DEPARTMENT OF OPERATIONS RESEARCH
UNIVERSITY OF AARHUS

Working Paper no. 2001/3

# Applying the minimum risk criterion in stochastic recourse programs

Morten Riis
Rüdiger Schultz

ISSN 1600-8987

Department of Mathematical Sciences
Telephone: +45 8942 3188
E-mail: institut@imf.au.dk

Building 530, Ny Munkegade
DK-8000 Aarhus C, Denmark
URL: www.imf.au.dk

# Applying the minimum risk criterion in stochastic recourse programs

Morten Riis*
Department of Operations Research
University of Aarhus, building 530
Ny Munkegade
DK - 8000 Århus C
Denmark

Rüdiger Schultz
Department of Mathematics
Gerhard-Mercator University Duisburg
Lotharstrasse 65
D - 47048 Duisburg
Germany

**Abstract**

In the setting of stochastic recourse programs, we consider the problem of minimizing the probability of total costs exceeding a certain threshold value. The problem is referred to as the minimum risk problem and is posed in order to obtain a more adequate description of risk aversion than that of the accustomed expected value problem. We establish continuity properties of the recourse function as a function of the first-stage decision, as well as of the underlying probability distribution of random parameters. This leads to stability results for the optimal solution of the minimum risk problem when the underlying probability distribution is subjected to perturbations. Furthermore, an algorithm for the minimum risk problem is elaborated and we present results of some preliminary computational experiments.

*Keywords:* Stochastic Programming; Risk Aversion; Continuity; Stability.

# 1 Introduction

Stochastic recourse programs arise as optimization problems in situations where some parameters of the underlying model are not known with certainty. Assuming that some information on the probability distribution of the unknown parameters is available, the objective is to formulate an optimization problem, explicitly taking all outcomes of the random parameters into account rather than simply replacing them by their expected values. It is assumed that some decisions must be taken before the outcome of random parameters is revealed and

*Corresponding author. Email: riis@imf.au.dk

hence must be based on the knowledge of the distribution of the parameters only. This is referred to as the first stage. In the second stage, outcomes of all random parameters have been observed and some corrective (or *recourse*) actions may be taken.

We start out by presenting a standard two-stage stochastic linear program with recourse in which the sum of first-stage costs and the expected value of second-stage costs is minimized. Assuming that randomness occurs only in the technology matrix and the second-stage right-hand side, this problem is stated as follows:

$$EVP \qquad \min\{c^T x + \mathcal{Q}_E(x) : x \in X\}, \tag{1.1}$$

where $\mathcal{Q}_E(x)$ denotes the expected second-stage cost given the first-stage decision $x$,

$$\mathcal{Q}_E(x) = \mathbb{E}_\omega \Phi\big(h(\omega) - T(\omega)x\big),$$

and the second-stage value function $\Phi$ is defined by

$$\Phi(\tau) = \min\{q^T y : Wy \geq \tau,\ y \in \mathbb{R}^{n_2}_+\}. \tag{1.2}$$

Here $X \subseteq \mathbb{R}^{n_1}$ is assumed to be non-empty and closed, $c \in \mathbb{R}^{n_1}$ and $q \in \mathbb{R}^{n_2}$ are known vectors and $W$ is a known rational $m \times n_2$-matrix. The technology matrix $T$ and the second-stage right-hand side $h$, on the other hand, are dependent on the outcome of a random event $\omega$. Denoting by $\mathbb{R}^{m\times n}$ the space of real $m\times n$-matrices, we assume that $T : \Omega \mapsto \mathbb{R}^{m\times n_1}$ and $h : \Omega \mapsto \mathbb{R}^m$ are measurable mappings defined on some probability space $(\Omega, \mathcal{F}, P)$.

Several objections may be put forward against the formulation of the expected value problem given by (1.1), a primary objection being that minimization of the expected cost does not always constitute an appropriate objective. The appropriateness of this criterion is dependent on the assumption that the decision process is to be repeated a great number of times, implying by the law of large numbers that, in the long run, average cost will be equal to the expected cost. This assumption, however, will frequently not be justified and consequently the expected cost may not be of much interest to the decision maker. Another major objection against the expected cost as the object of minimization, is the fact that the optimal solution of the expected value problem may only assure the achievement of the corresponding expected cost with a relatively small probability. These considerations imply that the risk averse decision maker will not consider the solution of the expected value problem to be "optimal". Instead, what may be desired is a solution ensuring a low probability of very large costs. This leads us to apply the minimum risk criterion (see e.g. Bereanu [4]) to the setting described above. The minimum risk problem is the problem of minimizing the probability of total costs exceeding some threshold value $\phi$. The threshold

value may be thought of as the level of bankruptcy or even just a budget limit. Formally, the problem is stated as follows:

$$MRP \qquad \min\{\mathcal{Q}_P(x) : x \in X\}, \tag{1.3}$$

where the recourse function $\mathcal{Q}_P(x)$ now denotes the probability of total costs incurred in the two stages, exceeding the threshold value $\phi$, given the first-stage decision $x$,

$$\mathcal{Q}_P(x) = P\big(\{\omega \in \Omega : c^T x + \Phi\big(h(\omega) - T(\omega)x\big) > \phi\}\big),$$

and the second-stage value function $\Phi$ is still defined by (1.2).

Since we are going to address the issue of stability of the minimum risk problem when the underlying distribution of $\omega$ is subjected to perturbations, we will be interested in the structural properties of the recourse function $Q_P$ as a function of $x$ as well as the distribution of $\omega$. To facilitate such an analysis it will be convenient to introduce the induced probability measure $\mu = P \circ (h,T)^{-1}$ on $\mathbb{R}^{m\times(1+n_1)}$ and restate the minimum risk problem as follows:

$$MRP(\mu) \qquad \min\{\mathcal{Q}(x,\mu) : x \in X\},$$

where

$$\mathcal{Q}(x,\mu) = \mu\big(\{(h,T) \in \mathbb{R}^{m\times(1+n_1)} : c^T x + \Phi\big(h - Tx\big) > \phi\}\big),$$

so that the recourse function $\mathcal{Q}$ now explicitly depends on the distribution $\mu$.

Structural properties of the expected value problem (1.1) have been studied by numerous authors. We refer to the textbooks by Birge and Louveaux [7], Kall and Wallace [15] and Prékopa [21] and research papers by e.g. Dupačová [9], Kall [14], Robinson and Wets [26], Römisch and Schultz [27, 28, 29], Shapiro [36] and Wets [39]. In this paper we will show that the minimum risk problem (1.3) is equivalent to an expected value problem in which a binary variable and an additional constraint have been included in the second stage. Thus, the minimum risk problem belongs to the general class of two-stage stochastic programs with mixed-integer recourse. Structural properties for this class of problems have been studied by e.g. Artstein and Wets [2], Klein Haneveld and van der Vlerk [10], Louveaux and van der Vlerk [20], Rinnooy Kan and Stougie [16], Schultz [31, 32, 33, 34] and Stougie [38]. The minimum risk problem, however, possesses a much simpler structure than that of a general stochastic program with mixed-integer recourse and in this paper we will show that some of the results previously established for this general class of problems, remain valid for the minimum risk problem under more general assumptions. Early results in this direction were obtained by Raik [23, 24] who established lower semicontinuity of the recourse function $\mathcal{Q}_P$, as well as a sufficient condition for continuity. (See also Kibzun and Kan [17].)

Also, the issue of solution procedures for the expected value problem has been the center of extensive research for a number of years. In 1969 van Slyke and Wets [37] introduced the L-shaped algorithm. Since then much work has been done to improve performance of this algorithm and to develop alternative solution procedures. Currently, the most powerful algorithms for this class of problems include regularized decomposition introduced by Ruszczynski [30] and stochastic decomposition introduced by Higle and Sen [12]. We will present an algorithm for the minimum risk problem, which may be seen as a specialized version of the L-shaped algorithm. Since regularized decomposition as well as stochastic decomposition are similar in spirit to this procedure, extensions of the algorithm incorporating these techniques are natural. Such an approach is not carried completely to an end in the present paper, although we will discuss the inclusion of some regularizing mechanism.

