



# Convex approximations of two-stage risk-averse mixed-integer recourse models

E. Ruben van Beesten<sup>1</sup> · Ward Romeijnders<sup>1</sup> · Kees Jan Roodbergen<sup>1</sup>

Received: 14 September 2021 / Accepted: 13 January 2024 / Published online: 13 February 2024  
© The Author(s) 2024

## Abstract

We consider two-stage risk-averse mixed-integer recourse models with law invariant coherent risk measures. As in the risk-neutral case, these models are generally non-convex as a result of the integer restrictions on the second-stage decision variables and hence, hard to solve. To overcome this issue, we propose a convex approximation approach. We derive a performance guarantee for this approximation in the form of an asymptotic error bound, which depends on the choice of risk measure. This error bound, which extends an existing error bound for the conditional value at risk, shows that our approximation method works particularly well if the distribution of the random parameters in the model is highly dispersed. For special cases we derive tighter, non-asymptotic error bounds. Whereas our error bounds are valid only for a continuously distributed second-stage right-hand side vector, practical optimization methods often require discrete distributions. In this context, we show that our error bounds provide statistical error bounds for the corresponding (discretized) sample average approximation (SAA) model. In addition, we construct a Benders' decomposition algorithm that uses our convex approximations in an SAA-framework and we provide a performance guarantee for the resulting algorithm solution. Finally, we perform numerical experiments which show that for certain risk measures our approach works even better than our theoretical performance guarantees suggest.

**Keywords** Stochastic programming · Mixed-integer recourse · Risk aversion · Convex approximations

---

✉ E. Ruben van Beesten  
vanbeesten@ese.eur.nl

Ward Romeijnders  
w.romeijnders@rug.nl

Kees Jan Roodbergen  
k.j.roodbergen@rug.nl

<sup>1</sup> Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE Groningen, The Netherlands

## 1 Introduction

Recourse models are a class of models within the field of stochastic optimization that deal with decision-making problems under uncertainty (see, e.g., [1]). The explicit way in which these models incorporate uncertainty make them suitable for a wide array of practical problems, ranging from energy optimization [2] to scheduling [3] and vehicle routing problems [4]. In this paper, we focus on (two-stage) risk-averse mixed-integer recourse (MIR) models. These are a class of recourse models with two decision stages that combine two difficulties: *integer restrictions* on decision variables and *risk aversion*. Integer restrictions are often necessary for accurate modeling of practical problem aspects, such as on/off decisions, indivisibilities, or fixed batch sizes. Furthermore, risk aversion is an important aspect of preferences of decision makers and is reflected in the way uncertainty is captured in the model. Compared with risk-neutral models (in which the *expected* total cost is minimized), risk-averse models put more emphasis on avoiding extremely undesirable outcomes.

Unfortunately, risk-averse MIR models are notoriously hard to solve. In particular, the inclusion of integer restrictions makes these models significantly harder to solve than their continuous counterparts. Indeed, for a wide class of risk measures used in practice, risk-averse *continuous* recourse models are convex, and efficient solution methods have been developed in the literature. These methods exploit convexity of the objective function and make use of solution methods based on convex optimization. See, e.g., Ahmed [5], Miller and Ruszczyński [6] and Noyan [7] for decomposition methods and Rockafellar [8] for a progressive hedging algorithm.

Risk-averse *mixed-integer* recourse models, however, are generally not convex. As a result, efficient methods from convex optimization cannot be applied and hence, alternative methods are required. In the literature, some authors have proposed solution methods based on large-scale mixed-integer linear programming (MILP) reformulations [9, 10] or heuristic methods for specific problem settings [11]. However, these methods only work for problems with a specific structure (e.g., disaster relief planning [11] or location-allocation problems [10]) or a particular choice of risk measure (for instance, CVaR MIR models can be reformulated as risk-neutral MIR models with an additional auxiliary variable [9]) and of a limited size (since they are based on computationally demanding non-convex optimization techniques [9]). To the best of our knowledge, efficient solution methods for general large-scale risk-averse MIR models are still lacking.

In this paper we take an alternative approach: rather than aiming for an exact solution, we propose to solve a *convex approximation* instead. That is, we approximate the objective function by a convex function. The resulting convex approximation model is amenable to convex optimization techniques and therefore much easier to solve. This idea has been developed over the past few decades for risk-neutral MIR models first [12–21], and recently also for risk-averse MIR models with a mean-CVaR risk measure [22]. In this paper we extend this literature to the more general class of risk-averse MIR models with an arbitrary law invariant coherent risk measure.

Our first contribution is the derivation of a performance guarantee for convex approximations of risk-averse MIR models. In particular, we derive an error bound: an upper bound on the approximation error. We do so by extending the error bound

from [22] for the case with CVaR to our more general setting by using the Kusuoka representation [23] of law invariant coherent risk measures. Like the error bounds from the literature on risk-neutral MIR models [14, 17], this error bound is asymptotic in nature and depends on the distribution of the second-stage right-hand side vector. If the distribution is highly dispersed, then the error bound is small and hence, our convex approximation is good. In comparison with the error bounds from [14] and [17], our error bound contains an additional factor associated with the risk measure  $\rho$ : the error bound is higher if the maximum probability distortion factor associated with  $\rho$  is higher. For the special case of risk-averse totally unimodular integer recourse models, we derive a specialized, tighter error bound whose dependence on the risk measure  $\rho$  is more favorable.

