

# On Deviation Measures in Stochastic Integer Programming

Andreas Märkert\* and Rüdiger Schultz†

Institute of Mathematics

University of Duisburg-Essen, Campus Duisburg

Lotharstr. 65, D-47048 Duisburg, Germany

May 10, 2004

## Abstract

We propose extensions of traditional expectation-based stochastic integer programs to mean-risk models. Risk is measured by expected deviations of suitable random variables from their means or from preselected targets. We derive structural properties of the resulting stochastic programs and present first algorithmic ideas to achieve problem decomposition.

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

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

## 1 Introduction

This paper deals with mean-risk extensions of the following two-stage mixed-integer linear stochastic program:

$$\min\{Q_{\mathcal{E}}(x, \mu) : x \in X\} \quad (1)$$

where

$$Q_{\mathcal{E}}(x, \mu) := \int_{\mathbb{R}^s} (c^\top x + \Phi(z - Tx)) \mu(dz) \quad (2)$$

and

$$\Phi(t) := \min\{q^\top y + q'^\top y' : Wy + W'y' = t, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'}\}. \quad (3)$$

All ingredients above are assumed to have conformable dimensions. Moreover,  $W, W'$  are rational matrices, and  $X \subseteq \mathbb{R}^m$  is a nonempty polyhedron, possibly involving integer requirements to components of  $x$ . The probability measure  $\mu$  belongs to  $\mathcal{P}(\mathbb{R}^s)$ , the set of all Borel probability measures on  $\mathbb{R}^s$ . Dependence of  $Q_{\mathcal{E}}$  on both  $x$  and  $\mu$  is marked explicitly since, later on,  $Q_{\mathcal{E}}$  and related objects will be studied both as functions in  $x$  and jointly in  $(x, \mu)$ . For the moment, let  $\mu \in \mathcal{P}(\mathbb{R}^s)$  be fixed.

The following assumptions ensure that the above model is well-defined in the sense that  $Q_{\mathcal{E}}(x, \mu) \in \mathbb{R}$  for all  $x \in \mathbb{R}^m$ . For details see [22].

(A1) (complete recourse)  $W(\mathbb{Z}_+^{\bar{m}}) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$ ,

(A2) (sufficiently expensive recourse)  $\{u \in \mathbb{R}^s : W^\top u \leq q, W'^\top u \leq q'\} \neq \emptyset$ ,

(A3) (finite first moment)  $\int_{\mathbb{R}^s} \|z\| \mu(dz) < +\infty$ .

The model (1)-(3) arises from two-stage mixed-integer linear programs under uncertainty. Decision variables fall into two categories. The first-stage decision  $x$  has to be taken in a here-and-now manner, before knowing the outcome of the random data  $z = z(\omega)$ . The second-stage decision  $(y, y')$  is taken after  $x$  has been fixed and  $z(\omega)$  has been observed. Assuming that  $(y, y')$  is selected best possible, this two-stage decision process leads to a random cost value that can be expressed as  $f(x, z(\omega)) := c^\top x + \Phi(z(\omega) - Tx)$ .

---

\*maerkert@math.uni-duisburg.de

†schultz@math.uni-duisburg.de, corresponding author

For further basic details of two-stage stochastic programming we refer to the books [5, 11, 18, 21]. The problem of finding a best here-and-now decision  $x$  in the above setting can be seen as finding a best random variable in the family  $\{f(x, z(\omega)) : x \in X\}$ . In (1) the traditional approach in two-stage stochastic programming is reflected, namely, the random variables  $f(x, z)$  are ranked by their expectations with respect to  $\mu$ . The present paper goes beyond this setting by ranking the above random variables according to weighted sums of their means and suitable terms expressing risk. This leads to mean-risk extensions

$$\min\{Q_{\mathbb{E}}(x, \mu) + \alpha \cdot Q_{\mathcal{R}}(x, \mu) : x \in X\} \quad (4)$$

of the model (1) with fixed weight factor  $\alpha \geq 0$  and risk term  $Q_{\mathcal{R}}$ .

We will study three versions of (4) which are given by the following specifications:

the central deviation, where

$$Q_{\mathcal{R}}(x, \mu) := Q_{\mathcal{D}}(x, \mu) := \int_{\mathbb{R}^s} |f(x, z) - Q_{\mathbb{E}}(x, \mu)| \mu(dz) \quad \text{and} \quad 0 \leq \alpha \leq \frac{1}{2}, \quad (5)$$

the semideviation, where

$$Q_{\mathcal{R}}(x, \mu) := Q_{\mathcal{D}^+}(x, \mu) := \int_{\mathbb{R}^s} \max\{f(x, z) - Q_{\mathbb{E}}(x, \mu), 0\} \mu(dz) \quad \text{and} \quad 0 \leq \alpha \leq 1, \quad (6)$$

and the expected excess of a given target  $\eta \in \mathbb{R}$ , where

$$Q_{\mathcal{R}}(x, \mu) := Q_{\mathcal{D}^\eta}(x, \mu) := \int_{\mathbb{R}^s} \max\{f(x, z) - \eta, 0\} \mu(dz) \quad \text{and} \quad \alpha \geq 0. \quad (7)$$

We call the above quantities deviation measures since they are based on expected deviations of the random variable from its mean or from some preselected target. Our emphasis on deviation measures determined by *piecewise linear* operations (here, taking the absolute value or the maximum) is mainly motivated algorithmically. We will see that, thanks to the piecewise linearity, problem (4) can be tackled successfully by extensions of mixed-integer linear programming techniques provided the underlying measure  $\mu$  is discrete.