This paper is organized as follows: in Section 2 we present some prerequisites from probability theory, needed for the subsequent analysis. Next, in Section 3 we establish lower semicontinuity and a sufficient condition for continuity of $\mathcal{Q}$ as a function of $x$, as well as a sufficient condition for joint continuity of $\mathcal{Q}$ as a function of $x$ and $\mu$. Furthermore, we present a quantitative continuity result for $\mathcal{Q}$ as a function of $\mu$. Having established these continuity properties of $\mathcal{Q}$ it is straightforward to arrive at stability results for the optimal solution value and the optimal solution set of $MRP(\mu)$ when the underlying probability measure is subjected to perturbations. Such an analysis is carried out in Section 4. In particular, the stability properties established in Section 4 justify numerical procedures that rely on approximating the distribution of $\mu$ by simpler (discrete) ones. In Section 5 such a procedure is elaborated. The procedure, which is a modified version of the L-shaped algorithm, was implemented in C++ and a number of computational experiments were performed to test its practicability. In Section 6 we describe some implementational details and report results of our computational experiments. Finally, in Section 7 we give some concluding remarks.

# 2 Prerequisites

In this section we present some basic concepts and results from probability theory used throughout the paper. For a more thourough discussion of these topics we refer to the textbook by Hoffmann-Jørgensen [13].

We shall be concerned with the set of all Borel probability measures on $\mathbb{R}^n$ which we denote by $\mathcal{P}(\mathbb{R}^n)$. We recall that a set $A \subseteq \mathbb{R}^n$ is said to be measurable if $A \in \mathcal{B}(\mathbb{R}^n)$, where $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel $\sigma$-algebra on $\mathbb{R}^n$. In particular, all open sets and all closed sets are measurable. Also, a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is said to be measurable if and only if

$$\{x \in \mathbb{R}^n : f(x) \leq a\} \in \mathcal{B}(\mathbb{R}^n) \quad \forall\, a \in \mathbb{R}$$

and hence, in particular, the indicator function of a measurable set is measurable.

Next, consider a sequence of measurable sets $\{A_n\}$ in $\mathcal{B}(\mathbb{R}^n)$. The lim inf and lim sup of this sequence are defined by

$$\liminf_{n\to\infty} A_n = \bigcup_{j\geq 1} \bigcap_{n\geq j} A_n \qquad \text{and} \qquad \limsup_{n\to\infty} A_n = \bigcap_{j\geq 1} \bigcup_{n\geq j} A_n. \tag{2.1}$$

Hence, the lim inf is the set of all those $x \in \mathbb{R}^n$ for which there exists $N \in \mathbb{N}$ such that $x \in A_n$ for all $n \geq N$, whereas the lim sup is the set of all those $x \in \mathbb{R}^n$ such that $x \in A_n$ for infintely many $n$. We will need the following result:

**Proposition 2.1.** *Let $\mu \in \mathcal{P}(\mathbb{R}^n)$, $A \in \mathcal{B}(\mathbb{R}^n)$ and let $\{A_n\}$ be a sequence of sets in $\mathcal{B}(\mathbb{R}^n)$. Then,*
*(a)* $\liminf_{n\to\infty} \mu(A_n) \geq \mu(\liminf_{n\to\infty} A_n)$;
*(b)* $\limsup_{n\to\infty} \mu(A_n) \leq \mu(\limsup_{n\to\infty} A_n)$,

cf. chapter 1.4 in Hoffman-Jørgensen [13].

While studying continuity properties of the recourse function $\mathcal{Q}$ and stability of the minimum risk problem $MRP(\mu)$ with respect to $\mu$ we shall adopt the notion of weak convergence on $\mathcal{P}(\mathbb{R}^n)$ defined as follows. Let $\mu \in \mathcal{P}(\mathbb{R}^n)$ and let $\{\mu_n\}$ be some sequence of probability measures in $\mathcal{P}(\mathbb{R}^n)$. If for any bounded continuous function, $g : \mathbb{R}^n \mapsto \mathbb{R}$, we have

$$\int_{\mathbb{R}^n} g(x)\mu_n(dx) \xrightarrow{n\to\infty} \int_{\mathbb{R}^n} g(x)\mu(dx),$$

then the sequence $\{\mu_n\}$ is said to converge weakly to $\mu$ and we write $\mu_n \xrightarrow{w} \mu$.

In the subsequent sections we will let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ describe the joint distribution of the technology matrix and the second-stage right-hand side. At some points we shall also consider marginal and conditional distributions. To this end, we denote by $\mu_1$ and $\mu_2$ the marginal distributions of $h$ and $T$, respectively, and for $T \in \mathbb{R}^{m\times n_1}$ we denote by $\mu_1^2(\cdot, T)$ the conditional distribution of $h$ given $T$. The marginal and conditional distributions possess the following properties:

- $\mu_1$ and $\mu_2$ are probability measures on $\mathbb{R}^m$ and $\mathbb{R}^{m\times n_1}$, respectively.
- $\mu_1^2(\cdot, T)$ is a probability measure on $\mathbb{R}^m$ for any $T \in \mathbb{R}^{m\times n_1}$.
- $\mu_1^2(A, \cdot)$ is a measurable function on $\mathbb{R}^{m\times n_1}$ for any $A \in \mathcal{B}(\mathbb{R}^m)$.
- For any $B \in \mathcal{B}(\mathbb{R}^{m\times(1+n_1)})$, we have $\mu(B) = \int_{\mathbb{R}^{m\times n_1}} \int_{\mathbb{R}^m} \mathbf{1}_B(h,T)\, \mu_1^2(dh,T)\mu_2(dT)$, where $\mathbf{1}_B$ denotes the indicator function of the set $B$.

# 3 Structural properties

In this section we discuss the structural properties of the minimum risk problem, and in particular of the recourse function $\mathcal{Q}$. We will show that the problem is equivalent to an expected value problem with mixed-integer second stage, and hence shares the structural properties established for this class of problems. These properties, however, will be shown to remain valid for the minimum risk problem under more general assumptions than those employed when studying the general class of problems. We will make just the following two assumptions, ensuring that the second-stage value function $\Phi$ is real-valued.

(A1) There exists a vector $u \in \mathbb{R}^m_+$ satisfying $W^T u \leq q$.

(A2) For all $t \in \mathbb{R}^m$ there exists a second-stage solution $y \in \mathbb{R}^{n_2}_+$ satisfying $Wy \geq t$.

Assumption (A1) is employed to ensure dual feasibility and hence boundedness of all second-stage problems while assumption (A2) is the assumption of complete recourse, ensuring feasibility of all second-stage problems for all possible right-hand sides. For practical purposes it is often sufficient to replace assumption (A2) by the weaker assumption of relative complete recourse, ensuring feasibility of all second-stage problems only for those right-hand sides corresponding to a feasible first-stage solution. That is, for all $\omega \in \Omega$ and for all $x \in X$ there is a second-stage solution $y \in \mathbb{R}^{n_2}_+$ satisfying $Wy \geq h(\omega) - T(\omega)x$. We note, that if relative complete recourse is not inherent in the problem, it may be established by the inclusion of feasibility cuts (see e.g. Birge and Louveaux [7]). Under assumptions (A1) and (A2) the second-stage value function $\Phi$ is a real-valued, piecewise linear and convex function on $\mathbb{R}^m$.

Evidently, the minimum risk problem (1.3) is equivalent to an expected value problem where the expectation is taken of an appropriately defined indicator function. Specifically, for $x \in \mathbb{R}^{n_1}$ we may define the set of all outcomes of random parameters yielding total costs exceeding the threshold value,

$$M(x) := \{(h,T) \in \mathbb{R}^{m\times(1+n_1)} : c^T x + \Phi\big(h - Tx\big) > \phi\}$$

and introduce the corresponding indicator function, $\psi : \mathbb{R}^{n_1} \times \mathbb{R}^m \times \mathbb{R}^{m\times n_1} \mapsto \mathbb{B}$. (Here $\mathbb{B} \subseteq \mathbb{R}$ denotes the subset of binary numbers.) Hence, we define

$$\psi(x,h,T) := \begin{cases} 1 & \text{if} \quad (h,T) \in M(x), \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

We now have

$$\mathcal{Q}_P(x) = E_\omega \psi(x, h(\omega), T(\omega)) = \int_\Omega \psi(x, h(\omega), T(\omega)) P(d\omega)$$

and

$$\mathcal{Q}(x,\mu) = \mu(M(x)) = \int_{\mathbb{R}^{m\times(1+n_1)}} \psi(x,h,T)\mu(d(h,T)).$$

One way to define such an indicator function is to let $\psi$ be the value function of the following mixed-integer program:

$$\psi(x,h,T) = \min\{z : Wy \geq h - Tx,\ q^T y - M_1 z \leq \phi - c^T x,\ y \in \mathbb{R}^{n_2}_+,\ z \in \mathbb{B}\}, \qquad (3.2)$$

where $M_1 > 0$ is some large number. Assuming that $c^T x \leq \phi$ for all $x \in X$, a sufficient condition for (3.2) to be feasible is to choose $M_1$ as an upper bound on the second-stage value function:

$$M_1 \geq \sup\{\Phi(h(\omega) - T(\omega)x) : x \in X,\ \omega \in \Omega\}.$$

Note that the supremum exists and is finite, assuming that $\Omega$ and $X$ are bounded. Hence we see that the minimum risk problem (1.3) is equivalent to a two-stage stochastic program with mixed-integer recourse. In particular, the continuity properties established for such programs by e.g. Schultz [32, 33] and Stougie [38] apply to the recourse function $\mathcal{Q}$. Obviously though, the structure of the minimum risk problem is much simpler than that of a general stochastic program with mixed-integer recourse and hence we may strengthen some of the previously established results. This will be the aim of our subsequent analysis. We will show, that in order to obtain the desired continuity properties of the recourse function $\mathcal{Q}$ and the resulting stability properties of the minimum risk problem $MRP(\mu)$, it is no longer necessary to assume that $\mu$ has finite first moment. This is a crucial assumption when studying the general two-stage stochastic program with mixed-integer recourse, which is needed to establish the existence of integrable minorants and majorants of the second-stage value functions.

