(x)$  is approximated by a sequence that is monotonically increasing with respect to set inclusion and that consists of closed polyhedra from  $\mathcal{B}_o$  which are contained in the relative interior of  $P_\iota(x)$ . Altogether, this allows to continue the estimate (24) as follows

$$\begin{aligned}
&\leq c_3 \cdot R^s \cdot \alpha_{\mathcal{B}_o}(\mu, \nu) + \frac{2C}{R^a} \\
&\leq c_3 \cdot \alpha_{\mathcal{B}_o}(\mu, \nu)^{-\frac{s}{s+a}+1} + 2C \cdot \alpha_{\mathcal{B}_o}(\mu, \nu)^{\frac{a}{s+a}} = (c_3 + 2C) \cdot \alpha_{\mathcal{B}_o}(\mu, \nu)^{\frac{a}{s+a}},
\end{aligned}$$

and the proof is complete.  $\square$

**Remark 3.13** In general,  $\alpha_{\mathcal{B}_o}$  need not define a metric on  $\mathcal{P}(\mathbb{R}^s)$  since  $\alpha_{\mathcal{B}_o}(\mu, \nu) = 0$  is possible with  $\mu \neq \nu$ . This can be overcome by enriching  $\mathcal{B}_o$ , for instance by adding the canonical basis vectors in  $\mathbb{R}^s$  to the set of relevant outer normals in the definition of  $\mathcal{B}_o$ . Then  $\alpha_{\mathcal{B}_o}$  majorizes the uniform distance of distribution functions which is known to be a metric on  $\mathcal{P}(\mathbb{R}^s)$ , and Proposition 3.12 holds without the restriction that  $\alpha_{\mathcal{B}_o}(\mu, \nu) \neq 0$ .

**Remark 3.14** Under suitable assumptions, weak convergence of probability measures implies convergence in  $\alpha_{\mathcal{B}_o}$ , and Proposition 3.12 can be seen as a quantification of Proposition 3.11: A class  $\mathcal{B}$  of Borel sets in  $\mathbb{R}^s$  is called a  $\mu$ -uniformity class if  $\alpha_{\mathcal{B}}(\mu_n, \mu) \rightarrow 0$  holds for every sequence  $\mu_n$  in  $\mathcal{P}(\mathbb{R}^s)$  converging weakly to  $\mu \in \mathcal{P}(\mathbb{R}^s)$ . By Theorem 2.11 in [8] the family of all convex Borel sets in  $\mathbb{R}^s$  is a  $\mu$ -uniformity class provided  $\mu$  has a density. Since all members of  $\mathcal{B}_o$  are convex and Borel, weak convergence of probability measures implies convergence in  $\alpha_{\mathcal{B}_o}$  if the limiting measure has a density.

**Remark 3.15** The fact that the number of affine halfspaces defining the members of  $\mathcal{B}_o$  is uniformly bounded implies speed-of-convergence estimates for  $\alpha_{\mathcal{B}_o}$  in the context of estimation by empirical measures. Given a sequence  $\xi_1, \xi_2, \dots, \xi_n, \dots$  of independent  $\mathbb{R}^s$ -valued random variables on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  with joint distribution  $\mu$ , the empirical measures  $\mu_n(\omega)$  ( $\omega \in \Omega, n \in \mathbb{N}$ ) are defined by

$$\mu_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i(\omega)}$$

where  $\delta_{\xi_i(\omega)}$  denotes the measure with unit mass at  $\xi_i(\omega)$  (cf. [15, 34, 48]). For the variational distance  $\alpha_{\mathcal{B}}(\mu, \mu_n(\omega))$  then the following law of iterated logarithm established in [26] holds, provided that  $\mathcal{B}$  is a so called Vapnik-Červonenkis class:

$$\limsup_{n \rightarrow \infty} \left( \frac{n}{2 \log \log n} \right)^{1/2} \cdot \alpha_{\mathcal{B}}(\mu, \mu_n(\omega)) \leq \frac{1}{2} \quad \text{for } \mathbb{P}\text{-almost all } \omega \in \Omega. \quad (25)$$

A family  $\mathcal{B}$  of Borel sets in  $\mathbb{R}^s$  is called a Vapnik-Červonenkis class if there exists an  $m_o \in \mathbb{N}$  such that for any finite set  $E \subset \mathbb{R}^s$  with  $m_o$  elements not every subset  $E_o$  of  $E$  arises as an intersection  $E_o = E \cap B$  for some  $B \in \mathcal{B}$ . The catch is now that, thanks to the uniform bound on the number of defining half spaces, the family  $\mathcal{B}_o$  is a Vapnik-Červonenkis class, for details see e.g. [34, 48]. Proposition 3.12 and (25) then provide a speed-of-convergence estimate for  $|Q_{\mathbb{P}}(x, \mu) - Q_{\mathbb{P}}(x, \mu_n(\omega))|$ .

Propositions 3.11 and 3.12 are the essential ingredients for studying the stability of optimal solutions to optimization problems whose objective function involves the excess probability functional  $Q_{\mathbb{P}}$ . Stability of the traditional expectation-based stochastic program (4) was studied in [1, 36, 46, 47]. We will close this section with some stability results for the risk minimization problem

$$P(\mu) \quad \min\{Q_{\mathbb{P}}(x, \mu) : x \in X\}.$$

This specific problem has been chosen to display the direct impact of Propositions 3.11 and 3.12 on stability. If one is interested in the stability of the mean-risk model (6) one has to combine the assumptions in the statements below with assumptions in [1, 36, 46, 47].

In general, the function  $Q_{\mathbb{P}}(\cdot, \mu)$  is nonconvex such that an analysis of local optimal solutions is appropriate. To this end we follow [40, 24] and consider localized optimal values and solution sets. With some subset  $V \subset \mathbb{R}^m$  we define

$$\begin{aligned} \varphi_V(\mu) &:= \inf\{Q_{\mathbb{P}}(x, \mu) : x \in X \cap \text{cl } V\}, \\ \Psi_V(\mu) &:= \{x \in X \cap \text{cl } V : Q_{\mathbb{P}}(x, \mu) = \varphi_V(\mu)\}. \end{aligned}$$

Given  $\mu \in \mathcal{P}(\mathbb{R}^s)$ , a nonempty set  $Z \subset \mathbb{R}^m$  is called a complete local minimizing set (CLM set) of  $P(\mu)$  with respect to  $V$  if  $V$  is open and  $Z = \Psi_V(\mu) \subset V$ . Examples for CLM sets are the set of global minimizers and isolated local minimizers. The basic feature of CLM sets is that they contain all local minimizers “nearby”. Without such a completeness property, pathologies may occur under perturbations, see [40, 24] for details.