Another motivation for considering the above deviation measures rests in the consistency with stochastic dominance they induce. Stochastic dominance is an established notion of partial order for random variables, see [9, 13, 15, 16, 17] where different aspects of stochastic dominance as such or in relation with stochastic programming are covered. As an example let us consider second degree stochastic dominance. When preferring small outcomes to big ones, as we do in view of our minimization setting, a (real-valued) random variable  $X$  is said to dominate a random variable  $Y$  to second degree ( $X \succeq_2 Y$ ) if  $\mathbb{E}h(X) \leq \mathbb{E}h(Y)$  for all nondecreasing convex functions  $h$  for which both expectations exist. The mean-risk model (4) is said to be consistent with (second degree) stochastic dominance if the ranking its objective incurs on the random variables  $f(x, z)$  inherits a ranking possibly already existing with respect to (second degree) stochastic dominance, more precisely, if  $f(x_1, z) \succeq_2 f(x_2, z)$  implies  $Q_{\mathbb{E}}(x_1, \mu) + \alpha Q_{\mathcal{R}}(x_1, \mu) \leq Q_{\mathbb{E}}(x_2, \mu) + \alpha Q_{\mathcal{R}}(x_2, \mu)$ . The catch is that the specifications (5)-(7), with  $\alpha$  restricted to the given intervals, all make (4) consistent with second degree stochastic dominance. For details see [9, 14, 16].

The purpose of this paper is to study mathematical structures of (4) with the specifications (5)-(7), and to present some first algorithmic ideas for the stochastic integer programs arising.

## 2 Prerequisites

A key prerequisite for analyzing mathematical structures in the objectives of (4) is the following proposition about the mixed-integer value function (3). The proposition collects results dating back to [3, 6].

**Proposition 2.1** *Assume that (A1), (A2) hold. Then*

- (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  $i$ ;

(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  $W'(\mathbb{R}_+^m)$  and  $t_i, t_{ij}$  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  $t_1, t_2 \in \mathbb{R}^s$ .

Our investigations will be directed to continuity of the objectives of (4), both as functions in  $x$  and jointly in  $(x, \mu)$ . The latter is motivated by stability analysis of (4) under perturbations of  $\mu$ . In real-life cases, the probability measure that underlies a stochastic program often is subjective. Moreover, computing the multivariate integrals in (2), (5)-(7) may require approximation of the underlying probability measure by simpler ones. This motivates searching for sufficient conditions under which “small” perturbations of  $\mu$  lead to only “small” perturbations of the solutions to (4). For the latter, joint continuity of the objective of (4) in  $(x, \mu)$  is of crucial importance. For recent surveys on stability analysis in stochastic programming see [20, 23].

For studying the outlined continuity a notion of convergence on the set of probability measures  $\mathcal{P}(\mathbb{R}^s)$  is needed. The following notion of weak convergence of probability measures [4] is both sufficiently general to cover important 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  $h : \mathbb{R}^s \rightarrow \mathbb{R}$  it holds  $\int_{\mathbb{R}^s} h(z) \mu_n(dz) \rightarrow \int_{\mathbb{R}^s} h(z) \mu(dz)$  as  $n \rightarrow \infty$ . Let us also introduce the notation  $\Delta_{p,K}(\mathbb{R}^s) := \{\nu \in \mathcal{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \|z\|^p \nu(dz) \leq K\}$  where  $p > 1$  and  $K > 0$  are fixed constants. The structure of the function  $Q_{\mathbf{E}}(x, \mu)$  of (2) was studied in [22] from where we quote the following.

### Proposition 2.2

- (i) Assume (A1)-(A3). Then  $Q_{\mathbf{E}}(\cdot, \mu)$  is real-valued and lower semicontinuous on  $\mathbb{R}^m$ .
- (ii) Assume (A1)-(A3) and that  $\mu(E(x)) = 0$  where  $E(x) = \{z \in \mathbb{R}^s : \Phi \text{ is discontinuous at } z - Tx\}$ . Then  $Q_{\mathbf{E}}(\cdot, \mu)$  is continuous at  $x$ .
- (iii) Assume (A1), (A2). Let  $\mu \in \Delta_{p,K}(\mathbb{R}^s)$  for some  $p > 1$  and  $K > 0$ , and  $\mu(E(x)) = 0$ . Then  $Q_{\mathbf{E}} : \mathbb{R}^m \times \Delta_{p,K}(\mathbb{R}^s) \rightarrow \mathbb{R}$  is continuous at  $(x, \mu)$ .

Note that the assumption  $\mu(E(x)) = 0$ , which might look rather implicit at first sight, is always fulfilled if  $\mu$  has a density. Indeed, by Proposition 2.1(iii) discontinuities of  $\Phi$  can only occur in a countable union of hyperplanes, and the latter has  $\mu$ -measure zero provided  $\mu$  has a density.