When studying the structural properties of $\mathcal{Q}$ as a function of $x$, we will find it convenient to define for $x \in \mathbb{R}^{n_1}$ the set $E(x)$ of all those $(h,T) \in \mathbb{R}^{m\times(1+n_1)}$ such that $\psi(\cdot,h,T)$ is discontinuous at $x$. By continuity of $\Phi$ this set is easily seen to be equal to the set of all outcomes of random parameters yielding total costs equal to the threshold value,

$$E(x) := \{(h,T) \in \mathbb{R}^{m\times(1+n_1)} : c^T x + \Phi(h - Tx) = \phi\}.$$

Using this definition and recalling the definition of lim inf and lim sup of sequences of sets in Section 2, we can prove the following lemma.

**Lemma 3.1.** *Assume (A1) and (A2), let $x \in \mathbb{R}^{n_1}$ and let $\{x_n\}$ be some sequence in $\mathbb{R}^{n_1}$ converging to $x$. Then,*
*(a)* $\liminf_{n\to\infty} M(x_n) \supseteq M(x)$;
*(b)* $\limsup_{n\to\infty} M(x_n) \subseteq M(x) \cup E(x)$.

*Proof.* (a) If $(h,T) \in M(x)$ we have by definition of $M(x)$ that $c^T x + \Phi(h - Tx) > \phi$. By continuity of $\Phi$ this means that there exists some $N \in \mathbb{N}$ such that for all $n > N$ we have $c^T x_n + \Phi(h - Tx_n) > \phi$ and hence $(h,T) \in M(x_n)$ for all $n > N$. The result follows immediately, cf. the definition of lim inf (2.1).
(b) Let $(h,T) \in \limsup_{n\to\infty} M(x_n) \setminus M(x)$. This means that $c^T x + \Phi(h - Tx) \leq \phi$ while $c^T x_n + \Phi(h - Tx_n) > \phi$ for infinitely many $n$, cf. the definition of lim sup (2.1). Now, by continuity of $\Phi$, we see that $c^T x + \Phi(h - Tx) = \phi$ and hence $(h,T) \in E(x)$. □

Lemma 3.1 is sufficient to establish the qualitative continuity properties of $\mathcal{Q}$ as a function of $x$ expressed in the following propositions.

**Proposition 3.1.** *Assume (A1) and (A2) and let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$. Then $\mathcal{Q}(\cdot,\mu)$ is a real-valued lower semicontinuous function on $\mathbb{R}^{n_1}$.*

*Proof.* By continuity of $\Phi$ it is easily seen that $M(x)$ is an open set and hence measurable for any $x \in \mathbb{R}^{n_1}$. Thus $\mathcal{Q}(\cdot,\mu)$ is well-defined and obviously real-valued. Now, let $x \in \mathbb{R}^{n_1}$ and let $\{x_n\}$ be a sequence in $\mathbb{R}^{n_1}$ converging to $x$. By Proposition 2.1 (a) and Lemma 3.1 (a) we now have

$$\mathcal{Q}(x,\mu) = \mu(M(x)) \leq \mu(\liminf_{n\to\infty} M(x_n)) \leq \liminf_{n\to\infty} \mu(M(x_n)) = \liminf_{n\to\infty} \mathcal{Q}(x_n,\mu).$$

Hence, $\mathcal{Q}(\cdot,\mu)$ is lower semicontinuous at $x$. □

**Proposition 3.2.** *Assume (A1) and (A2) and let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ and $x \in \mathbb{R}^{n_1}$ be such that $\mu(E(x)) = 0$. Then $\mathcal{Q}(\cdot,\mu)$ is continuous at $x$.*

*Proof.* Let $\{x_n\}$ be a sequence in $\mathbb{R}^{n_1}$ converging to $x$. By the assumption $\mu(E(x)) = 0$ we have $\mu(M(x)) = \mu(M(x) \cup E(x))$ and hence by Proposition 2.1 (b) and Lemma 3.1 (b) we get

$$\mathcal{Q}(x,\mu) = \mu(M(x)) \geq \mu(\limsup_{n\to\infty} M(x_n)) \geq \limsup_{n\to\infty} \mu(M(x_n)) = \limsup_{n\to\infty} \mathcal{Q}(x_n,\mu).$$

Hence, observing Proposition 3.1, we see that $\mathcal{Q}(\cdot,\mu)$ is continuous at $x$. □

Recalling the properties of the marginal and conditional distributions of the second-stage right-hand side and the technology matrix listed in Section 2, we obtain the following corollary as an immediate consequence of Proposition 3.2.

**Corollary 3.1.** *Assume (A1) and (A2) and let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ be such that $\mu_1^2(\cdot,T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$ for $\mu_2$-almost all $T$. Then $\mathcal{Q}(\cdot,\mu)$ is a continuous function on $\mathbb{R}^{n_1}$.*

*Proof.* Let $x \in \mathbb{R}^{n_1}$. Under assumptions (A1) and (A2) we have by linear programming duality that

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

where $d_1, \ldots, d_N$ are the vertices of the set $\{u \in \mathbb{R}^m_+ : W^T u \leq q\}$. Hence the set of all those $\tau \in \mathbb{R}^m$ such that $c^T x + \Phi(\tau) = \phi$ is contained in a finite union of hyperplanes $\mathcal{H} := \cup_{j=1}^N H_j$ in $\mathbb{R}^m$ where $H_j := \{\tau \in \mathbb{R}^m : d_j^T \tau = \phi - c^T x\}$. Thus we see that for any $x \in \mathbb{R}^{n_1}$ we have $E(x) \subseteq \{(h,T) \in \mathbb{R}^{m\times(1+n_1)} : h - Tx \in \mathcal{H}\}$ and hence

$$\begin{aligned} \mu(E(x)) &= \int_{\mathbb{R}^{m\times n_1}} \int_{\mathbb{R}^m} \mathbf{1}_{E(x)}(h,T)\, \mu_1^2(dh,T)\, \mu_2(dT) \\ &\leq \int_{\mathbb{R}^{m\times n_1}} \int_{Tx+\mathcal{H}} \mu_1^2(dh,T)\, \mu_2(dT) \\ &= 0, \end{aligned} \tag{3.3}$$

where the last equality follows since the inner integral is equal to zero $\mu_2$-almost surely under the assumption that $\mu_1^2(\cdot,T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$ for $\mu_2$-almost all $T$. Thus we may apply Proposition 3.2 to obtain the desired result. □

We now turn to the joint continuity of $\mathcal{Q}$ as a function of $x$ and $\mu$. As previously mentioned, we will restrict the feasible region of the probability measure $\mu$ to the set of all Borel probability measures on $\mathbb{R}^{m\times(1+n_1)}$ and adopt the notion of weak convergence on this set. To establish joint continuity of the recourse function of a general two-stage stochastic program with mixed-integer recourse, using this notion of convergence, one must further restrict the feasible region of the probability measure $\mu$ to ensure boundedness of certain integrals. As is evident from the following proposition, such a restriction is not necessary for the minimum risk problem.