**Proposition 3.16** *Assume that  $W(\mathbb{Z}_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$  and  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ , that  $q, q'$  are rational vectors, and that  $\mu \in \mathcal{P}(\mathbb{R}^s)$  has a density. Suppose further that there exists a subset  $Z \subset \mathbb{R}^m$  which is a CLM set for  $P(\mu)$  with respect to some bounded open set  $V \subset \mathbb{R}^m$ . Then it holds*

- (i) *the function  $\varphi_V : \mathcal{P}(\mathbb{R}^s) \rightarrow \mathbb{R}$  is continuous at  $\mu$ , where  $\mathcal{P}(\mathbb{R}^s)$  is equipped with weak convergence of probability measures,*
- (ii) *the multifunction  $\Psi_V : \mathcal{P}(\mathbb{R}^s) \rightarrow 2^{\mathbb{R}^m}$  is Berge upper semicontinuous at  $\mu$ , i.e., for any open set  $\mathcal{O}$  in  $\mathbb{R}^m$  with  $\mathcal{O} \supseteq \Psi_V(\mu)$  there exists a neighborhood  $\mathcal{N}$  of  $\mu$  in  $\mathcal{P}(\mathbb{R}^s)$ , again equipped with the topology of weak convergence of probability measures, such that  $\Psi_V(\nu) \subseteq \mathcal{O}$  for all  $\nu \in \mathcal{N}$ ,*
- (iii) *there exists a neighborhood  $\mathcal{N}'$  of  $\mu$  in  $\mathcal{P}(\mathbb{R}^s)$  such that for all  $\nu \in \mathcal{N}'$  the set  $\Psi_V(\nu)$  is a CLM set for  $P(\nu)$  with respect to  $V$ ,*
- (iv) *there exists a constant  $L_o > 0$  such that*

$$|\varphi_V(\mu) - \varphi_V(\nu)| \leq L_o \cdot \alpha_{\mathcal{B}_o}(\mu, \nu)^{\frac{a}{s+a}}$$

*for all  $\mu, \nu \in \Delta_{a,C}(\mathbb{R}^s)$  with  $\alpha_{\mathcal{B}_1}(\mu, \nu) \neq 0$ .*

Before proving the above proposition let us add a few comments. The above assertions are paradigmatic statements in the stability analysis of nonconvex optimization problems. Their proofs rely on well-established arguments that date back (at least) to Berge [7] and that were adapted and extended by many authors, cf. [3, 41], for instance. The main ingredients to make these arguments work are qualitative and quantitative continuity properties as established in Propositions 3.11 and 3.12 together with nonemptiness and compactness of the unperturbed solution set that, in Proposition 3.16, is hidden in the boundedness assumption on  $V$ . Therefore, we will refrain from presenting all details of the proof and merely outline its main ideas.

**Proof:** Using the joint continuity established in Proposition 3.11 the proof of (i) and (ii) follows the lines of Berge's theory as displayed, for instance, in the proof of Theorem 4.2.2 in [3].

To prove (iii), one first confirms the nonemptiness of  $\Psi_V(\nu)$  which is a consequence of the lower semicontinuity of  $Q_{\mathbb{P}}(\cdot, \nu)$ , see Proposition 3.3, together with the nonemptiness and compactness of  $X \cap cl V$ . The CLM property then follows from (ii).

For proving (iv) we, as in the proof of (iii), confirm that  $\Psi_V(\mu) \neq \emptyset$  and  $\Psi_V(\nu) \neq \emptyset$ . Let  $\mu, \nu \in \Delta_{a,C}(\mathbb{R}^s)$  and  $x_\nu \in \Psi_V(\nu)$ ,  $x_\mu \in \Psi_V(\mu)$ . Then it holds

$$\varphi_V(\mu) \leq Q_{\mathbb{P}}(x_\nu, \mu) \leq \varphi_V(\nu) + |Q_{\mathbb{P}}(x_\nu, \mu) - Q_{\mathbb{P}}(x_\nu, \nu)|$$

and

$$\varphi_V(\nu) \leq Q_{\mathbb{P}}(x_\mu, \nu) \leq \varphi_V(\mu) + |Q_{\mathbb{P}}(x_\mu, \nu) - Q_{\mathbb{P}}(x_\mu, \mu)|.$$

Together with Proposition 3.12 this implies

$$|\varphi_V(\nu) - \varphi_V(\mu)| \leq L_o \cdot \alpha_{\mathcal{B}_o}(\nu, \mu)^{\frac{a}{s+a}},$$

and the proof is complete.  $\square$

**Remark 3.17** *Due to the lower semicontinuity of  $Q_{\mathbb{P}}(\cdot, \nu)$  and the fact that  $X \cap cl V$  is nonempty and compact, nonemptiness of  $\Psi_V(\nu)$  is immediate. Not immediate, however, is that  $\Psi_V(\nu)$  consists of local minimizers to  $P(\nu)$ , i.e., when minimizing over  $X$ . The latter is confirmed by assertion (iii) above, which says that for all  $\nu \in \mathcal{N}'$  the set  $\Psi_V(\nu)$  is a CLM set and hence a set of local minimizers to  $P(\nu)$ .*

## 4 Algorithm and Computational Experiments

The following result establishes a useful link between two-stage stochastic programs with excess probabilities and traditional expectation-based two-stage models. Before stating the proposition we recall that the support  $supp \mu$  of  $\mu \in \mathcal{P}(\mathbb{R}^s)$  is the smallest closed subset of  $\mathbb{R}^s$  with  $\mu$ -measure 1.

**Proposition 4.1** *Assume that  $W(\mathbb{Z}_+^{\tilde{n}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ ,  $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$ , and that  $\mu$  has bounded support. Then the following holds.*

(i) *For any  $x \in \mathbb{R}^m$  there exists a constant  $M_x > 0$  such that*

$$Q_{\mathbb{P}}(x) = \tilde{Q}_{\mathbb{E}}(x) := \int_{\mathbb{R}^s} \tilde{\Phi}(h - Tx, c^T x - \varphi_o) \mu(dh) \quad (26)$$

where

$$\tilde{\Phi}(t_1, t_2) := \min\{\theta : Wy + W'y' = t_1, -q^T y - q'^T y' + (M_x - \varphi_o)\theta \geq t_2, \\ y \in \mathbb{Z}_+^{\tilde{n}}, y' \in \mathbb{R}_+^{m'}, \theta \in \{0, 1\}\}. \quad (27)$$

(ii) *If, in addition,  $X$  is bounded then  $M_x$  in (i) can be chosen as a uniform constant  $M$  for all  $x \in X$ , and the stochastic programs*

$$\min\{Q_{\mathbb{P}}(x) : x \in X\} \quad \text{and} \quad \min\{\tilde{Q}_{\mathbb{E}}(x) : x \in X\}$$

*are equivalent.*

**Proof:** To prove (i) we define

$$\mathbb{M}_x := \sup\{c^T x + \Phi(h - Tx) : h \in \text{supp } \mu\}.$$

This supremum is finite since, by Proposition 3.1(iv) and  $\Phi(0) = 0$ , it holds

$$|\Phi(h - Tx)| = |\Phi(h - Tx) - \Phi(0)| \leq \beta\|h\| + \beta\|Tx\| + \gamma, \quad (28)$$

and since  $\text{supp } \mu$  is bounded.