Our second contribution is the extension of our error bounds to a setting where the distribution of the second-stage right-hand side vector is discrete. Our original error bounds are only valid if this right-hand side vector has a continuous distribution. In practice, however, distributions are often discretized to avoid the numerically demanding task of computing multi-dimensional integrals. A particular popular method for discretization is sample average approximation (SAA) [24]. We show that our error bounds translate to *statistical* error bounds if the distribution is discretized using SAA. This result not only holds for the error bounds in our paper, but carries over to many error bounds for convex approximations of risk-neutral MIR models as well (e.g., those in [12, 14–17, 22]).

Our third contribution is the construction of an algorithm based on our convex approximations that finds a near-optimal solution for risk-averse MIR models. Such an algorithm is warranted, because solving the convex approximation model is not a trivial task due to the complicated way in which it is defined. We formulate an algorithm for our risk-averse setting by adapting the loose benders decomposition algorithm (LBDA) by van der Laan and Romeijnnders [17], which was originally defined for a risk-neutral setting. This algorithm uses a Benders' decomposition approach with so-called "loose" optimality cuts in an SAA framework. In our extension to the risk-averse case we pay particular attention to a linearization of the risk-averse objective function. To guarantee performance of the resulting algorithmic solution, we provide an asymptotic performance guarantee, based on the error bound for our convex approximation.

Our fourth contribution consists of a numerical analysis of the performance of our convex approximation approach. In the experiments, we focus on the relation between the choice of risk measure and the performance of our convex approximations and our algorithm. We show that in some cases, both the approximations themselves and the algorithm perform better than our performance guarantees suggest, in particular as a function of the risk measure  $\rho$ . This highlights the fact that our performance guarantees are *worst-case* error bounds. Moreover, we find evidence that our approach performs particularly well if the risk measure has an associated *risk spectrum* that is "smooth".

The remainder of this paper is organized as follows. In Sect. 2, we mathematically define risk-averse MIR models and their convex approximations. In Sect. 3, we derive a corresponding error bound. In Sect. 4, we derive a specialized bound for risk-averse totally unimodular integer models and we analyze an example with simple integer recourse. In Sect. 5, we show that the error bounds from Sects. 3 and 4 constitute statistical error bounds for the corresponding SAA models. In Sect. 6, we extend

the LBDA to our risk-averse setting and we derive a performance guarantee on the resulting solution. Section 7 contains our numerical experiments. Finally, in Sect. 8, we conclude and discuss directions for future research.

## 2 Problem definition

In this section we mathematically define risk-averse MIR models and we construct a convex approximation.

### 2.1 Risk-averse MIR model

We consider risk-averse MIR models of the form

$$\min_{x \in X} \{c^T x + \underbrace{\rho(v(\xi, x))}_{Q(x)}\}. \quad (1)$$

Here,  $x$  is the first-stage decision vector with unit cost  $c$ ,<sup>1</sup> to be chosen from a feasible set of the form  $X := \{x \in \mathbb{R}_+^{n_1} \mid Ax = b\}$  before learning the realization of the random vector  $\xi$ , defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with support  $\Xi$ . The value function  $v(\xi, x)$ , which we define below, represents the second-stage cost as a function of the first-stage decision  $x$  and the realization of the random vector  $\xi$ . Since the second-stage cost  $v(\xi, x)$  is a random variable, a risk measure  $\rho$  is employed that maps this random variable to a real number for every  $x \in X$ ; the resulting function  $Q$  is called the recourse function.

Formally, we define a risk measure as a function  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$ , on the space  $\mathcal{Z} := L_1(\Omega, \mathcal{F}, \mathbb{P})$  of random variables  $Z$  for which  $\mathbb{E}[|Z|] < +\infty$ . We make a number of assumptions ensuring that  $\rho$  “reasonably” quantifies risk. Specifically, we assume that the risk measure  $\rho$  is *law invariant* [23] and *coherent* [25]. Moreover, in some cases we assume that  $\rho$  is also *comonotone additive* [23], in which case we refer to  $\rho$  as a *spectral risk measure* [26].