**Proposition 3.3.** *Assume (A1) and (A2) and let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ and $x \in \mathbb{R}^{n_1}$ be such that $\mu(E(x)) = 0$. Then $\mathcal{Q} : \mathbb{R}^{n_1} \times \mathcal{P}(\mathbb{R}^{m\times(1+n_1)}) \mapsto \mathbb{R}$ is continuous at $(x,\mu)$.*

*Proof.* Let $\{x_n\}$ be a sequence in $\mathbb{R}^{n_1}$ converging to $x$ and let $\{\mu_n\}$ be a sequence in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ converging weakly to $\mu$. First, we introduce functions $f_n : \mathbb{R}^{m\times(1+n_1)} \mapsto \mathbb{R}$ and $f : \mathbb{R}^{m\times(1+n_1)} \mapsto \mathbb{R}$ defined by

$$f_n(h,T) = \psi(x_n,h,T) \quad \text{and} \quad f(h,T) = \psi(x,h,T).$$

Note that all these functions are measurable due to measurability of the sets $M(x_n)$ and $M(x)$, cf. the discussion in Section 2. Also, we define the set $E_0(x)$ consisting of all those

$(h,T) \in \mathbb{R}^{m\times(1+n_1)}$ for which there exists a sequence $\{(h_n,T_n)\}$ in $\mathbb{R}^{m\times(1+n_1)}$ such that $(h_n,T_n) \to (h,T)$ but $f_n(h_n,T_n) \not\to f(h,T)$ as $n \to \infty$.

We will show that $E_0(x) \subseteq E(x)$. So let $(h,T) \in E_0(x)$ and let $\{(h_n,T_n)\}$ be a sequence in $\mathbb{R}^{m\times(1+n_1)}$ converging to $(h,T)$ which satisfies $f_n(h_n,T_n) \not\to f(h,T)$ as $n \to \infty$. Assume that $(h,T) \in M(x)$, i.e. $c^Tx+\Phi(h-Tx) > \phi$. Now, by continuity of $\Phi$ there exists a number $N \in \mathbb{N}$ such that $(h_n,T_n) \in M(x_n)$ for all $n > N$ and hence $\lim_{n\to\infty} f_n(h_n,T_n) = 1 = f(h,T)$, a contradiction. Likewise, the assumption $c^Tx+\Phi(h-Tx) < \phi$ leads to a contradiction and hence $(h,T) \in E(x)$.

Thus, by the assumption $\mu(E(x)) = 0$ we get $\mu(E_0(x)) = 0$ and we may apply Rubin's Theorem (see e.g. Billingsley [6]) to obtain

$$\mu_n \circ f_n^{-1} \xrightarrow{w} \mu \circ f^{-1}, \text{ as } n \to \infty.$$

Now, we introduce a bounded continuous function $g : \mathbb{R} \mapsto \mathbb{R}$, satisfying $g(0) = 0$ and $g(1) = 1$. Note that we have $g(t) = t$, $\mu_n \circ f_n^{-1}$-almost surely for all $n$ and $g(t) = t$, $\mu \circ f^{-1}$-almost surely. Since the $\mu$-integral of two functions which are equal $\mu$-almost surely is the same, we have by the above weak convergence of $\mu_n \circ f_n^{-1}$ to $\mu \circ f^{-1}$ that

$$\int_{\mathbb{R}} t\, \mu_n \circ f_n^{-1}(dt) = \int_{\mathbb{R}} g(t)\, \mu_n \circ f_n^{-1}(dt) \xrightarrow{n\to\infty} \int_{\mathbb{R}} g(t)\, \mu \circ f^{-1}(dt) = \int_{\mathbb{R}} t\, \mu \circ f^{-1}(dt),$$

and we obtain the desired result by changing variables:

$$\int_{\mathbb{R}^{m\times(1+n_1)}} f_n(h,T)\mu_n(d(h,T)) \xrightarrow{n\to\infty} \int_{\mathbb{R}^{m\times(1+n_1)}} f(h,T)\mu(d(h,T)).$$

□

Once again, as an immediate consequence of Proposition 3.3, we obtain the following corollary:

**Corollary 3.2.** *Assume (A1) and (A2) and let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ be such that $\mu_1^2(\cdot,T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$ for $\mu_2$-almost all $T$. Then $\mathcal{Q} : \mathbb{R}^{n_1} \times \mathcal{P}(\mathbb{R}^{m\times(1+n_1)}) \mapsto \mathbb{R}$ is a continuous function on $\mathbb{R}^{n_1} \times \{\mu\}$.*

*Proof.* The proof is similar to that of Corollary 3.1. □

Quantitative continuity of $\mathcal{Q}$ as a function of the underlying measure $\mu$ relies on identifying a (pseudo-) distance on $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ that is properly adjusted to the definition of $\mathcal{Q}$ via probabilities of level sets of value functions. Adapting probability (pseudo-) distances to the underlying structures is a proven tool in quantitative stability analysis of stochastic programs. We refer to Rachev and Römisch [22] for a general framework and applications

to expectation-based recourse models and chance-constrained problems. For our purposes, the following discrepancy defined for $\mu, \nu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ will turn out useful:

$$\alpha_{\mathcal{B}_k}(\mu,\nu) := \sup\big\{|\mu(B) - \nu(B)| : B \in \mathcal{B}_k\big\}.$$

Here, $\mathcal{B}_k \subseteq \mathcal{B}(\mathbb{R}^{m\times(1+n_1)})$ denotes the family of all polyhedra in $\mathbb{R}^{m\times(1+n_1)}$ with at most $k$ faces, $k \in \mathbb{N}$. The discrepancy $\alpha_{\mathcal{B}_k}$ is easily seen to be a pseudometric on $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$. Furthermore, as we will discuss in Section 4, the class $\mathcal{B}_k$ is in fact a Vapnik-Červonenkis class. This fact will allow us to derive conclusions on the stability of optimal solutions for the minimum risk problem when the probability measure $\mu$ is estimated by empirical measures.

**Proposition 3.4.** *Assume (A1) and (A2). Then there exists a $k \in \mathbb{N}$ such that for all $x \in \mathbb{R}^n$ and all $\mu, \nu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ we have*

$$\big|\mathcal{Q}(x,\mu) - \mathcal{Q}(x,\nu)\big| \leq \alpha_{\mathcal{B}_k}(\mu,\nu).$$

*Proof.* Denoting by $M^c(x)$ the complement of $M(x)$ we have for any $x \in \mathbb{R}^n$

$$\begin{aligned} M^c(x) &= \big\{(h,T) \in \mathbb{R}^{m\times(1+n_1)} : c^T x + \Phi(h - Tx) \leq \phi\big\} \\ &= \big\{(h,T) \in \mathbb{R}^{m\times(1+n_1)} : d_j^T h - d_j^T T x \leq \phi - c^T x,\ j = 1,\dots N\big\}, \end{aligned}$$

where, once again, $d_1,\dots,d_N$ are the vertices of the set $\{u \in \mathbb{R}^m : W^T u \leq q\}$. It follows that $\{M^c(x) : x \in \mathbb{R}^n\}$ is a family of polyhedra in $\mathbb{R}^{m\times(1+n_1)}$ whose numbers of facets are bounded above by a uniform constant, i.e. a constant not depending on $x$. Hence, there exists a $k \in \mathbb{N}$ such that

$$\begin{aligned} \big|\mathcal{Q}(x,\mu) - \mathcal{Q}(x,\nu)\big| &= \big|\mu(M(x)) - \nu(M(x))\big| \\ &= \big|\mu(M^c(x)) - \nu(M^c(x))\big| \\ &\leq \sup\big\{|\mu(B) - \nu(B)| \ :\ B \in \mathcal{B}_k\big\}, \end{aligned}$$

and the proof is complete. □

A coherence between the discrepancy $\alpha_{\mathcal{B}_k}$ and weak convergence of probability measures may be established using the concept of a $\mu$-uniformity class defined for some probability measure $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ as follows. A class $\mathcal{B}_0 \subseteq \mathcal{B}(\mathbb{R}^{m\times(1+n_1)})$ is called a $\mu$-uniformity class if $\sup\big\{|\mu_n(B) - \mu(B)| : B \in \mathcal{B}_0\big\} \xrightarrow{n\to\infty} 0$ for every sequence of probability measures $\{\mu_n\}$ in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ converging weakly to $\mu$. According to Theorem 2.11 in Bhattacharya and Ranga Rao [5], the class $\mathcal{B}_c$ of all convex Borel sets in $\mathbb{R}^{m\times(1+n_1)}$ is a $\mu$-uniformity class for all those $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ that are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m\times(1+n_1)}$. Since $\mathcal{B}_k \subseteq \mathcal{B}_c$, we see that $\alpha_{\mathcal{B}_k}(\mu_n,\mu) \xrightarrow{n\to\infty} 0$ for any sequence of probability measures $\{\mu_n\}$ in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ converging weakly to such $\mu$. Thus, if $\mu$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m\times(1+n_1)}$ and $\mu_1^2(\cdot,T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$ for $\mu_2$-almost all $T$, then Proposition 3.4 may be seen as a quantification of the result in Corollary 3.2.