Next we verify that the integral in (26) is taken over a real-valued function, i.e., that  $\tilde{\Phi}(h - Tx, c^T x - \varphi_o) \in \mathbb{R}$  for all  $h \in \text{supp } \mu$ . By Proposition 3.1(i) it holds that  $\Phi(t) \in \mathbb{R}$  for all  $t \in \mathbb{R}^s$ . Hence, the optimization problem associated with  $\Phi(h - Tx)$ , cf. (2), is solvable for all  $h \in \text{supp } \mu$ . Let  $h \in \text{supp } \mu$  and  $(y, y') \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'}$  be an optimal solution associated with  $\Phi(h - Tx)$ . Then we have

$$Wy + W'y' = h - Tx \quad \text{and} \quad \mathbb{M}_x - \varphi_o \geq c^T x + \Phi(h - Tx) - \varphi_o = c^T x + q^T y + q'^T y' - \varphi_o.$$

Hence, the tuple  $(y, y', 1)$  is feasible for the optimization problem associated with  $\tilde{\Phi}(h - Tx, c^T x - \varphi_o)$ , cf. (27). Since this optimization problem has an objective with values in  $\{0, 1\}$  only, it is solvable, and  $\tilde{\Phi}(h - Tx, c^T x - \varphi_o) \in \mathbb{R}$ .

The integral in (26) makes sense for measurable functions only. Therefore, we have to show that  $\tilde{\Phi}$  is measurable. Since  $\tilde{\Phi}$  takes values in the finite set  $\{0, 1\}$ , it is sufficient to show measurability of the pre-images  $\tilde{\Phi}^{-1}(\{0\})$  and  $\tilde{\Phi}^{-1}(\{1\})$ . For these sets we have the following representations:

$$\begin{aligned} \tilde{\Phi}^{-1}(\{0\}) &= \{(t_1, t_2) \in \mathbb{R}^{s+1} : \exists (y, y') \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'} \quad Wy + W'y' = t_1, \quad q^T y + q'^T y' \leq -t_2\} \\ &= \{(t_1, t_2) \in \mathbb{R}^{s+1} : \Phi(t_1) \leq -t_2\} \end{aligned}$$

and

$$\begin{aligned} \tilde{\Phi}^{-1}(\{1\}) &= \tilde{\Phi}^{-1}(\{0, 1\}) \setminus \tilde{\Phi}^{-1}(\{0\}) \\ &= \left\{ (t_1, t_2) \in \mathbb{R}^{s+1} : \exists (y, y', \theta) \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'} \times \{0, 1\} \quad Wy + W'y' = t_1, \right. \\ &\quad \left. q^T y + q'^T y' \leq -t_2 + (\mathbb{M}_x - \varphi_o)\theta \right\} \\ &\quad \cap \left\{ (t_1, t_2) \in \mathbb{R}^{s+1} : \Phi(t_1) > -t_2 \right\} \\ &= \left[ \left\{ (t_1, t_2) \in \mathbb{R}^{s+1} : \exists (y, y') \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'} \quad Wy + W'y' = t_1, \quad q^T y + q'^T y' \leq -t_2 \right\} \right. \\ &\quad \left. \cup \left\{ (t_1, t_2) \in \mathbb{R}^{s+1} : \exists (y, y') \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'} \quad Wy + W'y' = t_1, \right. \right. \\ &\quad \left. \left. q^T y + q'^T y' \leq -t_2 + \mathbb{M}_x - \varphi_o \right\} \right] \\ &\quad \cap \left\{ (t_1, t_2) \in \mathbb{R}^{s+1} : \Phi(t_1) > -t_2 \right\} \\ &= \{(t_1, t_2) \in \mathbb{R}^{s+1} : \Phi(t_1) \leq -t_2 + \mathbb{M}_x - \varphi_o\} \cap \{(t_1, t_2) \in \mathbb{R}^{s+1} : \Phi(t_1) > -t_2\} \\ &= \{(t_1, t_2) \in \mathbb{R}^{s+1} : -t_2 < \Phi(t_1) \leq -t_2 + \mathbb{M}_x - \varphi_o\}. \end{aligned}$$

In view of Proposition 3.1(i) the function  $\Phi$  is lower semicontinuous and hence measurable. The above representations then yield measurability of the sets  $\tilde{\Phi}^{-1}(\{0\})$  and  $\tilde{\Phi}^{-1}(\{1\})$ . Note that in case  $\mathbb{M}_x - \varphi_o \leq 0$  we have  $\tilde{\Phi}^{-1}(\{1\}) = \emptyset$ . Since the integrand in (26) is globally bounded on its domain of finiteness, now measurability of  $\tilde{\Phi}$  implies that the integral in (26) is well-defined.

To check the asserted equality in (26) we denote by  $\chi_{M(x)}(h)$  the indicator function of  $M(x)$  and we show that  $\tilde{\Phi}(h - Tx, c^T x - \varphi_o) = \chi_{M(x)}(h)$  for all  $h \in \text{supp } \mu$ .

If  $\tilde{\Phi}(h - Tx, c^T x - \varphi_o) = 0$  then there exists  $(y, y') \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'}$  fulfilling  $Wy + W'y' = h - Tx$  and  $c^T x + q^T y + q'^T y' \leq \varphi_o$ . Hence  $c^T x + \Phi(h - Tx) \leq \varphi_o$ , implying  $h \notin M(x)$ , and we have  $\chi_{M(x)}(h) = 0$ .

If  $\tilde{\Phi}(h - Tx, c^T x - \varphi_o) = 1$  then  $c^T x + q^T y + q'^T y' > \varphi_o$  for all  $(y, y') \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'}$  fulfilling

$Wy + W'y' = h - Tx$ . Since the optimization problem associated with  $\Phi(h - Tx)$  is solvable, it follows that  $c^T x + \Phi(h - Tx) > \varphi_o$ . Therefore  $h \in M(x)$ , and we obtain  $\chi_{M(x)}(h) = 1$ . This completes the proof of (i).