**Definition 1** A risk measure  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  is said to be

- (1) *Law invariant* if  $\rho(Z) = \rho(Z')$  whenever  $Z, Z' \in \mathcal{Z}$  have the same cumulative distribution function;
- (2) *Coherent* if it satisfies the following axioms:
  - (i) *Convexity*:  $\rho(tZ + (1-t)Z') \leq t\rho(Z) + (1-t)\rho(Z')$ , for all  $Z, Z' \in \mathcal{Z}$ ,
  - (ii) *Monotonicity*:  $\rho(Z) \geq \rho(Z')$ , for all  $Z, Z' \in \mathcal{Z}$  satisfying  $Z \geq Z'$  a.s.,
  - (iii) *Translation equivariance*:  $\rho(Z+a) = \rho(Z) + a$ , for every  $Z \in \mathcal{Z}$  and  $a \in \mathbb{R}$ ,
  - (iv) *Positive homogeneity*:  $\rho(tZ) = t\rho(Z)$ , for all  $Z \in \mathcal{Z}$  and  $t > 0$ ;

<sup>1</sup> In fact, our main result in Theorem 1 also hold for the more general case with a random first-stage cost vector  $c(\xi)$ , as long as  $c(\xi)$  and  $h(\xi)$  are independent. However, for ease of presentation, we use a deterministic cost vector in this paper.

- (3) *Comonotone additive* if  $\rho(Z + Z') = \rho(Z) + \rho(Z')$  whenever  $Z, Z' \in \mathcal{Z}$  are comonotone random variables, i.e., whenever  $(Z, Z')$  has the same distribution as  $(F_Z^{-1}(U), F_{Z'}^{-1}(U))$ , where  $F_Z$  and  $F_{Z'}$  are the cumulative distribution functions of  $Z$  and  $Z'$ , respectively, and  $U$  is a uniformly distributed random variable on  $(0, 1)$ ;
- (4) *Spectral* if it is law invariant, coherent, and comonotone additive.

The second-stage value function  $v$  in (1) is defined by

$$v(\xi, x) := \min_{y \in Y} \{q(\xi)^T y \mid T(\xi)x + Wy = h(\xi)\}, \quad \xi \in \Xi, x \in X. \tag{2}$$

Here,  $y$  is a decision vector representing the second-stage *recourse* actions, to be decided after deciding on the first-stage decision vector  $x$  and observing the realization of the random vector  $\xi$  with range  $\Xi$ . The set  $Y := \mathbb{Z}_+^{n_2} \times \mathbb{R}_+^{\tilde{n}_2}$  models integer restrictions on some of the elements of  $y$ , hence the term *mixed-integer* recourse. The recourse actions come at a unit cost of  $q(\xi)$  and need to satisfy the constraint  $T(\xi)x + Wy = h(\xi)$ . Note that  $q, T$ , and  $h$  are allowed to depend on  $\xi$ . In the remainder of the paper, we write  $\xi = (q, T, h)$  with range  $\Xi = \Xi^q \times \Xi^T \times \Xi^h$ , and we drop the explicit dependence of  $q(\xi), T(\xi)$ , and  $h(\xi)$  on  $\xi$ , i.e., we simply write  $q, T$ , and  $h$ .

Throughout this paper we make the following assumptions.

**Assumption 1** We assume that

- (a) the recourse is relatively complete and sufficiently expensive, i.e.,  $-\infty < v(\xi, x) < \infty$ , for all  $\xi \in \Xi$  and  $x \in \mathbb{R}^{n_1}$ ,
- (b) the expectations of the  $\ell_1$  norm of  $h, q$ , and  $T$  are finite, i.e.,  $\mathbb{E}^{\mathbb{P}}[\|h\|_1] < \infty, \mathbb{E}^{\mathbb{P}}[\|q\|_1] < \infty$ , and  $\mathbb{E}^{\mathbb{P}}[\|T\|_1] < \infty$ ,
- (c) the recourse matrix  $W$  is integer,
- (d)  $h$  is continuously distributed with joint pdf  $f$ ,
- (e)  $(q, T)$  and  $h$  are pairwise independent.

Assumptions 1(a)–(b) ensure that the recourse function  $Q(x)$  is finite for every  $x \in \mathbb{R}^{n_1}$ . Assumption 1(c) is required for the derivation of our error bounds in Sect. 3 and 4. However, this assumption is not very restrictive, since any rational matrix can be transformed into an integer matrix by appropriate scaling. Assumption 1(d) restricts the right-hand side vector  $h$  to be *continuously* distributed. This is essential for the derivation of error bounds, which depend on the total variations of the pdf  $f$  of  $h$  (although recent numerical and theoretical results in [17, 27] suggest that the interpretation of total variation error bounds extends to settings where  $h$  is discrete, too). Finally, Assumption 1(e) is for ease of presentation; similar results can be derived under a less restrictive assumption, resulting in more elaborate expressions for the error bounds in terms of *conditional* density functions of  $h$ , given  $q$  and  $T$ .

### 2.2 Convex approximation

In this subsection we define our convex approximation of the risk-averse MIR model from Sect. 2.1. In particular, we construct a so-called  $\alpha$ -approximation  $\tilde{Q}_\alpha$  of the

recourse function  $Q$ . This approximation is based on a corresponding  $\alpha$ -approximation  $\tilde{v}_\alpha$  of the value function  $v$  from van der Laan and Romeijnders [17].