# 4 Stability

In this section, we turn to the issue of stability of optimal solution values and optimal solution sets of the minimum risk problem $MRP(\mu)$ when the underlying probability distribution is subjected to perturbations. Since the minimum risk problem is a non-convex problem, local minimizers should be included in the analysis. To this end, we introduce for any non-empty open set $V \subseteq \mathbb{R}^{n_1}$ a localized version, $\varphi_V : \mathcal{P}(\mathbb{R}^{m\times(1+n_1)}) \mapsto \mathbb{R}$, of the optimal-value function, defined by

$$\varphi_V(\mu) := \inf\{\mathcal{Q}(x,\mu) : x \in X \cap \operatorname{cl} V\},$$

and a localized version, $\Psi_V : \mathcal{P}(\mathbb{R}^{m\times(1+n_1)}) \mapsto \mathbb{R}^{n_1}$, of the solution set mapping, defined by

$$\Psi_V(\mu) := \{x \in X \cap \operatorname{cl} V : \mathcal{Q}(x,\mu) = \varphi_V(\mu)\},$$

where $\operatorname{cl} V$ denotes the closure of $V$. Note that if $X \cap \operatorname{cl} V$ is non-empty and bounded the infimum in the definition of $\varphi_V$ is always attained since we are minimizing a lower semicontinuous function over a compact set, and hence in this case $\Psi_V(\mu)$ is non-empty for any $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$.

Having established the joint continuity of $\mathcal{Q}$ with respect to $x$ and $\mu$ it is straightforward to prove continuity of $\varphi_V$ and Berge upper semicontinuity of $\Psi_V$ for any bounded open set $V \subseteq \mathbb{R}^{n_1}$. (Recall that the point-to-set mapping, $\Psi_V$, is Berge upper semicontinuous at some $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ if for any open set $G \subseteq \mathbb{R}^{n_1}$ with $\Psi_V(\mu) \subseteq G$ there exists some neighbourhood $U$ of $\mu$ in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ such that $\Psi_V(\nu) \subseteq G$ for all $\nu \in U$.) For the analysis of local minimizers, however, we will not find these properties quite sufficient in their own, the shortcoming being that they do not preclude certain pathologies that may occur when dealing with stability of local minimizers. Particularly, one easily constructs examples such that $\Psi_V(\mu)$ is a set of local minimizers of $MRP(\mu)$ for some $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ and $V \subseteq \mathbb{R}^{n_1}$ while for any neighbourhood $U$ of $\mu$ in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ there exists $\nu \in U$ such that $\Psi_V(\nu)$ does not contain any local minimizers of $MRP(\nu)$.

To preclude such pathologies Robinson [25] and Klatte [18] proposed a local stability analysis for non-convex problems, emphasizing the need for considerations to include all local minimizers that are, in some sense, nearby the minimizers one is interested in. The crucial concept is that of a complete local minimizing set, or simply a CLM set, which may be formulated as follows. Let $\mu$ be a Borel probability measure and let $M$ be a non-empty subset of $\mathbb{R}^{n_1}$. If there exists an open set $V \subseteq \mathbb{R}^{n_1}$ such that $M \subseteq V$ and $M = \Psi_V(\mu)$, then $M$ is called a CLM set for $MRP(\mu)$ with respect to $V$. Obvious examples of CLM sets are the set of global minimizers as well as any set of strict local minimizers. Hence, the subsequent propositions stated in general for CLM sets are valid in particular for the set of global minimizers and for any set of strict local minimizers.

**Proposition 4.1.** *Assume (A1) and (A2), let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ be such that $\mu_1^2(\cdot, T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$ for $\mu_2$-almost all $T$ and let $V \subseteq \mathbb{R}^{n_1}$ be a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $MRP(\mu)$ with respect to $V$. Then,*
*(a) $\varphi_V : \mathcal{P}(\mathbb{R}^{m\times(1+n_1)}) \mapsto \mathbb{R}$ is continuous at $\mu$;*
*(b) $\Psi_V : \mathcal{P}(\mathbb{R}^{m\times(1+n_1)}) \mapsto \mathbb{R}^{n_1}$ is Berge upper semicontinuous at $\mu$;*
*(c) There exists some neighbourhood $U$ of $\mu$ in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ such that $\Psi_V(\nu)$ is a CLM set for $MRP(\nu)$ with respect to $V$ for all $\nu \in U$.*

*Proof.* (a) Continuity of $\varphi_V$ is an immediate consequence of Corollary 3.2 and compactness of $X \cap \operatorname{cl} V$, cf. e.g. the proof of Theorem 4.2.2 in Bank et al. [3].
(b) Let $\{\mu_n\}$ be a sequence in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ converging weakly to $\mu$ and let $x_n \in \Psi_V(\mu_n)$ for all $n$ such that the sequence $\{x_n\}$ converges to some $x \in \mathbb{R}^{n_1}$. By Corollary 3.2 and continuity of $\varphi_V$ we now have

$$\mathcal{Q}(x, \mu) = \lim_{n\to\infty} \mathcal{Q}(x_n, \mu_n) = \lim_{n\to\infty} \varphi_V(\mu_n) = \varphi_V(\mu).$$

Thus $x \in \Psi_V(\mu)$ implying that $\Psi_V$ is a closed mapping and hence Berge upper semicontinuous by compactness of $X \cap \operatorname{cl} V$, cf. Lemma 2.2.3 in Bank et al. [3].
(c) By Berge upper semicontinuity of $\Psi_V$ there exists some neighbourhood $U$ of $\mu$ in $\mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ such that $\Psi_V(\nu) \subseteq V$ for all $\nu \in U$. Non-emptyness of $\Psi_V(\nu)$ for $\nu \in U$ follows from non-emptyness of $\Psi_V(\mu)$, boundedness of $V$ and lower semicontinuity of $\mathcal{Q}(\cdot, \nu)$. □

Once again we may quantify the result in Proposition 4.1 (a) using the pseudometric $\alpha_{\mathcal{B}_k}$.

**Proposition 4.2.** *Assume (A1) and (A2), let $\mu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ and let $V \subseteq \mathbb{R}^{n_1}$ be a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $MRP(\mu)$ with respect to $V$. Then there exists a $k \in \mathbb{N}$ such that for all $\nu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$, we have*

$$\left|\varphi_V(\mu) - \varphi_V(\nu)\right| \leq \alpha_{\mathcal{B}_k}(\mu, \nu).$$

*Proof.* Let $\nu \in \mathcal{P}(\mathbb{R}^{m\times(1+n_1)})$ and note once again that $\Psi_V(\nu)$ is non-empty by non-emptyness of $\Psi_V(\mu)$, boundedness of $V$ and lower semicontinuity of $\mathcal{Q}(\cdot, \nu)$. Now, let $x_\mu \in \Psi_V(\mu)$ and $x_\nu \in \Psi_V(\nu)$ and apply Proposition 3.4 to obtain the existence of some $k \in \mathbb{N}$ such that

$$\varphi_V(\mu) - \varphi_V(\nu) \leq \left|Q(x_\nu, \mu) - Q(x_\nu, \nu)\right| \leq \alpha_{\mathcal{B}_k}(\mu, \nu)$$

and

$$\varphi_V(\nu) - \varphi_V(\mu) \leq \left|Q(x_\mu, \nu) - Q(x_\mu, \mu)\right| \leq \alpha_{\mathcal{B}_k}(\mu, \nu).$$