To verify (ii) we observe that the estimate (28) yields a uniform upper bound  $\mathbf{M}$  for  $\sup\{c^T x + \Phi(h - Tx) : h \in \text{supp } \mu, x \in X\}$ , provided that  $\text{supp } \mu$  and  $X$  are bounded. Equivalence of the listed stochastic programs then is a direct consequence of (i).  $\square$

**Remark 4.2** *As a particular consequence of Proposition 4.1 we obtain that the stochastic program  $\min\{\tilde{Q}_{\mathbb{E}}(x) : x \in X\}$  has relative complete mixed-integer recourse, meaning that for any  $h \in \text{supp } \mu$  and any  $x \in X$  there exists a feasible tuple  $(y, y', \theta)$  to the optimization problem associated with  $\tilde{\Phi}(h - Tx, c^T x - \varphi_o)$ . On the other hand, the stochastic program fails to have complete mixed-integer recourse. Namely, if we fix  $t_1 \in \mathbb{R}^s$ , consider a sequence  $(t_2^n)_{n \in \mathbb{N}}$  tending to  $+\infty$ , and assume that there exist  $(y_n, y'_n, \theta_n) \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'}$  such that*

$$Wy_n + W'y'_n = t_1 \quad \text{and} \quad -q^T y_n - q'^T y'_n + (\mathbf{M} - \varphi_o)\theta_n \geq t_2^n,$$

*then we have found  $(y_n, y'_n) \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'}$  for which  $Wy_n + W'y'_n = t_1$  and  $q^T y_n + q'^T y'_n \rightarrow -\infty$  as  $n \rightarrow \infty$ . This contradicts the fact that  $\min\{q^T y + q'^T y' : Wy + W'y' = t_1, (y, y') \in \mathbb{Z}_+^{\tilde{m}} \times \mathbb{R}_+^{m'}\}$  is solvable, or, in other words, that  $\Phi(t_1) \in \mathbb{R}$ , cf. Proposition 3.1(i).*

*This lack of complete mixed-integer recourse prevented the application of existing results on structure and stability of expectation-based two-stage stochastic integer programs (see [46, 47]) in our analysis of Section 3. In addition, of course, Proposition 4.1 requires the underlying probability measures to have bounded support while the analysis of Section 3 does not.*

In the remainder of this section we assume that the set  $X$  is bounded and closed, and arises as a solution set to a system of linear inequalities, possibly involving integer requirements to components of  $x$ . Moreover, we assume that the underlying probability measure  $\mu$  is discrete with finitely many realizations (or scenarios)  $h_j$  and probabilities  $\pi_j, j = 1, \dots, J$ . Clearly, the support of  $\mu$  is bounded then, and we obtain the following corollary to Proposition 4.1.

**Corollary 4.3** *Adopt the setting of Proposition 4.1 and let  $X, \mu$  be as above. Then there exists a constant  $\mathbf{M} > 0$  such that the stochastic program*

$$\min\{Q_{\mathbb{P}}(x) : x \in X\} \tag{29}$$

*can be equivalently restated as*

$$\begin{aligned} \min_{x, y, y', \theta} \left\{ \sum_{j=1}^J \pi_j \theta_j \quad : \quad Wy_j + W'y'_j = h_j - Tx, \quad -q^T y_j - q'^T y'_j + (\mathbf{M} - \varphi_o)\theta_j \geq c^T x - \varphi_o, \right. \\ \left. x \in X, \quad y_j \in \mathbb{Z}_+^{\tilde{m}}, \quad y'_j \in \mathbb{R}_+^{m'}, \quad \theta_j \in \{0, 1\}, \quad j = 1, \dots, J \right\}. \end{aligned} \tag{30}$$

Problem (30) quickly becomes large-scale such that general-purpose mixed-integer linear programming algorithms and software fail. On the other hand, the constraint matrix of (30) obeys the same block-angular structure as with traditional expectation-based linear two-stage stochastic programs. Second-stage variables  $(y_j, y'_j, \theta_j)$  for different scenarios are not linked in explicit constraints but only through the scenario-independent first stage variable  $x$ .

In analogy to the traditional expectation-based model, cf. [13], this suggests the following algorithmic approach to (30) via scenario decomposition, i.e., Lagrangian relaxation of nonanticipativity.

Introduce in (30) copies  $x_j, j = 1, \dots, J$ , referring to the number of scenarios, and add the nonanticipativity constraints  $x_1 = \dots = x_J$  (or an equivalent system), for which we use the notation  $\sum_{j=1}^J H_j x_j = 0$  with proper  $(l, m)$ -matrices  $H_j, j = 1, \dots, J$ . Problem (30) then becomes

$$\begin{aligned} \min_{x, y, y', \theta} \left\{ \sum_{j=1}^J \pi_j \theta_j \quad : \quad Wy_j + W'y'_j = h_j - Tx_j, \quad -q^T y_j - q'^T y'_j + (\mathbf{M} - \varphi_o)\theta_j \geq c^T x_j - \varphi_o, \right. \\ \left. x_j \in X, \quad y_j \in \mathbb{Z}_+^{\tilde{m}}, \quad y'_j \in \mathbb{R}_+^{m'}, \quad \theta_j \in \{0, 1\}, \quad j = 1, \dots, J, \quad \sum_{j=1}^J H_j x_j = 0 \right\}. \end{aligned} \tag{31}$$

The constraint system of (31) can be decoupled by Lagrangian relaxation of  $\sum_{j=1}^J H_j x_j = 0$ . To this end, we consider for  $\lambda \in \mathbb{R}^l$  the functions

$$L_j(x_j, y_j, y'_j, \theta_j, \lambda) := \pi_j \theta_j + \lambda^T H_j x_j, \quad j = 1, \dots, J,$$

and form the Lagrangian

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

The Lagrangian dual of (31) reads

$$\max\{D(\lambda) : \lambda \in \mathbb{R}^l\} \tag{32}$$

where

$$D(\lambda) = \min\left\{\sum_{j=1}^J L_j(x_j, y_j, y'_j, \theta_j, \lambda) : \begin{aligned} &W y_j + W' y'_j = h_j - T x_j, \\ &-q^T y_j - q'^T y'_j + (\mathbf{M} - \varphi_o) \theta_j \geq c^T x_j - \varphi_o, \\ &x_j \in X, \quad y_j \in \mathbb{Z}_+^{\bar{m}}, \quad y'_j \in \mathbb{R}_+^{m'}, \quad \theta_j \in \{0, 1\}, \quad j = 1, \dots, J. \end{aligned}\right.$$