As a starting point, let  $q \in \Xi^q$  be fixed and consider the dual representation of the LP-relaxation  $v_{LP}$  of the value function  $v$ :

$$v_{LP}(\xi, x) = \max_{\lambda \in \mathbb{R}^m} \{ \lambda^T (h - Tx) \mid \lambda^T W \leq q \} = \max_{k \in K^q} (\lambda^k)^T (h - Tx),$$

where  $\lambda^k := q_{B^k} (B^k)^{-1}$ , is the vertex of the dual feasible region corresponding to the dual feasible basis matrix  $B^k$ ,  $k \in K^q$ . Identifying the optimal vertex for each value of  $h - Tx$  yields the following result from the literature.

**Lemma 1** *Consider the LP-relaxation  $v_{LP}$  of the value function  $v$  and let  $q \in \Xi^q$  be fixed. Then, there exist dual feasible basis matrices  $B^k$ , with associated cones  $\Lambda^k := \{s \in \mathbb{R}^m \mid (B^k)^{-1}s \geq 0\}$ ,  $k \in K^q$ , with  $\cup_{k \in K^q} \Lambda^k = \mathbb{R}^m$ , and whose interiors are pairwise disjoint, such that  $v_{LP}(\xi, x) = (\lambda^k)^T (h - Tx)$  if  $h - Tx \in \Lambda^k$ .*

**Proof** See Theorem 2.9 in [14]. □

Hence, for every fixed  $q \in \Xi^q$ , we can partition  $\mathbb{R}^m$  into the cones  $\Lambda^k$ ,  $k \in K^q$ , each associated with a single dual optimal basis matrix  $B^k$  and vertex  $\lambda^k$ . Romeijnders et al. [14] prove that a similar partition also exists for mixed-integer linear programs. In particular, they show that there exists  $\sigma^k \in \Lambda^k$  such that for all  $h - Tx \in \sigma^k + \Lambda^k$ ,

$$v(\xi, x) = v_{LP}(\xi, x) + \psi^k(\xi, x)$$

where  $\psi^k$ ,  $k \in K^q$ , represents the ‘‘penalty’’ incurred as a result of the integer restrictions on  $y$ . Specifically,  $\psi^k$  is defined as

$$\begin{aligned} \psi^k(\xi, x) = \min_{\substack{y_{B^k} \\ y_{N^k}}} \{ & q_{B^k}^T y_{B^k} + q_{N^k}^T y_{N^k} : B^k y_{B^k} + N^k y_{N^k} = h - Tx, (y_{B^k}, y_{N^k}) \in \hat{Y} \} \\ & - q_{B^k}^T (B^k)^{-1} (h - Tx), \end{aligned}$$

where  $\hat{Y}$  imposes the same integrality and non-negativity restrictions on the variables but with the non-negativity constraints of the basic variables  $y_{B^k}$  relaxed. Romeijnders et al. [14] prove that for  $q$  fixed,  $\psi^k$  is a ‘‘ $B^k$ -periodic’’ function (see Definition 2.2 in [14]).

**Lemma 2** *Consider the setting of Lemma 1 and let  $q \in \Xi^q$  be fixed. Then, there exist vectors  $\sigma^k \in \Lambda^k$ ,  $k \in K^q$ , such that*

$$v(\xi, x) = (\lambda^k)^T (h - Tx) + \psi^k(h - Tx), \quad \text{if } h - Tx \in \sigma^k + \Lambda^k,$$

where  $\psi^k$  is a  $B^k$ -periodic function.

**Proof** See Theorem 2.9 in [14]. □

From this lemma it is clear that for  $h - Tx \in \sigma^k + \Lambda^k$ , the non-convexity of  $v(\xi, x)$  as a function of  $x$  is a result of the periodicity of  $\psi^k$ . To eliminate this non-convexity, van der Laan and Romeijnders [17] propose to replace  $\psi^k(h - Tx)$  by  $\psi^k(h - \alpha)$ , for some constant vector  $\alpha \in \mathbb{R}^m$ . Taking the maximum over  $k \in K^q$ , and performing this procedure for every  $q \in \Xi^q$  then yields the  $\alpha$ -approximation of  $v$ .

**Definition 2** Consider the value function  $v$  and let  $\lambda^k$ , and  $\psi^k$ ,  $k \in K^q$ , denote the dual vertices and  $B^k$ -periodic functions from Lemma 2. Then, for any  $\alpha \in \mathbb{R}^m$ , we define the  $\alpha$ -approximation of  $v$  as

$$\tilde{v}_\alpha(\xi, x) := \max_{k \in K^q} \{(\lambda^k)^T (h - Tx) + \psi^k(h - \alpha)\}, \quad \xi \in \Xi, x \in \mathbb{R}^{n_1}.$$

It is not hard to see that  $\tilde{v}_\alpha(\xi, x)$  is indeed convex in  $x$ , as it is the pointwise maximum of a finite number of affine functions. Applying the risk measure  $\rho$  to  $\tilde{v}_\alpha(\xi, x)$  yields the  $\alpha$ -approximation of the recourse function  $Q$ .