## 3 Structural Properties

Straightforward calculations provide the following identities

$$\begin{aligned}
& Q_{\mathbf{E}}(x, \mu) + \alpha Q_{\mathcal{D}}(x, \mu) \\
&= \mathbb{E}f(x, z) + \alpha \mathbb{E}|f(x, z) - \mathbb{E}f(x, z)| \\
&= \mathbb{E}f(x, z) + \alpha \left( \mathbb{E} \max \{f(x, z) - \mathbb{E}f(x, z), 0\} + \mathbb{E} \max \{\mathbb{E}f(x, z) - f(x, z), 0\} \right) \\
&= \mathbb{E}f(x, z) + \alpha \left( \mathbb{E} \max \{f(x, z), \mathbb{E}f(x, z)\} - \mathbb{E}f(x, z) + \mathbb{E}(\max\{\mathbb{E}f(x, z), f(x, z)\} - f(x, z)) \right) \\
&= (1 - 2\alpha) \mathbb{E}f(x, z) + 2\alpha \mathbb{E} \max \{f(x, z), \mathbb{E}f(x, z)\}, \tag{8}
\end{aligned}$$

$$\begin{aligned}
& Q_{\mathbf{E}}(x, \mu) + \alpha Q_{\mathcal{D}^+}(x, \mu) \\
&= \mathbb{E}f(x, z) + \alpha \mathbb{E} \max \{f(x, z) - \mathbb{E}f(x, z), 0\} \\
&= (1 - \alpha) \mathbb{E}f(x, z) + \alpha \mathbb{E} \max \{f(x, z), \mathbb{E}f(x, z)\}, \tag{9}
\end{aligned}$$

$$\begin{aligned}
& Q_{\mathbf{E}}(x, \mu) + \alpha Q_{\mathcal{D}^\eta}(x, \mu) \\
&= \mathbb{E}f(x, z) + \alpha \mathbb{E} \max \{f(x, z) - \eta, 0\} \\
&= \mathbb{E}f(x, z) + \alpha \mathbb{E} \max \{f(x, z), \eta\} - \alpha \eta. \tag{10}
\end{aligned}$$

Hence, for  $\alpha$  in the intervals of (5)-(7), the resulting objective functions of (4) all essentially are nonnegative linear combinations of  $Q_{\mathbb{E}}(x, \mu)$  and the function

$$Q_{max}(x, \mu) := \int_{\mathbb{R}^s} \max\{f(x, z), Q_{\mathbb{E}}(x, \mu)\} \mu(dz), \quad (11)$$

or its close, and simpler to analyze, relative

$$Q_{max,\eta}(x, \mu) := \int_{\mathbb{R}^s} \max\{f(x, z), \eta\} \mu(dz). \quad (12)$$

Our structural investigations, hence, will focus on  $Q_{max}(x, \mu)$ , and then, in view of Proposition 2.2, the desired results for the objectives in (4) will follow immediately.

**Proposition 3.1** *Assume (A1)-(A3). Then  $Q_{max}(\cdot, \mu)$  is real-valued and lower semicontinuous on  $\mathbb{R}^m$ .*

**Proof:** The lower semicontinuity of  $\Phi$  (Proposition 2.1(i)) implies that for each  $x \in \mathbb{R}^m$  the integrand in (11) is a measurable function in  $z$ . Finiteness of the integral then follows from (A3) and the estimate

$$\begin{aligned} |\max\{f(x, z), Q_{\mathbb{E}}(x, \mu)\}| &\leq |f(x, z)| + |Q_{\mathbb{E}}(x, \mu)| \\ &\leq |\Phi(z - Tx) - \Phi(0)| + |c^\top x| + |Q_{\mathbb{E}}(x, \mu)| \\ &\leq \beta\|z\| + \beta\|Tx\| + \gamma + |c^\top x| + |Q_{\mathbb{E}}(x, \mu)|. \end{aligned}$$

In this estimate we have used that (A2) implies  $\Phi(0) = 0$ , and we have employed Proposition 2.1(iv). To show lower semicontinuity let  $x \in \mathbb{R}^m$  and  $x_n \rightarrow x$ . Our aim is to apply Fatou's Lemma to the functions  $h_n(z) := \max\{f(x_n, z), Q_{\mathbb{E}}(x_n, \mu)\}$ . Using the above arguments again we obtain an integrable minorant for these functions:

$$\max\{f(x_n, z), Q_{\mathbb{E}}(x_n, \mu)\} \geq f(x_n, z) \geq -(\beta\|z\| + \beta\|Tx_n\| + \gamma + |c^\top x_n|) \geq -\beta\|z\| + \kappa$$

where  $\kappa$  is a lower bound for the (convergent) sequence  $\{-(\beta\|Tx_n\| + \gamma + |c^\top x_n|)\}$ . Now the following holds

$$\begin{aligned} \liminf_{x_n \rightarrow x} Q_{max}(x_n, \mu) &\geq \int_{\mathbb{R}^s} \liminf_{x_n \rightarrow x} \max\{f(x_n, z), Q_{\mathbb{E}}(x_n, \mu)\} \mu(dz) \\ &\geq \int_{\mathbb{R}^s} \max\left\{\liminf_{x_n \rightarrow x} f(x_n, z), \liminf_{x_n \rightarrow x} Q_{\mathbb{E}}(x_n, \mu)\right\} \mu(dz) \\ &\geq \int_{\mathbb{R}^s} \max\{f(x, z), Q_{\mathbb{E}}(x, \mu)\} \mu(dz) = Q_{max}(x, \mu), \end{aligned}$$

establishing the desired lower semicontinuity. In the first estimate above we have used Fatou's Lemma, and in the third that  $\Phi$  as well as  $Q_{\mathbb{E}}(\cdot, \mu)$  are lower semicontinuous (Propositions 2.1(i) and 2.2(i)).  $\square$

**Proposition 3.2** *Assume (A1)-(A3) and that  $\mu(E(x)) = 0$  where  $E(x) = \{z \in \mathbb{R}^s : \Phi \text{ is discontinuous at } z - Tx\}$ . Then  $Q_{max}(\cdot, \mu)$  is continuous at  $x$ .*