Separability yields

$$D(\lambda) = \sum_{j=1}^J D_j(\lambda) \tag{33}$$

where

$$D_j(\lambda) = \min\{L_j(x_j, y_j, y'_j, \theta_j, \lambda) : \begin{aligned} &W y_j + W' y'_j = h_j - T x_j, \\ &-q^T y_j - q'^T y'_j + (\mathbf{M} - \varphi_o) \theta_j \geq c^T x_j - \varphi_o, \\ &x_j \in X, \quad y_j \in \mathbb{Z}_+^{\bar{m}}, \quad y'_j \in \mathbb{R}_+^{m'}, \quad \theta_j \in \{0, 1\}. \end{aligned}\} \tag{34}$$

$D(\lambda)$  is the pointwise minimum of affine functions in  $\lambda$ . Therefore it is piecewise affine and concave. Thus, (32) is a nonsmooth concave maximization (or convex minimization) problem that can be solved by bundle methods from nondifferentiable optimization, for instance by the conic bundle method of [17] or the proximal bundle method of [22, 23]. At each iteration, these methods require the objective value and one subgradient of  $D$ . The structure of  $D$ , cf. (33), enables substantial decomposition, since the single-scenario problems (34) 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 (32) provides a lower bound to the optimal value  $z$  of problem (30). From integer programming ([30]) 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 (30). 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 ( $x_1 = \dots = x_N$ ), ideas for such heuristics come up straightforwardly. For example, examine the  $x_j$ -components,  $j = 1, \dots, J$ , of solutions to (34) for optimal or nearly optimal  $\lambda$ , and decide for the most frequent value arising or average and round if necessary.

If the heuristic yields a feasible solution to (30), 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 for (29) seen as a nonconvex global optimization problem. Let  $\mathbf{P}$  denote the list of current problems and  $z_{LD} = z_{LD}(P)$  the Lagrangian lower bound for  $P \in \mathbf{P}$ . The algorithm then proceeds as follows.

#### Algorithm 4.4

**Step 1 (Initialization):** Set  $\bar{z} = +\infty$  and let  $\mathbf{P}$  consist of problem (31).

**Step 2 (Termination):** If  $\mathbf{P} = \emptyset$  then the solution  $\hat{x}$  that yielded  $\bar{z} = Q_{\mathbf{P}}(\hat{x})$  is optimal.

**Step 3 (Node selection):** Select and delete a problem  $P$  from  $\mathbf{P}$  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  $x_j$ ,  $j = 1, \dots, J$ , are identical: If  $Q_{\mathcal{P}}(x_j) < \bar{z}$  then let  $\bar{z} = Q_{\mathcal{P}}(x_j)$  and delete from  $\mathbf{P}$  all problems  $P'$  with  $z_{LD}(P') \geq \bar{z}$ . Go to Step 2.
2. The scenario solutions  $x_j$ ,  $j = 1, \dots, J$  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  $Q_{\mathcal{P}}(\bar{x}^R) < \bar{z}$  then let  $\bar{z} = Q_{\mathcal{P}}(\bar{x}^R)$  and delete from  $\mathbf{P}$  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  $\mathbf{P}$  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 is obviously finite if all  $x$ -components have to be integers. (Recall that  $X$  is bounded !) If  $x$  is mixed-integer some stopping criterion to avoid endless branching on the continuous components has to be employed.

As already mentioned, the algorithm follows the same lines as the algorithm for  $\min\{Q_{\mathcal{E}}(x) : x \in X\}$  developed in [13]. In a straightforward manner this leads to a scenario decomposition algorithm for the mean-risk model  $\min\{Q_{\mathcal{E}}(x) + \rho Q_{\mathcal{P}}(x) : x \in X\}$ . At the end of the present section we will report some initial computational experience with this algorithm.

**Relations with Efficient Points in Multiobjective Optimization** The mean-risk model (6) can be seen as a scalarization of the multiobjective optimization problem

$$\min \{ (Q_{\mathcal{E}}(x), Q_{\mathcal{P}}(x)) : x \in X \}. \quad (35)$$

A common notion of optimality in multiobjective optimization is efficiency (or nondominance). In terms of (35) a point  $x^* \in X$  is called efficient if there is no  $x \in X$  fulfilling  $Q_{\mathcal{E}}(x) \leq Q_{\mathcal{E}}(x^*)$  and  $Q_{\mathcal{P}}(x) \leq Q_{\mathcal{P}}(x^*)$ , with at least one strict inequality. For basic facts of multiobjective optimization we refer to [5, 25] and the references therein. Given  $\rho_1, \rho_2 \in \mathbb{R}_+$ , every optimal solution to

$$\min\{\rho_1 Q_{\mathcal{E}}(x) + \rho_2 Q_{\mathcal{P}}(x) : x \in X\} \quad (36)$$

is efficient. This result enables computation of efficient points by solving scalar optimization problems. In general, only a subset of the efficient points of (35) can be computed via (36). For computing the full efficiency set via (36), additional assumptions, e.g., convexity of the individual objectives and the feasible set, are needed. Although we did not elaborate this in Section 3, it is quite easy to confirm that neither  $Q_{\mathcal{E}}$  nor  $Q_{\mathcal{P}}$  are convex in general. The following example demonstrates that, indeed, there exist efficient points for (35) not computable as solutions to (36) for any  $\rho_1, \rho_2 \in \mathbb{R}_+$ .

**Example 4.5** We specify (1) and (2) as follows. Let  $m = 1, \bar{m} = 1, m' = 2, s = 1$ , and  $c = 0, T = -1, X = \{\frac{4}{12}, \frac{6}{12}, \frac{7}{12}\}$ . The random variable  $h(\omega)$  is given by the realizations  $0, \frac{4}{12}, \frac{6}{12}$  with the probabilities  $\frac{2}{5}, \frac{2}{5}, \frac{1}{5}$ . The second stage is defined by

$$\begin{aligned} \Phi(t) &= \min\{\frac{1}{2}y_1 + y'_1 + y'_2 : y_1 + y'_1 - y'_2 = t, y_1 \in \mathbb{Z}_+, (y'_1, y'_2) \in \mathbb{R}_+^2\} \\ &= \min\{\frac{1}{2}y_1 + |t - y_1| : y_1 \in \mathbb{Z}_+\}. \end{aligned}$$