**Definition 3** Consider the recourse function  $Q$  from (1). Then, its  $\alpha$ -approximation is defined by

$$\tilde{Q}_\alpha(x) := \rho(\tilde{v}_\alpha(\xi, x)), \quad x \in \mathbb{R}^{n_1}.$$

Since  $\rho$  is a coherent risk measure, it is convexity-preserving (in particular, this is a consequence of the axioms monotonicity and convexity; see [25]). Hence, the approximating recourse function  $\tilde{Q}_\alpha$  is indeed a convex function.

### 3 Asymptotic error bound

In this section we derive a performance guarantee for our  $\alpha$ -approximation model. We do this by deriving an asymptotic error bound for the  $\alpha$ -approximation  $\tilde{Q}_\alpha$  from Definition 3. That is, we derive an upper bound on the maximum approximation error

$$\|Q - \tilde{Q}_\alpha\|_\infty := \sup_{x \in \mathbb{R}^{n_1}} |Q(x) - \tilde{Q}_\alpha(x)|.$$

Similar as in the literature on risk-neutral models, we aim for an asymptotic error bound, in the sense that the bound converges to zero as the *total variations* of the conditional density functions  $f_i(\cdot|h_{-i})$ ,  $i = 1, \dots, m$ , of the right-hand-side vector  $h$  go to zero.

**Definition 4** Let  $f : D \rightarrow \mathbb{R}$  be a real-valued function on an interval  $D \subseteq \mathbb{R}$  and let  $I \subseteq D$  be a given subinterval. Let  $\Pi(I)$  denote the set of all finite ordered sets  $P = \{z_1, \dots, z_{N+1}\}$  with  $z_1 < \dots < z_{N+1}$  in  $I$ . Then, the *total variation*, *total increase*, and *total decrease* of  $f$  on  $I$ , denoted by  $|\Delta|f(I)$ ,  $\Delta^+ f(I)$ , and  $\Delta^- f(I)$ , respectively, are defined by

$$|\Delta|f(I) := \sup_{P \in \Pi(I)} V_f(P), \quad \Delta^+ f(I) := \sup_{P \in \Pi(I)} V_f^+(P), \quad \Delta^- f(I) := \sup_{P \in \Pi(I)} V_f^-(P),$$

where  $V_f(P) := \sum_{i=1}^N |f(z_{i+1}) - f(z_i)|$ ,  $V_f^+(P) := \sum_{i=1}^N (f(z_{i+1}) - f(z_i))^+$ , and  $V_f^-(P) := \sum_{i=1}^N (f(z_{i+1}) - f(z_i))^-$ . We use short-hand notation  $|\Delta|f := |\Delta|f(D)$  and  $|\Delta|f(\cup_{i=1, \dots, n} I_i) := \sum_{i=1}^n |\Delta|f(I_i)$ , if  $I_1, \dots, I_n$  are disjoint intervals in  $D$ , with straightforward analogues for  $\Delta^+$  and  $\Delta^-$ . Finally, we say that  $f$  is of *bounded variation* if  $|\Delta|f < +\infty$ .

Observe that for one-dimensional unimodal density functions  $f$ , the total variation  $|\Delta|f$  is equal to two times the value at the mode. For example, for a normal distribution with arbitrary mean  $\mu$  and standard deviation  $\sigma$ , we have  $|\Delta|f = \sigma^{-1} \sqrt{2/\pi}$ . We restrict ourselves to joint density functions  $f$  for which the corresponding conditional density functions are of bounded variation.

**Definition 5** We denote by  $\mathcal{H}^m$  the set of all  $m$ -dimensional joint pdfs  $f$  whose conditional density functions  $f_i(\cdot|t_{-i})$  are of bounded variation for all  $t_{-i} \in \mathbb{R}^{m-1}$ ,  $i = 1, \dots, m$ .

To derive an asymptotic error bound, we take the error bound from [22] for the special case  $\rho = \text{CVaR}_\beta$  as a starting point. We generalize this bound to our risk-averse MIR setting, where  $\rho$  can be any law invariant, coherent risk measure. For this purpose, we use the so-called *Kusuoka representation* [28] of law invariant coherent risk measures.

**Lemma 3** (Kusuoka representation) *Let  $\rho$  be a law invariant coherent risk measure. Then, there exists a set  $\mathcal{M}$  of probability measures on  $[0, 1]$  such that*

$$\rho(Z) = \sup_{\mu \in \mathcal{M}} \int_0^1 \text{CVaR}_\beta(Z) d\mu(\beta), \quad Z \in \mathcal{Z},$$

where  $\text{CVaR}_\beta$  represents the conditional value-at-risk with parameter  $\beta \in [0, 1)$ , defined as

$$\text{CVaR}_\beta(Z) = \min_{\zeta \in \mathbb{R}} \left\{ \zeta + \frac{1}{1 - \beta} \mathbb{E}[(Z - \zeta)^+] \right\}, \quad Z \in \mathcal{Z}.$$

If, in addition,  $\rho$  is comonotone additive (i.e., if  $\rho$  is a spectral risk measure), then  $\mathcal{M}$  reduces to a singleton.