**Proof:** Let  $x \in \mathbb{R}^m$  and  $x_n \rightarrow x$ . We apply Lebesgue's Dominated Convergence Theorem to the functions  $h_n(z) := \max\{f(x_n, z), Q_{\mathbb{E}}(x_n, \mu)\}$ . As in the previous proof we obtain

$$|\max\{f(x_n, z), Q_{\mathbb{E}}(x_n, \mu)\}| \leq \beta\|z\| + \beta\|Tx_n\| + \gamma + |c^\top x_n| + |Q_{\mathbb{E}}(x_n, \mu)|.$$

According to Proposition 2.2(ii),  $Q_{\mathbb{E}}(\cdot, \mu)$  is continuous at  $x$ . Hence there exists an upper bound  $\kappa$  for the sequence  $\{\beta\|Tx_n\| + \gamma + |c^\top x_n| + |Q_{\mathbb{E}}(x_n, \mu)|\}$ , yielding the integrable majorant  $\beta\|z\| + \kappa$  for  $|h_n(z)|$ . Again by the continuity of  $Q_{\mathbb{E}}(\cdot, \mu)$  at  $x$ , and by  $\mu(E(x)) = 0$ , the pointwise convergence  $h_n(z) \rightarrow h(z) := \max\{f(x, z), Q_{\mathbb{E}}(x, \mu)\}$  fails at most on a set with  $\mu$ -measure zero. Lebesgue's Theorem now provides

$$\begin{aligned} \lim_{x_n \rightarrow x} Q_{max}(x_n, \mu) &= \int_{\mathbb{R}^s} \lim_{x_n \rightarrow x} \max\{f(x_n, z), Q_{\mathbb{E}}(x_n, \mu)\} \mu(dz) \\ &= \int_{\mathbb{R}^s} \max\{f(x, z), Q_{\mathbb{E}}(x, \mu)\} \mu(dz) = Q_{max}(x, \mu), \end{aligned}$$

and the proof is complete.  $\square$

**Proposition 3.3** *Assume (A1), (A2). Let  $\mu \in \Delta_{p,K}(\mathbb{R}^s)$  for some  $p > 1$  and  $K > 0$ , and  $\mu(E(x)) = 0$ . Then  $Q_{max} : \mathbb{R}^m \times \Delta_{p,K}(\mathbb{R}^s) \rightarrow \mathbb{R}$  is continuous at  $(x, \mu)$ .*

**Proof:** Let  $x, x_n \in \mathbb{R}^m, x_n \rightarrow x$  as well as  $\mu, \mu_n \in \Delta_{p,K}(\mathbb{R}^s), \mu_n \xrightarrow{w} \mu$ . We introduce the functions  $h_n(z) := \max\{f(x_n, z), Q_E(x_n, \mu_n)\}$  and  $h(z) := \max\{f(x, z), Q_E(x, \mu)\}$ . Let  $E := \{z \in \mathbb{R}^s : \exists z_n \rightarrow z \text{ such that } h_n(z_n) \not\rightarrow h(z)\}$ . Then it holds  $E \subseteq E(x)$ . Indeed, if  $z \notin E(x)$  then  $f(x_n, z_n) \rightarrow f(x, z)$ , and moreover, by Proposition 2.2(iii),  $Q_E(x_n, \mu_n) \rightarrow Q_E(x, \mu)$ . This implies  $z \notin E$ . In view of  $\mu(E(x)) = 0$  we now have  $\mu(E) = 0$ . Rubin's theorem on weak convergence of image measures ([4]) then implies  $\mu_n \circ h_n^{-1} \xrightarrow{w} \mu \circ h^{-1}$ . The proof is completed by Theorem 5.4 in [4] stating that  $\mu_n \circ h_n^{-1} \xrightarrow{w} \mu \circ h^{-1}$  and the uniform integrability

$$\limsup_{a \rightarrow \infty} \sup_n \int_{|h_n(z)| \geq a} |h_n(z)| \mu_n(dz) = 0 \quad (13)$$

together imply  $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^s} h_n(z) \mu_n(dz) = \int_{\mathbb{R}^s} h(z) \mu(dz)$  which is just the assertion. To show (13) notice first that

$$\int_{\mathbb{R}^s} |h_n(z)|^p \mu_n(dz) \geq a^{p-1} \int_{|h_n(z)| \geq a} |h_n(z)| \mu_n(dz) \quad (14)$$

and, in addition,

$$\begin{aligned} |h_n(z)|^p &= |\max\{c^\top x_n + \Phi(z - Tx_n), Q_E(x_n, \mu_n)\}|^p \\ &\leq \max\{|c^\top x_n| + |\Phi(z - Tx_n) - \Phi(0)|, |Q_E(x_n, \mu_n)|\}^p \\ &\leq (\beta\|z\| + \beta\|Tx_n\| + \gamma + |c^\top x_n| + |Q_E(x_n, \mu_n)|)^p. \end{aligned}$$

The continuity of  $Q_E$  at  $(x, \mu)$ , see Proposition 2.2(iii), now implies that there is an upper bound  $\kappa$  for the sequence  $\{\beta\|Tx_n\| + \gamma + |c^\top x_n| + |Q_E(x_n, \mu_n)|\}$ . The above estimate thus can be continued:

$$|h_n(z)|^p \leq (\beta\|z\| + \kappa)^p \leq (2\beta)^p \|z\|^p + (2\kappa)^p. \quad (15)$$