Hence,

$$\left|\varphi_V(\mu) - \varphi_V(\nu)\right| \leq \alpha_{\mathcal{B}_k}(\mu, \nu),$$

and the proof is complete. □

Now, let us briefly discuss an interesting implication of the result in Proposition 4.2 for the asymptotic convergence of optimal solutions to $MRP(\mu)$ when the probability measure $\mu$ is estimated by empirical measures. We let $\{(h_n, T_n)\}$ be a sequence of independent and identically distributed $m \times (1 + n_1)$-dimensional random vectors defined on the probability space $(\Omega, \mathcal{F}, P)$ and consider the corresponding sequence of empirical measures on $\mathbb{R}^{m\times(1+n_1)}$ defined by

$$\mu_n(\omega) := \frac{1}{n}\sum_{i=1}^{n} \delta_{(h_i(\omega),T_i(\omega))} \qquad \text{for } \omega \in \Omega,\ n \in \mathbb{N},$$

where $\delta_{(h_i(\omega),T_i(\omega))}$ denotes the measure with unit mass at $(h_i(\omega), T_i(\omega))$. Also, we denote by $\mu$ the common distribution of the random vectors $(h_n, T_n)$, $n \in \mathbb{N}$. Following the lines of Schultz [34] it is easily seen that the class $\mathcal{B}_k$ is a Vapnik-Červonenkis class, i.e. there exists some number $r$ such that for any finite set $E \subseteq \mathbb{R}^{m\times(1+n_1)}$ with $r$ elements, not all subsets of $E$ are of the form $E \cap B$, $B \in \mathcal{B}_k$. Next, using this fact and establishing measurability of $\alpha_{\mathcal{B}_k}(\mu_n(\omega), \mu)$ as a function of $\omega$, Schultz applies well-known results to show that $\alpha_{\mathcal{B}_k}(\mu_n(\omega), \mu) \xrightarrow{n\to\infty} 0$ for $P$-almost all $\omega \in \Omega$. Hence, if $V \subseteq \mathbb{R}^{n_1}$ is a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $MRP(\mu)$ with respect to $V$, we get by Proposition 4.2 that

$$\varphi_V(\mu_n(\omega)) \xrightarrow{n\to\infty} \varphi_V(\mu) \qquad \text{for } P\text{-almost all } \omega \in \Omega.$$

Furthermore, it is straightforward to establish the CLM property of $\Psi_V(\mu_n(\omega))$ for large $n$ for $P$-almost all $\omega \in \Omega$. For further details we refer to Schultz [34].

# 5 An algorithm for the minimum risk problem

In this section we elaborate a solution procedure for the minimum risk problem. For practical purposes we need to make the following assumptions:

(A3) The first-stage solution set $X$ is non-empty and compact.

(A4) The probability distribution $P$ is discrete and has finite support, say $\Omega = \{\omega^1, \ldots, \omega^S\}$ with corresponding probabilities $P\big(\{\omega^1\}\big) = \pi^1, \ldots, P\big(\{\omega^S\}\big) = \pi^S$.

We shall refer to a possible outcome $(h(\omega^s), T(\omega^s))$ of the random parameters, corresponding to some elementary event $\omega^s \in \Omega$, as a scenario and denote it simply by $(h^s, T^s)$. Note that assumption (A4) may be justified by the results established in the previous section. Thus, according to Proposition 4.1, the solution of a problem with continuous distribution of random parameters may be approximated to any given accuracy by solutions of problems using only discrete distributions.

As already pointed out, the minimum risk problem is equivalent to an expected value problem where the expectation is taken of an appropriately defined indicator function. A possible way to go is to define such an indicator function by (3.2), including a binary variable and an additional constraint in the second stage, and solve the problem by one of the solution procedures developed for the general class of two-stage stochastic programs with mixed-integer recourse. (See e.g. Ahmed, Tawarmalani and Sahimidis [1], Carøe and Schultz [8], Hemmecke and Schultz [11], Løkketangen and Woodruff [19] or Schultz, Stougie and van der Vlerk [35].) Including a binary variable in the second stage, however, drastically increases the problem complexity. In particular, the second-stage value function of the expected value problem is piecewise linear and convex, whereas that of the miminum risk problem would lose these appealing properties. In this section we show how one may avoid the inclusion of a binary second-stage variable by solving the minimum risk problem using a modified version of the L-shaped procedure. The idea is for each scenario $(h^s, T^s)$ to represent the general indicator function $\psi(\cdot, h^s, T^s)$, defined by (3.1), by a binary variable and a number of optimality cuts, similar to those used for ordinary two-stage stochastic linear programs. By not applying the definition of the indicator function given by (3.2), we obtain a formulation in which integer variables occur only in a master problem and not in the second-stage subproblems.

Given a first-stage solution $x \in X$ and a scenario $(h^s, T^s)$, the optimality cuts needed to represent $\psi(\cdot, h^s, T^s)$ at $x$ are derived considering the following linear programming problem:

$$\min\{e^T t + t_0 : Wy + It \geq h^s - T^s x,\ q^T y - t_0 \leq \phi - c^T x,\ y \in \mathbb{R}^{n_2}_+,\ (t, t_0) \in \mathbb{R}^{m+1}_+\} \quad (5.1)$$

and the corresponding dual problem

$$\max\{(h^s - T^s x)^T u + (c^T x - \phi)u_0 : W^T u - u_0 q \leq 0,\ Iu \leq e,\ u_0 \leq 1, (u, u_0) \in \mathbb{R}^{m+1}_+\} \quad (5.2)$$

where $e \in \mathbb{R}^m$ is a vector of 1's and $I \in \mathbb{R}^{m \times m}$ is the identity matrix. Note that both the primal problem (5.1) and the dual problem (5.2) are always feasible and have optimal value equal to zero if and only if $\psi(x, h^s, T^s) = 0$. In the following we will denote by $D$ the feasible region of the dual problem (5.2). Also, we let $M_2 > 0$ be some large number bounding from above the optimal value of the dual problem:

$$M_2 \geq \sup\Big\{(h(\omega) - T(\omega)x)^T u + (c^T x - \phi)u_0 : x \in X,\ (u, u_0) \in D,\ \omega \in \Omega\Big\}.$$

Note once again that, employing assumptions (A3) and (A4), the supremum exists and is finite, since $D$ is obviously bounded.

The following lemmas which elucidate the basic structure of the optimality cuts are an immediate consequence of the definition of $M_2$ and the previously discussed relationship between the optimal value of the problems (5.1) and (5.2) and the indicator function (3.1).

**Lemma 5.1.** *Let $(u, u_0) \in D$ and $x \in X$. Then for any scenario $(h^s, T^s)$ the indicator function $\psi(x, h^s, T^s)$ should satisfy the following inequality:*

$$M_2\psi(x, h^s, T^s) \geq (h^s - T^s x)^T u + (c^T x - \phi)u_0.$$

**Lemma 5.2.** *Let $x \in X$ be such that $\psi(x, h^s, T^s) = 1$ for some scenario $(h^s, T^s)$. Then there exists an extreme point $(u, u_0) \in D$ such that $(h^s - T^s x)^T u + (c^T x - \phi)u_0 > 0$.*

Using Lemma 5.1 and Lemma 5.2 it is easily seen that the minimum risk problem is equivalent to the following mixed-integer program:

$$\begin{aligned} MRP \qquad z^* = \min & \sum_{s=1}^{S} \pi^s \theta^s \\ \text{s.t. } & (h^s - T^s x)^T u + (c^T x - \phi)u_0 \leq M_2\theta^s \quad \forall (u, u_0) \in D, \\ & x \in X,\ \theta \in \mathbb{B}^S. \end{aligned}$$

The algorithm progresses by sequentially solving a master problem and adding violated optimality cuts generated through the solution of subproblems (5.1). Assuming that the number of optimality cuts generated before iteration $\nu$ is $J_\nu$, the current master problem is:

$$\begin{aligned} z_\nu = \min & \sum_{s=1}^{S} \pi^s \theta^s \\ \text{s.t. } & (h^s - T^s x)^T u^j + (c^T x - \phi)u_0^j \leq M_2\theta^s \quad \forall j = 1, \ldots, J_\nu, \\ & x \in X,\ \theta \in \mathbb{B}^S. \end{aligned} \tag{5.3}$$

We are now in a position to present a modified version of the L-shaped algorithm for the minimum risk problem.

**Algorithm 1**

**Step 1** (*Initialization*) Set $\nu = 0$ and $J_0 = 0$.

**Step 2** (*Solve master problem*) Solve the current master problem (5.3) and let $(x_\nu, \theta_\nu)$ be an optimal solution vector.

**Step 3** (*Solve subproblems*) Solve the second-stage problem (5.1) corresponding to all scenarios for which $\theta_\nu^s = 0$. Consider the following situations:

1. If all of these problems have optimal value equal to zero, then the current solution $x_\nu$ is optimal for the minimum risk problem (1.3) and $z^* = z_\nu$.
2. If some, say $k$, of these problems have optimal value greater than zero, then an equal number of extreme points $(u^j, u_0^j) \in D$, each of which satisfy $(h^s - T^s x_\nu)^T u^j + (c^T x_\nu - \phi)u_0^j > 0$ for some scenario $(h^s, T^s)$, are identified and the corresponding violated optimality cuts are added to the master. Set $J_{\nu+1} = J_\nu + k$ and $\nu = \nu + 1$; go to step 2.