**Proof** See, e.g., Shapiro [23]. (The original result by Kusuoka [28] holds in a more restrictive setting.) □

Lemma 3 shows that every law invariant coherent risk measure can be represented as a supremum over convex combinations of CVaRs. Combining this fact with the error bound for CVaR from [22], we can derive our error bound.

**Theorem 1** *Consider the recourse  $Q$  from (1) and its  $\alpha$ -approximation  $\tilde{Q}_\alpha$  from Definition 3 in a setting with an underlying law invariant, coherent risk measure  $\rho$ . Let*



$\mathcal{M}$  be the set of probability measures on  $[0, 1]$  from the Kusuoka representation of  $\rho$ . For every  $\mu \in \mathcal{M}$ , define the function  $\phi_\mu : [0, 1] \rightarrow \mathbb{R}$  as

$$\phi_\mu(p) := \int_{[0,p]} (1 - \beta)^{-1} d\mu(\beta), \quad p \in [0, 1]. \tag{3}$$

Then, there exists a constant  $C > 0$  such that for every  $f \in \mathcal{H}^m$ , we have

$$\|Q - \tilde{Q}_\alpha\|_\infty \leq \sup_{\mu \in \mathcal{M}} \phi_\mu(1) \cdot C \cdot \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta| f_i(\cdot | h_{-i})].$$

**Proof** Let  $x \in \mathbb{R}^{n_1}$  be given. Then, applying Lemma 3 to both  $Q(x)$  and  $\tilde{Q}_\alpha(x)$ , we have

$$\begin{aligned} |Q(x) - \tilde{Q}_\alpha(x)| &= \left| \sup_{\mu \in \mathcal{M}} \left\{ \int_0^1 \text{CVaR}_\beta(v(\xi, x)) d\mu(\beta) \right\} \right. \\ &\quad \left. - \sup_{\mu \in \mathcal{M}} \left\{ \int_0^1 \text{CVaR}_\beta(\tilde{v}_\alpha(\xi, x)) d\mu(\beta) \right\} \right| \\ &\leq \sup_{\mu \in \mathcal{M}} \left\{ \int_0^1 |\text{CVaR}_\beta(v(\xi, x)) - \text{CVaR}_\beta(\tilde{v}_\alpha(\xi, x))| d\mu(\beta) \right\}. \tag{4} \end{aligned}$$

By Theorem 1 in [22], there exists a constant  $C > 0$ , not depending on  $x$ , such that

$$|\text{CVaR}_\beta(v(\xi, x)) - \text{CVaR}_\beta(\tilde{v}_\alpha(\xi, x))| \leq (1 - \beta)^{-1} C \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta| f_i(\cdot | h_{-i})].$$

Here,  $C$  is finite by Theorem 2.1 in [29]. Substituting this inequality into the integral in (4), we obtain

$$\begin{aligned} |Q(x) - \tilde{Q}_\alpha(x)| &\leq \sup_{\mu \in \mathcal{M}} \left\{ \int_0^1 \left( (1 - \beta)^{-1} C \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta| f_i(\cdot | h_{-i})] \right) d\mu(\beta) \right\} \\ &= \sup_{\mu \in \mathcal{M}} \left\{ \int_0^1 (1 - \beta)^{-1} d\mu(\beta) \right\} \cdot C \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta| f_i(\cdot | h_{-i})] \\ &= \sup_{\mu \in \mathcal{M}} \phi_\mu(1) \cdot C \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta| f_i(\cdot | h_{-i})], \end{aligned}$$

where the last step follows from the definition of the risk spectrum  $\phi_\mu$  in (3). The result now follows from the observation that the upper bound above does not depend on the value of  $x$ . □

Similar to the error bounds from the literature, the error bound from Theorem 1 depends on the total variations of the one-dimensional conditional density functions

of the right-hand side vector  $h$ .<sup>2</sup> As these total variations go to zero, the error bound converges to zero. Intuitively, this can be interpreted as follows. Note that a density function with a small total variation is relatively “flat”. As the total variation decreases further, the distribution becomes ever flatter. This means that the distribution of  $h$  becomes more highly dispersed. So on an intuitive level, the error bound above shows that, if the distribution of  $h$  is unimodal, then our convex approximations are good if the distribution of  $h$  is relatively highly dispersed. A final interesting interpretation of Theorem 1 is that  $Q$  is “almost convex” (i.e., close to a convex function) if the distribution of  $h$  is highly dispersed (and unimodal).

Comparing the error bound from Theorem 1 with the error bound for the risk-neutral case in [22], we see that we have an additional factor  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1)$ . This expression depends on the specific risk measure  $\rho$ . To interpret this expression, consider the spectral case, where  $\mathcal{M}$  is a singleton  $\{\mu\}$ . Then, the function  $\phi_\mu$  from Theorem 1 is called the *risk spectrum* [26] associated with  $\rho$  and  $\rho(Z)$ ,  $Z \in \mathcal{Z}$ , can be represented as an expected value under a *distorted* distribution, where the risk spectrum  $\phi_\mu(p)$  represents the distortion (i.e., the factor by which probabilities are multiplied) at the  $p$ -quantile of the distribution of  $Z$ . Since  $\phi_\mu$  is non-decreasing,  $\phi_\mu(1)$  represents the maximum distortion of probabilities associated with the risk measure  $\rho_\mu$ . It follows that for general law invariant coherent risk measures  $\rho$ , the value  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1)$  is the maximum possible distortion associated with the risk measure  $\rho$ .