Finally, the probability threshold is selected as  $\varphi_o = \frac{7}{12}$ . We have just three points in  $X$  such that the image set  $(Q_{\mathcal{E}}, Q_{\mathcal{P}})(X)$  can be computed explicitly. It holds

$$(Q_{\mathcal{E}}, Q_{\mathcal{P}})(X) = \left\{ \left(\frac{35}{60}, 0\right), \left(\frac{34}{60}, \frac{2}{5}\right), \left(\frac{32}{60}, \frac{3}{5}\right) \right\}.$$

Clearly, all three members of the image set are efficient. One confirms that the point  $(\frac{33}{60}, \frac{2}{5})$  is located at the straight line passing through  $(\frac{35}{60}, 0)$  and  $(\frac{32}{60}, \frac{2}{5})$ . Hence there is no straight line supporting (from below)  $(Q_{\mathcal{E}}, Q_{\mathcal{P}})(X)$  and passing through  $(\frac{34}{60}, \frac{2}{5})$ . In other words, the efficient point  $(\frac{34}{60}, \frac{2}{5})$  is not computable as a solution to (36) for any  $\rho_1, \rho_2 \in \mathbb{R}_+$ .



A prominent example of efficiency in the context of mean-risk models is induced by Markowitz' mean-variance model for portfolio selection, [28, 44]. The model aims at finding an optimal asset allocation where the quality of the allocation is judged by both expectation and variance of the return. The total return being the sum of individual returns multiplied by the allocation proportions, both expectation and variance of the return are convex functions of the allocation. Hence the full set of efficient points, also called the efficient frontier, can be traced by solving scalarizations as in (36).

Due to lacking convexity and the above example we cannot hope to be able to trace the full efficient set, or efficient frontier, of (35) by solving scalarizations (36). However, Algorithm 4.4 bears the potential of tracing the supported part of the efficient frontier, i.e., those efficient points that arise as optimal solutions to scalarizations (36). To this end, it is sufficient to vary the parameter  $\rho$  in (6) within the nonnegative reals. For every individual  $\rho$ , Algorithm 4.4 then provides a global solution to the nonconvex optimization problem (6). The numerics of tracing nonsupported parts of efficient frontiers to nonconvex multiobjective optimization problems still is a widely open field, cf. [25] for an account of existing methods.

**Modeling Background for Computational Tests** To illustrate our initial computational experience we will report tracing of supported efficient points at an example from chemical engineering. The modeling background is given by a real-life multi-product batch plant producing expandable polystyrene (EPS). A detailed description of the EPS production process can be found in [16].

The process consists of preparation, polymerization, and finishing. During preparation different kinds of intermediates are produced. In certain mixtures depending on a finite number of recipes the intermediates are fed batch-wise into the polymerization reactors. After termination of each polymerization its product is transferred immediately into a mixing tank of a finishing line, leading to a discontinuous inflow into these tanks. Each finishing line further consists of a separation stage where different grain sizes of EPS are separated from each other. These grain sizes are the final product of the process and have to match customer demand. The separation stages are driven continuously. Shut-down and start-up procedures for separation stages are time consuming, expensive, and have to adhere to minimum up- and down-times of the stages.

The EPS process is controlled by fixing starting times and choices of recipes for the polymerizations and by selecting start-up and shut-down times as well as feed rates for the separation stages. A typical scheduling horizon is given by two weeks, with a time discretization into five equidistant intervals. The major source of uncertainty is customer demand. Optimization aims at minimizing a weighted sum of costs caused by running the polymerizations, switching the separation stages, and compensating deficit between production and customer demand.

The above setting gives rise to different two-stage stochastic integer programs, see [16] for details. For the numerical tests in the present paper we have used a planning model where the first-stage variables are given by the states of the separation stages. This places emphasis on the qualitative aspect that a smooth operation of the EPS process is desired, which is achieved by fixing the states of the most sensible part of the plant as early as possible.

**Tracing of Supported Efficient Points** To trace the supported parts of nonconvex, nonconnected efficient frontiers we have formulated instances of the EPS problem with 10, 20, 50, and 100 scenarios. With the mentioned extension of Algorithm 4.4 we then have solved to global optimality instances of the mean-risk model  $\min\{Q_E(x) + \rho Q_P(x) : x \in X\}$  for suitable values of  $\rho$ .

Tracing starts with  $\rho = 0$ , i.e., with solving  $\min\{Q_E(x) : x \in X\}$ . If the optimal solution is unique then it has to be efficient as well. Since we have no indication about unicity of optimal solutions, we solve the mean-risk model again with "small"  $\rho$ , say  $\underline{\rho} = 0.001$ . If the  $Q_E$ -value of the optimal solution remains the same, the optimal solution has to be efficient, and there are no further supported efficient points for  $0 < \rho < \underline{\rho}$ .

An analogous procedure is carried out at the "upper end". We solve  $\min\{Q_P(x) : x \in X\}$  and check with a "big"  $\rho$  ( $\bar{\rho} = 1000$ ) for efficiency.

Suppose the "lower" and "upper end" procedures have resulted in two distinct efficient points  $x', x''$  with distinct values of  $(Q_E, Q_P)(x)$ . We calculate the normal vector  $(1, \hat{\rho})^T$  of the straight line passing through  $(Q_E, Q_P)(x')$  and  $(Q_E, Q_P)(x'')$  and solve  $\min\{Q_E(x) + \hat{\rho}Q_P(x) : x \in X\}$ . The optimal solution  $x'''$  is a supported efficient point. If  $Q_E(x''') + \hat{\rho}Q_P(x''')$  coincides with the identical values  $Q_E(x') + \hat{\rho}Q_P(x') = Q_E(x'') + \hat{\rho}Q_P(x'')$  then it is clear that, up to equality of the value  $Q_E(x) + \hat{\rho}Q_P(x)$ ,

there are no further supported efficient points for  $\underline{\rho} < \rho < \bar{\rho}$ . Otherwise, the search continues with the intervals  $\underline{\rho} \leq \rho \leq \hat{\rho}$  and  $\hat{\rho} \leq \rho \leq \bar{\rho}$ .

The procedure is iterated at subintervals for  $\rho$  where further supported efficient points still may be expected. It terminates when no such intervals exist any longer. With a discrete probability distribution,  $Q_{IP}$  attains only finitely many values, implying that the procedure terminates after finitely many steps. Table 1 documents our computations. The parameter  $M$ , cf. (30), was put to 1000 in all instances. The threshold values  $\varphi_o$  are 102.52, 102.55, 102.07, and 102.39 for the 10-, 20-, 50-, and 100-scenario instances, respectively.