Recall that  $p > 1$  and that  $\mu_n \in \Delta_{p,K}(\mathbb{R}^s)$ . By (14) and (15) we obtain

$$\int_{|h_n(z)| \geq a} |h_n(z)| \mu_n(dz) \leq \frac{1}{a^{p-1}} \int_{\mathbb{R}^s} |h_n(z)|^p \mu_n(dz) \leq \frac{1}{a^{p-1}} ((2\beta)^p K + (2\kappa)^p),$$

with the rightmost term tending to zero if  $a \rightarrow \infty$ . This verifies (13) and completes the proof.  $\square$

**Remark 3.4** *Propositions 3.1-3.3 carry over in a one-to-one manner to the function  $Q_{max,\eta}$  defined in (12). The proofs actually remain the same with the constant  $\eta$  replacing the entities  $Q_E(x_n, \mu)$  and  $Q_E(x_n, \mu_n)$ , respectively.*

Altogether we obtain the following corollary on basic structural properties of the objectives in (4).

**Corollary 3.5** *Let  $Q(x, \mu) := Q_E(x, \mu) + \alpha Q_{\mathcal{R}}(x, \mu)$  with  $Q_{\mathcal{R}}$  and  $\alpha$  following one of the specifications in (5)-(7). Then all statements about  $Q_{max}$  made in Propositions 3.1-3.3 are also valid for  $Q$ .*

To illustrate possible consequences of the joint continuity of  $Q$  in  $(x, \mu)$  for stability analysis let us consider (4) as the following parametric optimization problem with parameter  $\mu$ .

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

Non-convexity of  $Q(\cdot, \mu)$  and the resulting relevance of local solutions lead to considering the following localized optimal values and solution sets

$$\begin{aligned} \varphi_V(\mu) &:= \inf\{Q(x, \mu) : x \in X \cap cl V\}, \\ \Psi_V(\mu) &:= \{x \in X \cap cl V : Q(x, \mu) = \varphi_V(\mu)\} \end{aligned}$$

where  $V \subset \mathbb{R}^m$ . 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  $Z = \Psi_V(\mu) \subset V$ . Roughly speaking, a set of local minimizers has the CLM property if it contains all ‘‘nearby’’ local minimizers. Without this property pathologies under perturbations of  $\mu$  may occur. Isolated local minimizers and the set of global minimizers are examples for CLM sets, while strict local minimizers not necessarily obey the CLM property, see [19] for details.

**Proposition 3.6** *Assume (A1), (A2). Let  $\mu \in \Delta_{p,K}(\mathbb{R}^s)$ , for some  $p > 1$  and  $K > 0$ , have 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  $\varphi_V : \Delta_{p,K}(\mathbb{R}^s) \rightarrow \mathbb{R}$  is continuous at  $\mu$ , where  $\Delta_{p,K}(\mathbb{R}^s)$  is equipped with weak convergence of probability measures,*
- (ii) *the multifunction  $\Psi_V : \Delta_{p,K}(\mathbb{R}^s) \rightarrow 2^{\mathbb{R}^m}$  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  $\Delta_{p,K}(\mathbb{R}^s)$ , 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  $\Delta_{p,K}(\mathbb{R}^s)$  such that for all  $\nu \in \mathcal{N}'$  the set  $\Psi_V(\nu)$  is a CLM set for  $P(\nu)$  with respect to  $V$ .*

Joint continuity of  $Q$  in  $(x, \mu)$  established, the proof of the above proposition follows the lines of Berge's theory as, for instance, in the proof of Proposition 4.2.2 in [2]. It is essential that the unperturbed CLM set is compact which is hidden in the boundedness assumption on  $V$ . In (iii), nonemptiness of  $\Psi_V(\nu)$  then is immediate. Here, the point is that  $\Psi_V(\nu)$  is a CLM set and hence a set of local minimizers.

Proposition 3.6 gives rise to conclusions involving specific modes of weak convergence such as discretizations via conditional expectations, pointwise almost surely converging densities, or empirical measures. In this way, known stability results for the expectation model (1), cf. [20, 23], extend to the mean-risk models in the present paper.

## 4 Algorithmic Issues

Assume that the probability measure  $\mu$  is discrete with realizations  $z_j$  and probabilities  $\pi_j, j = 1, \dots, J$ . By (8)-(10) the specifications (5)-(7) of (4) then turn into equivalent mixed-integer linear programs. The central-deviation model

$$\min\{Q_{\mathbb{E}}(x, \mu) + \alpha \cdot Q_{\mathcal{D}}(x, \mu) : x \in X\}, \quad \text{with } 0 \leq \alpha \leq \frac{1}{2}, \quad (16)$$

is equivalent to

$$\begin{aligned} \min\{(1 - 2\alpha)c^\top x + (1 - 2\alpha) \sum_{j=1}^J \pi_j (q^\top y_j + q'^\top y'_j) + 2\alpha \cdot \sum_{j=1}^J \pi_j v_j : \\ Tx + Wy_j + W'y'_j = z_j, \\ c^\top x + q^\top y_j + q'^\top y'_j \leq v_j, \\ c^\top x + \sum_{i=1}^J \pi_i (q^\top y_i + q'^\top y'_i) \leq v_j, \\ x \in X, y_j \in \mathbb{Z}_+^{\bar{m}}, y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}, j = 1, \dots, J\}. \end{aligned} \quad (17)$$