Since for every  $\mu \in \mathcal{M}$ ,  $\phi_\mu$  is a non-negative non-decreasing function on  $[0, 1]$  that integrates to one, it follows that  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1) \geq 1$ . Hence, our error bound is at least as large as its risk-neutral analogue, with the minimum value attained for the expected value:  $\rho = \mathbb{E}$ . As long as  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1)$  is finite, our error bound has the asymptotic interpretation described above: the approximation error goes to zero as the distribution of  $h$  becomes more dispersed. For instance, this will hold if  $\mathcal{M}$  consists of a single element  $\mu$ , (i.e., if  $\rho$  is a spectral risk measure), and  $\phi_\mu(1)$  is finite. Note that  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1)$  can be finite even if  $\mathcal{M}$  contains uncountably many elements, as we illustrate in the following example.

**Example 1** Consider the law invariant risk measure *mean-upper semideviation* (mean-USD) with parameter  $c \geq 0$ , defined as

$$\rho(Z) := \mathbb{E}[Z] + c \cdot \mathbb{E}\left[(Z - \mathbb{E}[Z])^+\right], \quad Z \in \mathcal{Z}.$$

By Example 2.46 in [30], this is a coherent risk measure for  $0 \leq c \leq 1$ , and the set  $\mathcal{M}$  of probability measures on  $[0, 1)$  from its Kusuoka representation is given by  $\mathcal{M} = \{\mu \in \mathcal{P}([0, 1)) \mid \phi_\mu(1) \leq 1 + c\}$ . It follows that  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1) = 1 + c$ .

Finally, we provide an example showing that the error bound from Theorem 1 can be tight.

<sup>2</sup> Note that, as in [22], the error bound does not depend on the distribution of  $T$ , as the error bound is uniform over all  $x$ , and  $T$  and  $x$  always appear together as  $Tx$ . Moreover, the dependence of the error bound on  $q$  is captured in the constant  $C$ ; see van Beesten and Romeijnders [21] for a recent publication investigating the dependence of  $C$  on  $q$  in detail.

**Example 2** Consider a so-called *simple integer recourse* setting. That is, consider the model from (1), with the assumptions that  $n_1 = m = 1$ ,  $q = (1, 0)$  and  $T = [1]$  are fixed,  $Y = \mathbb{Z}_+ \times \mathbb{R}_+$ , and  $W = [1 \ -1]$ . In other words, we assume that the value function can be written as

$$v(h, x) = \min_{y_1 \in \mathbb{Z}_+} \{y_1 \mid y_1 \geq h - x\} = [h - x]^+, \quad h \in \mathbb{R}^m, x \in \mathbb{R},$$

where  $[s]^+ := \max\{[s], 0\}$ ,  $s \in \mathbb{R}$ . Moreover, assume that  $\rho = \text{CVaR}_\beta$  with  $\beta = 3/4$ , and let the density function  $f$  of  $h$  be given by  $f(t) = 1/4$  if  $0 \leq t < 3$ ,  $f(t) = 1/2$  if  $3 \leq t < 3.5$ , and  $f(t) = 0$  otherwise. We show that the error bound from Theorem 1 is tight.

Leveraging a result in Romeijnders et al. [16], the relevant constant is  $C = 1/8$ . Moreover, it is not hard to show that  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1) = (1 - \beta)^{-1} = 4$  and that  $|\Delta|f = 1$ . Thus, the error bound reduces to  $\|Q - \tilde{Q}_\alpha\|_\infty \leq 4 \cdot \frac{1}{8} \cdot 1 = 1/2$ . Next, we show that there exist  $x \in \mathbb{R}$  and  $\alpha \in [0, 1)$  such that  $|Q(x) - \tilde{Q}_\alpha(x)|$  attains the error bound. Since  $v$  and  $\tilde{v}_\alpha$  are monotonely non-decreasing in  $h$ , we have that  $Q(x) = \int_{-\infty}^\infty v(t, x)g(t)dt$  and  $\tilde{Q}_\alpha(x) = \int_{-\infty}^\infty \tilde{v}_\alpha(t, x)g(t)dt$ , where the pdf  $g$  represents the distribution of  $h$ , restricted to the fraction  $1 - \beta = 1/4$  of its highest outcomes. That is,  $g(t) = 2$  if  $3 \leq t < 3.5$ , and  $g(t) = 0$  otherwise. For  $x = 0$  and  $\alpha = 0.5$  we have  $|Q(x) - \tilde{Q}_\alpha(x)| = \left| \int_{-\infty}^\infty (v(t, x) - \tilde{v}_\alpha(t, x))g(t) \right| = \left| \int_3^{3.5} (4 - 3.5) \cdot 2 \right| = 1/2$ . This proves that the error bound is tight.