It is easily seen that Algorithm 1 terminates finitely:

**Proposition 5.1.** *Assume (A1)-(A4). Then Algorithm 1 terminates with an optimal solution in a finite number of iterations.*

*Proof.* By assumption (A3) and Proposition 3.1 an optimal solution of the minimum risk problem exists. Let $x^*$ be one such solution. First of all note that the optimal solution value $z_\nu$ of the master problem in iteration $\nu$ is a lower bound on the optimal solution value of the minimum risk problem, since the master problem is a relaxation of $MRP$:

$$\sum_{s=1}^{S} \pi^s \theta_\nu^s = z_\nu \leq z^* = \sum_{s=1}^{S} \pi^s \psi(x^*, h^s, T^s).$$

Now, suppose in some iteration $\nu$ for some scenario $(h^s, T^s)$ that $\theta_\nu^s < \psi(x_\nu, h^s, T^s)$. By Lemma 5.2 a violated optimality cut, cutting off the current solution $(x_\nu, \theta_\nu)$, is identified in step 3 and the algorithm proceeds. Since the number of dual extreme points is finite this can only happen a finite number of times and we will eventually have

$$\theta_\nu^s \geq \psi(x_\nu, h^s, T^s) \quad s = 1, \ldots, S,$$

at which point the current solution, $x_\nu$ is optimal. □

Note that Algorithm 1 works equally well for problems which do not satisfy the (relative) complete recourse property. The optimal first-stage solution determined by the algorithm, however, may not guarantee feasibility for all second-stage problems in this case. Still, the algorithm may easily be modified to accomodate the possible requirement of feasibility of all second-stage problems by using feasibility cuts as in the original L-shaped algorithm.

# 6 Computational experiments

Algorithm 1 was implemented in C++ using procedures from the callable library of CPLEX 7.0. As previously pointed out, the algorithm is similar in spirit to the ordinary L-shaped algorithm and hence it is bound to suffer from some of the same drawbacks. Of particular importance in this respect, is the fact that early iterations will usually be quite inefficient since solutions tend to oscillate heavily. This deplorable behaviour may be surmounted by adding to the master objective a regularizing term, penalizing divergence from the current solution, cf. the regularized decomposition procedure introduced in the setting of the expected value problem by Ruszczynski [30]. Furthermore, adding a regularizing term to the master objective potentially allows the algorithm to take advantage of a starting solution $x_0$. Regularized decomposition as introduced by Ruszczynski as well as most bundle methods for

nonsmooth optimization, gain considerable advantage by adding a quadratic penalty term to the objective. To avoid a nonlinear mixed-integer formulation of the master problem, however, we chose the following linear formulation:

$$
\begin{aligned}
z_\nu = \min \; & \sum_{s=1}^{S} \pi^s \theta^s + \gamma e^T d \\
\text{s.t. } & (h^s - T^s x)^T u^j + (c^T x - \phi) u_0^j \leq M_2 \theta^s \quad \forall j = 1, \ldots, J_\nu, \\
& d \geq x - x_{\nu-1}, \\
& d \geq x_{\nu-1} - x, \\
& x \in X, \; \theta \in \mathbb{B}^S, \; d \in \mathbb{R}^{n_1}_+.
\end{aligned}
\tag{6.1}
$$

where $e \in \mathbb{R}^{n_1}$ is a vector of 1's and $\gamma$ is a scaling factor. Because of the mixed-integer nature of the minimum risk problem, the inclusion of the regularizing term is not theoretically justified as in regularized decomposition, in the sense that convergence of optimal solutions of problem (6.1) to an optimal solution of the minimum risk problem cannot be established in general. In practice, we chose to start off the algorithm with the regularizing term included in the objective and remove the regularization once solutions had stabilized. In most cases the algorithm terminated with an optimal solution after just one additional iteration but on a few occasions several iterations had to be performed after removing the regularization.

To investigate the practicability of Algorithm 1, we used three sets of problem instances subsequently referred to as EPS1, EPS2 and EPS3, respectively. The problems were obtained as linear programming relaxations of certain mixed-integer programs arising as scheduling problems in chemical production. We ran the algorithm with varying number of scenarios $S$ as well as varying values of the threshold value $\phi$. Each run was performed with two different versions of the algorithm. In the following, MRP refers to the algorithm as presented in Section 5 while MRPREG refers to the algorithm using the regularization of the master problem described above. At termination of each run we recorded the optimal value, the number of iterations performed, the number of generated cuts and the CPU time spent by the procedure. Finally, we fixed the first-stage variables at their values from the optimal solution of the expected value problem to calculate what we refer to as the value of the EVP solution (VEVP). All computational experiments were carried out on a SUN Enterprise 450, 300 MHz Ultra-SPARC.

Let us first consider the EPS1 instance. This problem contains 2 constraints and 3 variables in the first stage and 50 constraints and 51 variables in the second stage. For this instance we ran the algorithm with 20, 50, 100, 200 and 500 scenarios and each time we chose the threshold value close to the optimal value of the expected value problem. Results are reported in Table 1.

Table 1: Computational results for the EPS1 instance

| | | | | MRP | | | MRPREG | | |
|---|---|---|---|---|---|---|---|---|---|
| $S$ | $\phi$ | Opt. | VEVP | Ite. | Cuts | CPU | Ite. | Cuts | CPU |
| 20 | 18.9 | 0.250 | 0.250 | 5 | 73 | 0.31 | 6 | 58 | 0.29 |
| 50 | 50.9 | 0.120 | 0.140 | 5 | 110 | 0.51 | 7 | 110 | 0.63 |
| 100 | 32.7 | 0.860 | 0.860 | 4 | 300 | 0.88 | 6 | 300 | 0.93 |
| 200 | 28.4 | 0.115 | 1.000 | 8 | 940 | 8.15 | 8 | 761 | 6.55 |
| 500 | 757.5 | 0.048 | 0.048 | 5 | 1518 | 6.56 | 6 | 1518 | 6.62 |

Next we turn to the EPS2 instance. This problem contains 5 constraints and 12 variables in the first stage and 157 constraints and 164 variables in the second stage. For this instance we always used 100 scenarios and solved the problem for a number of different threshold values surrounding the optimal value of the expected value problem which was 65.4. Results are reported in table 2.

Table 2: Computational results for the EPS2 instance

| | | | | MRP | | | MRPREG | | |
|---|---|---|---|---|---|---|---|---|---|
| $S$ | $\phi$ | Opt. | VEVP | Ite. | Cuts | CPU | Ite. | Cuts | CPU |
| 100 | 60 | 0.91 | 0.92 | 19 | 851 | 36.98 | 14 | 728 | 22.93 |
| 100 | 63 | 0.81 | 0.81 | 19 | 1098 | 61.11 | 15 | 920 | 28.51 |
| 100 | 65 | 0.71 | 0.71 | 18 | 1146 | 196.96 | 14 | 868 | 53.77 |
| 100 | 67 | 0.17 | 0.17 | 24 | 1811 | 34.46 | 12 | 880 | 8.54 |
| 100 | 70 | 0.08 | 0.08 | 18 | 1322 | 12.12 | 13 | 782 | 10.50 |

Finally, we consider the EPS3 instance. This problem contains 9 constraints and 30 variables in the first stage and 280 constraints and 326 variables in the second stage. Once again, we always used 100 scenarios and solved the problem for a number of different threshold values surrounding the optimal value of the expected value problem which in this case was 191.3. Results are reported in table 3.

Table 3: Computational results for the EPS3 instance

| | | | | MRP | | | MRPREG | | |
|---|---|---|---|---|---|---|---|---|---|
| $S$ | $\phi$ | Opt. | VEVP | Ite. | Cuts | CPU | Ite. | Cuts | CPU |
| 100 | 170 | 0.94 | 0.97 | 41 | 1407 | 8164.01 | 46 | 1093 | 1858.37 |
| 100 | 180 | 0.85 | 0.88 | 45 | 1418 | 5564.62 | 30 | 1326 | 4632.01 |
| 100 | 190 | 0.59 | 0.60 | 37 | 1996 | 4716.74 | 25 | 1328 | 2019.36 |
| 100 | 200 | 0.13 | 0.14 | 44 | 2317 | 178.90 | 24 | 1280 | 88.26 |
| 100 | 210 | 0.02 | 0.02 | 62 | 2656 | 142.89 | 23 | 1455 | 36.30 |

We note that the optimal value of the expected value problem of the EPS1 instance with

200 scenarios was 28.7 with all scenarios having costs above 28.4 and hence the seemingly strange result for this instance reported in table 1 was obtained. Apart from this instance, however, the value of the EVP solution is always relatively close to the optimal value of the minimum risk problem and hence the gain of solving the minimum risk problem rather than the expected value problem is negligible for the instances considered here. We did, however, also test the algorithm on the linear programming relaxation of a small stochastic program, previously used as test instance in the papers by Carøe and Schultz [8] and Schultz, Stougie and van der Vlerk [35], and for this problem the gain of solving the minimum risk problem rather than the expected value problem was more significant as is evident from table 4. We should mention that this instance has 2 variables and no constraints in the first stage and 4 variables and 2 constraints in the second stage.