The semideviation model

$$\min\{Q_{\mathbb{E}}(x, \mu) + \alpha \cdot Q_{\mathcal{D}^+}(x, \mu) : x \in X\}, \quad \text{with } 0 \leq \alpha \leq 1, \quad (18)$$

is equivalent to

$$\begin{aligned} \min\{(1 - \alpha)c^\top x + (1 - \alpha) \sum_{j=1}^J \pi_j (q^\top y_j + q'^\top y'_j) + \alpha \cdot \sum_{j=1}^J \pi_j v_j : \\ Tx + Wy_j + W'y'_j = z_j, \\ c^\top x + q^\top y_j + q'^\top y'_j \leq v_j, \\ c^\top x + \sum_{i=1}^J \pi_i (q^\top y_i + q'^\top y'_i) \leq v_j, \\ x \in X, y_j \in \mathbb{Z}_+^{\bar{m}}, y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}, j = 1, \dots, J\}. \end{aligned} \quad (19)$$

The expected-excess model

$$\min\{Q_{\mathbb{E}}(x, \mu) + \alpha \cdot Q_{\mathcal{D}\eta}(x, \mu) : x \in X\}, \quad \text{with } \alpha \geq 0 \text{ and } \eta \in \mathbb{R}, \quad (20)$$

is equivalent to

$$\begin{aligned} \min\{c^\top x &+ \sum_{j=1}^J \pi_j(q^\top y_j + q'^\top y'_j) + \alpha \cdot \sum_{j=1}^J \pi_j v_j : \\ &Tx + Wy_j + W'y'_j = z_j, \\ &c^\top x + q^\top y_j + q'^\top y'_j - \eta \leq v_j, \\ &x \in X, y_j \in \mathbb{Z}_+^{\bar{m}}, y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}_+, j = 1, \dots, J\}. \end{aligned} \quad (21)$$

Algorithmically, it will be beneficial that in the constraints of (21) there is no explicit linkage between  $(y, y')$ -variables for different realizations  $j \in \{1, \dots, J\}$ . These variables are linked only implicitly via the first-stage variable  $x$ . In (17) and (19) explicit linkage occurs due to the conditions

$$c^\top x + \sum_{i=1}^J \pi_i(q^\top y_i + q'^\top y'_i) \leq v_j, \quad j = 1, \dots, J. \quad (22)$$

In principle, problems (17), (19), and (21) could be tackled with general mixed-integer linear programming solvers. With growing number  $J$  of scenarios however, such solvers typically come to their limits. This motivates approaches benefitting in one or another way from decomposition effects in the above models. Our point of departure is to see (16), (18), and (20) as non-convex global optimization problems and to tackle them with traditional branch-and-bound. This means to partition  $X$  by linear inequalities stepwise with increasing granularity, to find upper and lower bounds of the optimal objective values on current elements of the partition, and to apply coordination rules for guiding the branching and for pruning due to infeasibility, inferiority, or optimality.

In what follows, emphasis will be on bounding, in fact on lower bounds, since this is where decomposition becomes relevant. It will pay to pass from the compound problems (16), (18), and (20) to their expanded equivalents (17), (19), and (21). We will drop details about coordination where we have employed established rules or their analogues.

Let us first consider the expected-excess model (21). We replace  $x$  by copies  $x_j, j = 1, \dots, J$ , according to the number of scenarios and add the constraints  $x_1 = \dots = x_J$  which we assume to be represented as  $\sum_{j=1}^J H_j x_j$  with suitable  $l \times m$  matrices  $H_j$ . Except for  $x_1 = \dots = x_J$ , the constraints in (21) now are fully separable with respect to  $j = 1, \dots, J$ . The objective being separable anyhow, this suggests Lagrangian relaxation of  $x_1 = \dots = x_J$  leading to the following Lagrangian function and Lagrangian dual

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

with

$$L_j(x_j, y_j, y'_j, v_j, \lambda) := \pi_j(c^\top x_j + q^\top y_j + q'^\top y'_j + \alpha v_j) + \lambda^\top H_j x_j, \quad j = 1, \dots, J,$$

and

$$\max\left\{\sum_{j=1}^J D_j(\lambda) : \lambda \in \mathbb{R}^l\right\} \quad (23)$$

with

$$\begin{aligned} D_j(\lambda) = \min\{L_j(x_j, y_j, y'_j, v_j, \lambda) &: Tx_j + Wy_j + W'y'_j = z_j, \\ &c^\top x_j + q^\top y_j + q'^\top y'_j - \eta \leq v_j, \\ &x_j \in X, y_j \in \mathbb{Z}_+^{\bar{m}}, y'_j \in \mathbb{R}_+^{m'}, v_j \in \mathbb{R}_+\}. \end{aligned} \quad (24)$$

During the branch-and-bound procedure sketched above these formulas apply at each node of the branching tree, with  $X$  in (24) replaced by the corresponding element of the partition. Integer programming

theory says that the optimal value of (23) provides a lower bound to the optimal objective value that is never worse than the bound obtainable by relaxing integrality. The non-smooth concave maximization (or convex minimization) problem (23) can be solved by subgradient methods ([10, 12]) that require one function value and one subgradient per iteration. By the additive nature of  $\sum_{j=1}^J D_j(\lambda)$  these entities are computable by solving scenario-wise the mixed-integer linear programs (24), which is the announced decomposition effect.

For solving problems (17), (19), the principal bounding approach to introduce copies of  $x$  and continue with Lagrangian relaxation of  $x_1 = \dots = x_J$  is working, too. However, the constraints (22) then prevent the observed decoupling in the Lagrangian dual which is the key benefit above. This motivates to derive separable lower bounds for the objectives in (16), (18). A straightforward such bound clearly is  $Q_{\mathcal{E}}(x, \mu)$ . The following lemma identifies separable lower bounds strengthening this straightforward bound.