Table 1: Computational results for the EPS problem

Scenarios	Cont/Int/Bin	Constraints	$\rho$	$(Q_E, Q_{IP})$	Time (h:mm)	CPLEX
10	500/400/352	1370	0	(114.81, 1.00)	0:11	3.23%
	500/400/362	1380	0.001	(114.81, 1.00)	0:12	3.16%
	500/400/362	1380	44.17	(141.31, 0.40)	0:22	14.17%
	500/400/362	1380	94.78	(141.31, 0.40)	0:21	7.22%
	500/400/362	1380	246.60	(141.31, 0.40)	0:19	11.77%
	500/400/362	1380	1000	(190.64, 0.20)	0:10	0.67%
	500/400/362	1380	$+\infty$	(302.52, 0.20)	0:01	$\infty$
20	1000/800/692	2740	0	(120.50, 1.00)	0:21	4.79%
	1000/800/712	2760	0.001	(120.50, 1.00)	0:21	5.51%
	1000/800/712	2760	66.57	(120.50, 1.00)	0:37	11.87%
	1000/800/712	2760	67.80	(137.14, 0.75)	0:38	12.24%
	1000/800/712	2760	68.68	(161.18, 0.40)	0:38	12.03%
	1000/800/712	2760	177.13	(161.18, 0.40)	0:37	9.68%
	1000/800/712	2760	833.08	(161.18, 0.40)	0:37	23.56%
	1000/800/712	2760	1000	(244.49, 0.30)	0:34	7.18%
1000/800/712	2760	$+\infty$	(402.55, 0.30)	0:01	$\infty$	
50	2500/2000/1712	6850	0	(124.00, 1.00)	0:59	5.77%
	2500/2000/1762	6900	0.001	(124.00, 1.00)	1:00	6.23%
	2500/2000/1762	6900	111.25	(166.27, 0.62)	1:53	29.73%
	2500/2000/1762	6900	222.05	(166.27, 0.62)	1:40	28.03%
	2500/2000/1762	6900	413.43	(257.23, 0.40)	1:48	22.80%
	2500/2000/1762	6900	1000	(257.23, 0.40)	1:04	24.86%
	2500/2000/1762	6900	$+\infty$	(502.07, 0.40)	0:09	$\infty$
100	5000/4000/3412	13700	0	(122.10, 1.00)	2:21	5.98%
	5000/4000/3512	13800	0.001	(122.10, 1.00)	2:14	13.40%
	5000/4000/3512	13800	80.33	(153.43, 0.61)	3:27	31.81%
	5000/4000/3512	13800	153.57	(153.43, 0.61)	3:09	33.11%
	5000/4000/3512	13800	263.42	(221.92, 0.35)	3:18	$\infty$
	5000/4000/3512	13800	1000	(221.92, 0.35)	1:45	$\infty$
	5000/4000/3512	13800	$+\infty$	(452.39, 0.35)	0:22	$\infty$

The  $\rho$ -column displays the values that were necessary for the search. Let us explain at the 10-scenario instance:  $\rho = 0$  and  $\rho = +\infty$  correspond to minimizing  $Q_E$  and  $Q_{IP}$ , respectively. The values  $\underline{\rho} = 0.001$  and  $\bar{\rho} = 1000$  are the mentioned safeguards for efficiency at the “lower” and “upper ends”. Simultaneously, they serve to initialize the search interval for  $\rho$ . The first value for  $\hat{\rho}$  is 94.78. It yields an efficient point whose value of  $Q_E(x) + \hat{\rho}Q_{IP}(x)$  is distinct from the corresponding value of the two efficient points already found. Hence, the subintervals  $[\underline{\rho}, \hat{\rho}]$  and  $[\hat{\rho}, \bar{\rho}]$  must be considered further, what is done with the updated  $\hat{\rho}$ -values 44.17 and 246.60, respectively. In both cases, the optimal values of  $Q_E(x) + \hat{\rho}Q_{IP}(x)$  coincide with  $(Q_E + \hat{\rho}Q_{IP})$ -values of efficient points already known. (In fact, even the optimal solution points coincide with an efficient point already known.) The search terminates. Altogether, we have found three supported efficient points with  $(Q_E, Q_{IP})$ -values of (114.81, 1.00), (141.31, 0.40), and (190.64, 0.20). Up to equality of values  $Q_E(x) + \hat{\rho}Q_{IP}(x)$  for  $\hat{\rho} \in \{44.17, 246.60\}$ , these are all supported efficient points of the instance.

The Time-column displays the time needed by our decomposition algorithm to find an optimal point and prove its optimality with a relative gap of 0.001%.

The CPLEX-column shows the relative optimality gaps achieved by CPLEX 8.0 (with default parameters) in the time listed in the column before. The symbol  $\infty$  indicates that no feasible solution was found in this time. All computations were carried out on a Linux PC with an AMD Athlon XP 2200+ processor (1.8 GHz) and 1 GB RAM.

The Cont/Int/Bin-column has the numbers of continuous, integer (nonbinary), and binary variables in the block-angular mixed-integer linear programs corresponding to the models. The numbers of constraints are listed in the next column.

Altogether, the table confirms that our algorithm is able to trace with reasonable effort supported parts of efficient frontiers of realistic instances of the nonconvex multi-objective program (35).

**Acknowledgment.** We wish to thank Christoph Helmberg (Technical University of Chemnitz) for giving us access to the implementation of his conic bundle method. Further thanks are due to Andreas Märkert (Gerhard-Mercator University Duisburg) for stimulating discussions and his support in designing and performing the computational tests. Moreover, we are grateful to two anonymous referees and the Editorial Board Member for their comments that helped to improve the final version of our paper.