Table 4: Computational results for a small test instance

| | | | | MRP | | | MRPREG | | |
|---|---|---|---|---|---|---|---|---|---|
| $S$ | $\phi$ | Opt. | VEVP | Ite. | Cuts | CPU | Ite. | Cuts | CPU |
| 4 | 50 | 0.500 | 0.750 | 3 | 3 | 0.07 | 4 | 3 | 0.10 |
| 9 | 50 | 0.111 | 0.556 | 4 | 9 | 0.09 | 5 | 9 | 0.08 |
| 36 | 50 | 0.167 | 0.306 | 3 | 17 | 0.11 | 4 | 17 | 0.12 |
| 121 | 50 | 0.132 | 0.182 | 5 | 58 | 0.34 | 6 | 58 | 0.38 |
| 441 | 50 | 0.120 | 0.152 | 6 | 216 | 1.65 | 8 | 216 | 2.06 |

# 7 Conclusions

In the setting of two-stage stochastic recourse programs we have sought a problem formulation containing a more adequate description of risk aversion than that of the accustomed expected value problem. This lead to the formulation of the minimum risk problem in which the probability of total cost exceeding a certain threshold value is minimized. The recourse function $\mathcal{Q}(\cdot, \mu)$ of this problem was shown to be well-defined and lower semicontinuous under basic assumptions guaranteeing feasibility and boundedness of second-stage problems (Proposition 3.1). Furthermore, assuming that the probability of random parameters yielding optimal cost equal to the threshold value is zero, the function $\mathcal{Q}(\cdot, \mu)$ was shown to be continuous (Proposition 3.2). Equipping the space of underlying probability measures with the notion of weak convergence, this assumption is in fact sufficient for the joint continuity of $\mathcal{Q}$ as a function of the first-stage decision and the probability measure (Proposition 3.3). Having established the joint continuity of $\mathcal{Q}$ it is a small step to arrive at stability results for the minimum risk problem in the form of continuity of local optimal values and upper

semicontinuity of local optimal solutions (Proposition 4.1). Finally, the results of Proposition 3.3 and 4.1 were quantified using a certain variational distance of probability measures (Propositions 3.4 and 4.2).

Assuming that the distribution of random parameters is discrete, an assumption which is justified by the results in Proposition 4.1, we elaborated an algorithm for the minimum risk problem. The algorithm was shown to terminate with an optimal solution in a finite number of iterations under the additional assumption of a compact first-stage solution set (Proposition 5.1) and preliminary computational experiments exhibit promising results, in particular when a regularizing mechanism is incorporated.

# References


[1] S. Ahmed, M. Tawarmalani, and N.V. Sahimidis. A finite branch and bound algorithm for two-stage stochastic integer programs. Submitted to Mathematical Programming. Available at http://www.isye.gatech.edu/~sahmed, 2000.

[2] Z. Artstein and R.J-B Wets. Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM Journal on Optimization*, 4:537–550, 1994.

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

[4] B. Bereanu. Minimum risk criterion in stochastic optimization. *Economic Computation and Economic Cybernetics Studies and Research*, 2:31–39, 1981.

[5] R.N. Bhattacharya and R. Ranga Rao. *Normal Approximation and Assymptotic Expansions*. John Wiley, New York, 1976.

[6] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[7] J.R. Birge and F.V. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, Berlin Heidelberg New York, 1997.

[8] C.C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24:37–45, 1999.

[9] J. Dupačová. Stochastic programming with incomplete information: A survey of results on postoptimization and sensitivity analysis. *Optimization*, 18:507–532, 1987.

[10] W.K. Klein Haneveld and M.H. van der Vlerk. On the expected value function of a simple integer recourse problem with random technology matrix. *Journal of Computational and Applied Mathematics*, 56:45–53, 1994.

[11] R. Hemmecke and R. Schultz. Decomposition of test sets in stochastic integer programming. Technical report, Gerhard Mercator Universität Duisburg, 2000.

[12] J.L. Higle and S. Sen. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16:650–669, 1991.

[13] J. Hoffmann-Jørgensen. *Probability with a View Toward Statistics*, volume 1. Chapman & Hall, New York, 1994.

[14] P. Kall. On approximations and stability in stochastic programming. In J. Guddat, H.Th. Jongen, B. Kummer, and F. Nožička, editors, *Parametric Optimization and Related Topics*, pages 387–407. Akademie Verlag, Berlin, 1987.

[15] P. Kall and S.W. Wallace. *Stochastic Programming.* John Wiley and Sons, Chichester, UK, 1994.

[16] A.H.G. Rinnooy Kan and L. Stougie. Stochastic integer programming. In Y. Ermoliev and R.J-B Wets, editors, *Numerical Techniques for Stochastic Optimization*, pages 201–213. Springer-Verlag, Berlin, 1988.

[17] A.I. Kibzun and Y.S. Kan. *Stochastic Programming Problems with Probability and Quantile Functions.* John Wiley and Sons, Chichester, UK, 1996.

[18] D. Klatte. A note on quantitative stability results in nonlinear optimization. In K. Lommatzsch, editor, *Proceedings of the 19. Jahrestagung Mathematische Optimierung*, Seminarbericht Nr. 90, pages 77–86, Sektion Mathematik, Humboldt-Universität Berlin, 1987.

[19] A. Løkketangen and D.L. Woodruff. Progressive hedging and tabu search applied to mixed integer (0,1) multistage stochastic programming. *Journal of Heuristics*, 2:111–128, 1996.

[20] F.V. Louveaux and M.H. van der Vlerk. Stochastic programming with simple integer recourse. *Mathematical Programming*, 61:301–325, 1993.

[21] A. Prékopa. *Stochastic Programming.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1995.

[22] S.T. Rachev and Wömisch. Quantitative stability in stochastic programming: The method of probability metrics. Preprint 2000-22, Humboldt-Universität zu Berlin, Institut für Mathematik, 2000.

[23] E. Raik. Qualitative research into the stochastic nonlinear programming problems. *Eesti NSV Teaduste Akademia Toimetised (News of the Estonian Academy of Sciences) Füüs Mat.*, 20:8–14, 1971. In Russian.

[24] E. Raik. On the stochastic programming problem with the probability and quantile functionals. *Eesti NSV Teaduste Akademia Toimetised (News of the Estonian Academy of Sciences) Füüs Mat.*, 21:142–148, 1972. In Russian.

[25] S.M. Robinson. Local epi-continuity and local optimization. *Mathematical Programming*, 37:208–222, 1987.

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

[27] W. Römisch and R. Schultz. Distribution sensitivity in stochastic programming. *Mathematical Programming*, 50:197–226, 1991.

[28] W. Römisch and R. Schultz. Stability analysis for stochastic programs. *Annals of Operations Research*, 30:241–266, 1991.

[29] W. Römisch and R. Schultz. Stability of solutions for stochastic programs with complete recourse. *Mathematics of Operations Research*, 18:590–609, 1993.

[30] A. Ruszczynski. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35:309–333, 1986.

[31] R. Schultz. Continuity and stability in two-stage stochastic integer programming. In K. Marti, editor, *Stochastic Optimization, Numerical Methods and Technical Applications*, volume 379 of *Lecture Notes in Economics and Mathematical Systems*, pages 81–92. Springer-Verlag, Berlin, 1992.

[32] R. Schultz. Continuity properties of expectation functions in stochastic integer programming. *Mathematics of Operations Research*, 18:578–589, 1993.

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

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

[35] R. Schultz, L. Stougie, and M.H. van der Vlerk. Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions. *Mathematical Programming*, 83:229–252, 1998.

[36] A. Shapiro. Quantitative stability in stochastic programming. *Mathematical Programming*, 67:99–108, 1994.

[37] R.M. Van Slyke and R.J-B Wets. L-shaped linear programs with applications to optimal control and stochastic linear programming. *SIAM Journal of Applied Mathematics*, 17:638–663, 1969.

[38] L. Stougie. *Design and analysis of algorithms for stochastic integer programming*. CWI Tract, 37. Center for Mathematics and Computer Science, Amsterdam, 1987.

[39] R.J-B Wets. Stochastic programs with fixed recourse: the equivalent deterministic program. *SIAM Review*, 16:309–339, 1974.