**Lemma 4.1** *Assume (A1)-(A3), fix  $x \in X$ , and let  $\eta \leq Q_{\mathcal{E}}(x, \mu)$ . Then the following is valid*

$$Q_{\mathcal{E}}(x, \mu) \leq (1 - 2\alpha)Q_{\mathcal{E}}(x, \mu) + 2\alpha Q_{\mathcal{D}^\eta}(x, \mu) + 2\alpha\eta \leq Q_{\mathcal{E}}(x, \mu) + \alpha Q_{\mathcal{D}}(x, \mu), \quad (25)$$

$$Q_{\mathcal{E}}(x, \mu) \leq (1 - \alpha)Q_{\mathcal{E}}(x, \mu) + \alpha Q_{\mathcal{D}^\eta}(x, \mu) + \alpha\eta \leq Q_{\mathcal{E}}(x, \mu) + \alpha Q_{\mathcal{D}^+}(x, \mu), \quad (26)$$

for  $0 \leq \alpha \leq \frac{1}{2}$  in (25) and  $0 \leq \alpha \leq 1$  in (26), respectively.

**Proof:** Without restriction on  $\eta$  it holds that  $\max\{f(x, z), \eta\} \geq f(x, z)$ , implying  $\mathbb{E} \max\{f(x, z), \eta\} \geq \mathbb{E}f(x, z)$ . With (10) this extends into the estimate

$$2\alpha Q_{\mathcal{D}^\eta}(x, \mu) + 2\alpha\eta = 2\alpha \mathbb{E} \max\{f(x, z), \eta\} \geq 2\alpha \mathbb{E}f(x, z),$$

yielding the first inequality in (25). Now let  $\eta \leq Q_{\mathcal{E}}(x, \mu)$ . Employing (8) and (10) we obtain

$$\begin{aligned} Q_{\mathcal{E}}(x, \mu) + \alpha Q_{\mathcal{D}}(x, \mu) &= (1 - 2\alpha)\mathbb{E}f(x, z) + 2\alpha \mathbb{E} \max\{f(x, z), \mathbb{E}f(x, z)\} \\ &\geq (1 - 2\alpha)\mathbb{E}f(x, z) + 2\alpha \mathbb{E} \max\{f(x, z), \eta\} = (1 - 2\alpha)Q_{\mathcal{E}}(x, \mu) + 2\alpha Q_{\mathcal{D}^\eta}(x, \mu) + 2\alpha\eta \end{aligned}$$

which is the second inequality in (25). The proof of (26) is analogous.  $\square$

The lemma leaves room for studying different specifications of  $\eta$  together with their computational impacts. In particular, it is desirable to find specifications where the first inequalities in (25), (26) have potential to hold strictly. A candidate for the selection of  $\eta$  is the so-called wait-and-see solution ([5, 11, 18, 21]). It is given as the expected value of

$$\Phi_{WS}(z) := \min\{c^\top x + q^\top y + q'^\top y' : Tx + Wy + W'y' = z, x \in X, y \in \mathbb{Z}_+^m, y' \in \mathbb{R}_+^{m'}\}.$$

Clearly,  $\Phi_{WS}(z) \leq f(x, z)$  such that  $\eta := \mathbb{E}\Phi_{WS}(z) \leq Q_{\mathcal{E}}(x, \mu)$  for arbitrary  $x \in X$ , showing that the wait-and-see solution is a feasible choice in Lemma 4.1. With this selection of  $\eta$ , and for fixed  $x \in X$ , a sufficient condition for the first inequalities in (25), (26) to be strict is that the set  $\{z \in \mathbb{R}^s : \mathbb{E}\Phi_{WS}(z) > f(x, z)\}$  has positive  $\mu$ -measure.

Let us return to the branch-and-bound approach for solving (16), (18), (20) and have a quick look at upper bounds of optimal objective values on the elements of the partition. The basic idea is to find promising feasible points. When applying lower bounding via the Lagrangian dual (23), either directly to (21) or induced by Lemma 4.1 to (17), (19), the results of the dual optimization provide starting points for upper bounding heuristics. Consider  $x_j$ -parts ( $j = 1, \dots, J$ ) of optimal solutions to the MILPs in (24) for values of  $\lambda$  that are optimal or nearly optimal in (23). See these  $x_j$ -parts as proposals for a nonanticipative first-stage solution  $x$  that is behind the relaxed requirement  $x_1 = \dots = x_J$ . Heuristically reestablish  $x_1 = \dots = x_J$  (primal feasibility) on the basis of the proposals, for instance, by deciding for the most frequent one arising or by averaging and rounding to integers if necessary.

We conclude with some preliminary computational experiences with an implementation of the sketched branch-and-bound scheme to solve expected-excess models. Lower bounding follows (23), (24). For upper bounding a heuristic combining selection by frequency and rounding was used. Note that Lemma 4.1 enables solution of central- and semideviation models by expected-excess methodology. The main aim of our experiments was to study performance improvement of the algorithm when replacing the trivial lower bound of  $Q_{\mathcal{E}}(x, \mu)$  by the bound proposed in Lemma 4.1, with  $\eta$  selected as the wait-and-see solution. Experiments were carried out with a semideviation model that is a mean-risk extension of a two-stage



linear mixed-integer stochastic program originating from a chemical engineering application. The latter is described in detail in [8]. The model has  $m=24$  first-stage variables, all integer or binary, and 3 first-stage constraints. In the second stage there are  $\bar{m}=108$  integer or binary and  $m'=224$  continuous variables, together with 311 constraints. Each test instance has  $J=10$  scenarios.