## References

- [1] Artstein, Z.; Wets, R.J-B: Stability results for stochastic programs and sensors, allowing for discontinuous objective functions, *SIAM Journal on Optimization* 4 (1994), 537-550.
- [2] Artzner, P.; Delbaen, F.; Eber, J.-M.; Heath, D.: Coherent measures of risk, *Mathematical Finance* 9 (1999), 203-228.
- [3] Bank, B.; Guddat, J.; Klatte, D.; Kummer, B.; Tammer, K.: *Non-linear Parametric Optimization*, Akademie-Verlag, Berlin, 1982.
- [4] Bank, B.; Mandel, R.: *Parametric Integer Optimization*, Akademie-Verlag, Berlin, 1988.
- [5] Benson, H.P.: Multi-objective optimization: Pareto optimal solutions, properties, *Encyclopedia of Optimization*, Vol. III (C.A. Floudas, P.M. Pardalos, Eds.), Kluwer, Dordrecht, 2001, 489-493.
- [6] Boreanu, B.: Minimum risk criterion in stochastic optimization, *Economic Computation and Economic Cybernetics Studies and Research* 2 (1981), 31-39.
- [7] Berge, C.: *Topological Spaces*, Macmillan, New York, 1963.
- [8] Bhattacharya, R.N.; Ranga Rao, R.: *Normal Approximation and Asymptotic Expansions*, Wiley, New York, 1976.
- [9] Billingsley, P.: *Convergence of Probability Measures*, Wiley, New York, 1968.
- [10] Birge, J.R.; Louveaux, F.: *Introduction to Stochastic Programming*, Springer, New York, 1997.
- [11] Blair, C.E.; Jeroslow, R.G.: The value function of a mixed integer program: I, *Discrete Mathematics* 19 (1977), 121-138.
- [12] Buck, R.C.: Partition of space, *American Mathematical Monthly* 50 (1943), 541-544.
- [13] Carøe, C.C.; Schultz, R.: Dual decomposition in stochastic integer programming, *Operations Research Letters* 24 (1999), 37-45.
- [14] Using the CPLEX Callable Library, CPLEX Optimization, Inc. 2001.
- [15] Dudley, R.M.: *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove, 1989.
- [16] Engell, S.; Märkert, A.; Sand, G.; Schultz, R.; Schulz, Ch.: Online scheduling of multiproduct batch plants under uncertainty, *Online Optimization of Large Scale Systems* (M. Grötschel, S.O. Krumke, J. Rambau, Eds.), Springer, Berlin, 2001, 649-676.

- [17] Helmberg, C.; Kiwiel, K.C.: A spectral bundle method with bounds, *Mathematical Programming* (to appear).
- [18] Hemmecke, R.; Schultz, R.: Decomposition of test sets in stochastic integer programming, *Mathematical Programming* (to appear).
- [19] Kall, P.; Wallace, S.W.: *Stochastic Programming*, Wiley, Chichester, 1994.
- [20] Kibzun, A.I.; Kan, Y.S.: *Stochastic Programming Problems with Probability and Quantile Functions*, Wiley, Chichester, 1996.
- [21] King, A.J.; Takriti, S.; Ahmed, S.: Issues in risk modeling for multi-stage systems, *IBM Research Report*, RC-20993, Yorktown Heights, 1997.
- [22] Kiwiel, K. C.: Proximity control in bundle methods for convex nondifferentiable optimization, *Mathematical Programming* 46 (1990), 105-122.
- [23] 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.
- [24] Klatte, D.: On the stability of local and global optimal solutions in parametric problems of nonlinear programming, Part I and Part II, *Seminarbericht Nr. 75, Sektion Mathematik, Humboldt-Universität zu Berlin*, Berlin, 1985, 1-39.
- [25] Korhonen, P.: Multiple objective programming support, *Encyclopedia of Optimization*, Vol. III (C.A. Floudas, P.M. Pardalos, Eds.), Kluwer, Dordrecht, 2001, 566-574.
- [26] Kuelbs, J.; Dudley, R.M.: Log log laws for empirical measures, *Annals of Probability* 8 (1980), 405-418.
- [27] Märkert, A.; Schultz, R.: Variance and two-stage stochastic programs, manuscript in preparation.
- [28] Markowitz, H.M.: Portfolio selection, *Journal of Finance* 7 (1952), 77-91.
- [29] Mulvey, J.M.; Vanderbei, R.J.; Zenios, S.A.: Robust optimization of large-scale systems, *Operations Research* 43 (1995), 264-281.
- [30] Nemhauser, G.L.; Wolsey, L.A.: *Integer and Combinatorial Optimization*, Wiley, New York 1988.
- [31] Ogryczak, W.; Ruszczyński, A.: From stochastic dominance to mean-risk models: Semideviations as risk measures, *European Journal of Operational Research* 116 (1999), 33-50.
- [32] Ogryczak, W.; Ruszczyński, A.: Dual stochastic dominance and related mean-risk models, *SIAM Journal on Optimization* 13 (2002), 60-78.
- [33] Orlik, P.; Terao, H.: *Arrangements of Hyperplanes*, Springer, Berlin, 1992.
- [34] Pollard, D.: *Convergence of Stochastic Processes*, Springer, New York, 1984.
- [35] Prékopa, A.: *Stochastic Programming*, Kluwer, Dordrecht, 1995.
- [36] Rachev, S.T.; Römisch, W.: Quantitative stability in stochastic programming: the method of probability metrics, *Mathematics of Operations Research* (to appear).
- [37] Raik, E.: Qualitative research into the stochastic nonlinear programming problems, *Eesti NSV Teaduste Akademia Toimetised / Füüsika, Matemaatika (News of the Estonian Academy of Sciences / Physics, Mathematics)* 20 (1971), 8-14. In Russian.
- [38] Raik, E.: On the stochastic programming problem with the probability and quantile functionals, *Eesti NSV Teaduste Akademia Toimetised / Füüsika, Matemaatika (News of the Estonian Academy of Sciences / Physics, Mathematics)* 21 (1971), 142-148. In Russian.
- [39] Riis, M.; Schultz, R.: Applying the minimum risk criterion in stochastic recourse programs, *Computational Optimization and Applications* 24 (2003), 267-287.

- [40] Robinson, S.M.: Local epi-continuity and local optimization, *Mathematical Programming* 37 (1987), 208-222.
- [41] Rockafellar, R.T.; Wets, R.J-B.: *Variational Analysis*, Springer, Berlin, 1998.
- [42] Römisch, W.; Schultz, R.: Stability analysis for stochastic programs, *Annals of Operations Research* 30 (1991), 241-266.
- [43] 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, Berlin, 2001, 581-600.
- [44] Ruszczyński, A.; Vanderbei, R.J.: *Frontiers of stochastically nondominated portfolios*, Research Report ORFE 1-2002, Department of Operations Research and Financial Engineering, Princeton University, 2002.
- [45] Schultz, R.: Continuity properties of expectation functions in stochastic integer programming, *Mathematics of Operations Research* 18 (1993), 578-589.
- [46] Schultz, R.: On structure and stability in stochastic programs with random technology matrix and complete integer recourse, *Mathematical Programming* 70 (1995), 73-89.
- [47] Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse, *SIAM Journal on Optimization* 6 (1996), 1138-1152.
- [48] Shorack, G.R.; Wellner, J.A.: *Empirical Processes with Applications to Statistics*, Wiley, New York, 1986.
- [49] Takriti, S.; Ahmed, S.: On robust optimization of two-stage systems, *Mathematical Programming* (to appear).