For 10 test instances Table 1 displays relative gaps in % after 4 hours of cpu time on a Sun V880 with a 880 MHz processor and 4 GB of main memory. The gaps were calculated from the differences between the least upper and biggest lower bounds to the problems' optimal values that were obtained by the different methods. The CPLEX column reports direct application of the ILOG-CPLEX 8.1 mixed-integer linear programming solver to the equivalent given by (19). The B&B/ENH and B&B/EXP columns show the gaps for our branch-and-bound algorithm with lower bounding enhanced by Lemma 4.1 and with lower bounding by  $Q_{\mathbb{E}}(x, \mu)$ , respectively. The table indicates positive impact of both decomposition in principle and of the enhanced bounds from Lemma 4.1. We have observed (see instances 2, 5, 9) that the enhancement is particularly efficient if there is substantial dispersion among  $f(x, z_j), j = 1, \dots, J$ , for relevant  $x$ .

The above algorithmic ideas contribute to the literature on decomposition of stochastic integer programs, see for instance [1, 7]. In particular, they extend techniques for expectation models from [7] to the mean-risk models in the present paper.

Instance	CPLEX	B&B/ENH	B&B/EXP
1	86.40	3.01	5.05
2	94.30	16.16	47.41
3	57.80	4.02	6.94
4	10.99	4.31	4.43
5	89.26	7.49	20.86
6	8.73	4.46	7.54
7	6.06	3.62	7.41
8	5.31	5.34	8.64
9	5.34	1.18	5.45
10	97.03	3.87	6.79

Table 1: Gaps in %, semideviation model, cpu time: 4h

**Acknowledgement.** We are grateful to Christoph Helmberg (TU Chemnitz) for giving us access to his CONIC-BUNDLE Code for non-smooth optimization.

## References

- [1] Alonso-Ayuso, A.; Escudero, L.F.; Ortuño, M.T.: BFC, A branch-and-fix coordination algorithmic framework for solving some types of stochastic pure and mixed 0-1 programs, *European Journal of Operational Research* 151 (2003), 503-519.
- [2] Bank, B.; Guddat, J.; Klatte, D.; Kummer, B.; Tammer, K.: *Non-linear Parametric Optimization*, Akademie-Verlag, Berlin, 1982.
- [3] Bank, B.; Mandel, R.: *Parametric Integer Optimization*, Akademie-Verlag, Berlin, 1988.
- [4] Billingsley, P.: *Convergence of Probability Measures*, Wiley, New York, 1968.
- [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] Engell, S.; Märkert, A.; Sand, G.; Schultz, R.; Schulz, Ch.: Online scheduling of multiproduct batch plants under uncertainty, In: M. Grötschel, S.O. Krumke, J. Rambau (eds.) *Online Optimization of Large Scale Systems*, Springer, Berlin, 2001, 649-676.
- [9] Fishburn, P.C.: Mean-risk analysis with risk associated with below-target returns, *American Economic Review* 67 (1977), 116-126.
- [10] Helmborg, C.; Kiwiel, K.C.: A spectral bundle method with bounds, *Mathematical Programming* 93 (2002), 173-194.
- [11] Kall, P.; Wallace, S.W.: *Stochastic Programming*, Wiley, Chichester, 1994.
- [12] Kiwiel, K. C.: Proximity control in bundle methods for convex nondifferentiable optimization, *Mathematical Programming* 46 (1990), 105-122.
- [13] Levy, H.: Stochastic dominance and expected utility: survey and analysis, *Management Science* 38 (1992), 555-593.
- [14] Märkert, A.: Deviation measures in stochastic programming with mixed-integer recourse, PhD Thesis, Institute of Mathematics, University of Duisburg-Essen, 2004.
- [15] Müller, A.; Stoyan, D.: *Comparison Methods for Stochastic Models and Risk*, Wiley, Chichester, 2002.
- [16] Ogryczak, W.; Ruszczyński, A.: From stochastic dominance to mean-risk models: Semideviations as risk measures, *European Journal of Operational Research* 116 (1999), 33-50.
- [17] Ogryczak, W.; Ruszczyński, A.: Dual stochastic dominance and related mean-risk models, *SIAM Journal on Optimization* 13 (2002), 60-78.
- [18] Prékopa, A.: *Stochastic Programming*, Kluwer, Dordrecht, 1995.
- [19] Robinson, S.M.: Local epi-continuity and local optimization, *Mathematical Programming* 37 (1987), 208-222.
- [20] Römisch, W.: Stability of stochastic programming problems, In: A. Ruszczyński, A. Shapiro (eds.) *Handbooks in Operations Research and Management Science*, 10: *Stochastic Programming*, Elsevier, Amsterdam, 2003, 483-554.
- [21] Ruszczyński, A.; Shapiro, A. (Eds.): *Handbooks in Operations Research and Management Science*, 10: *Stochastic Programming*, Elsevier, Amsterdam, 2003.
- [22] Schultz, R.: On structure and stability in stochastic programs with random technology matrix and complete integer recourse, *Mathematical Programming* 70 (1995), 73-89.
- [23] Schultz, R.: Some aspects of stability in stochastic programming, *Annals of Operations Research* 100 (2000), 55-84.