We conclude this section with an analogue of Theorem 1 for the law invariant, non-coherent risk measure *expected excess (EE)*. This risk measure represents the expected amount by which a random variable exceeds a predefined threshold  $\tau$ . Even though EE is not coherent, it has relevant practical applications in settings where we are interested in the expected amount by which a budget  $\tau$  is exceeded. The result below illustrates that our approach is not necessarily limited to coherent risk measures. Moreover, we will use it in our proof of Theorem 3 in Sect. 5.

**Definition 6** Let  $Z \in \mathcal{Z}$  be a random variable. Then, its *expected excess (EE)* with parameter  $\tau \in \mathbb{R}$  is given by

$$EE_\tau(Z) := \mathbb{E}[(Z - \tau)^+].$$

**Proposition 1** Consider the recourse function  $Q$  from (1) and its  $\alpha$ -approximation  $\tilde{Q}_\alpha$  from Definition 3. Assume that the underlying risk measure satisfies  $\rho = EE_\tau$  for some  $\tau \in \mathbb{R}$ . Then, there exists a constant  $C > 0$  such that for every  $f \in \mathcal{H}^m$ , we have

$$\|Q - \tilde{Q}_\alpha\|_\infty \leq C \cdot \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta|f_i(\cdot|h_{-i})].$$

**Proof** Let  $x \in \mathbb{R}^{n_1}$  be given. Then,

$$|Q(x) - \tilde{Q}_\alpha(x)| = \left| \mathbb{E}[(v(\xi, x) - \tau)^+] - \mathbb{E}[(\tilde{v}_\alpha(\xi, x) - \tau)^+] \right|,$$

which equals  $|\mathbb{E}[v^\tau(\xi, x)] - \mathbb{E}[\tilde{v}_\alpha^\tau(\xi, x)]| = |R^*(x, \tau) - \tilde{Q}_\alpha^*(x, \tau)|$  in the notation of [22]. By the second part of the proof of Theorem 1 in that paper, it follows that there exists a constant  $C > 0$ , not depending on  $\tau$ , such that for every  $f \in \mathcal{H}^m$ ,

$$|Q(x) - \tilde{Q}_\alpha(x)| \leq C \cdot \sum_{i=1}^m \mathbb{E}_{h_{-i}}[|\Delta| f_i(\cdot | h_{-i})].$$

The result follows from the observation that the right-hand side here does not depend on the value of  $x$  or  $\tau$ . □

### 4 Non-asymptotic error bounds for special cases

Theorem 1 provides a worst-case error bound that holds for the general risk-averse MIR model from (1). A downside of the generality of this result is that it may be very conservative in specific settings: in particular, the constant  $C$  in Theorem 1 might be larger than necessary. In Sect. 4.1 we tackle this issue for the special case of totally unimodular integer recourse by deriving a specialized, tighter error bound. This bound also turns out to behave more favorably as a function of  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1)$  than the bound from Theorem 1. Inspired by this observation, in Sect. 4.2 we provide a simple integer recourse example for which we derive an error bound that does not depend on the choice of risk measure  $\rho$  at all. Both results show that our convex approximation approach might perform better in practice than the asymptotic error bound from Theorem 1 suggests.

#### 4.1 Totally unimodular integer recourse

We consider a risk-averse totally unimodular (TU) integer recourse model. That is, we consider the model (1) and we assume that the second-stage problem can be written as

$$v(\xi, x) = \min_{y \in Y} \{q^T y \mid Wy + Tx \geq h\}, \tag{5}$$

where  $W$  is a totally unimodular matrix and where all second-stage variables are restricted to integers, i.e.,  $Y = \mathbb{Z}_+^{n_2}$ . We derive a non-asymptotic error bound for this setting that behaves more favorably as a function of  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1)$  than the error bound from Theorem 1.

Our analysis consists of three steps. First, we use the Kusuoka representation of  $\rho$  and the distortion interpretation of spectral risk measures to write  $\rho$  as a supremum over expected values under distorted probability distributions. Second, we apply an error bound from [22] for risk-neutral TU integer recourse models on this expected value expression. The resulting error bound is written in terms of the distorted probability distribution. Finally, we rewrite the result in terms of the original distribution. For this purpose, we use the following steps. First, we observe that in this TU integer setting, the value function  $v(\xi, x)$  is non-decreasing as a function of  $h_i$ . Second, we show

that as a result, the distortion factor in the pdf of  $h_i$  must be non-decreasing as well. Finally, we use the lemma below to find an upper bound on the total variation of the distorted pdf in terms of the total variation of the original pdf  $f$ .

**Lemma 4** *Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a pdf of bounded variation and let  $\zeta : \mathbb{R} \rightarrow \mathbb{R}$  be a non-negative non-decreasing bounded function satisfying  $\zeta(t) \leq \zeta^*$ , for all  $t \in \mathbb{R}$ . Define  $g : \mathbb{R} \rightarrow \mathbb{R}$  by  $g(t) = \lambda(t)f(t)$ ,  $t \in \mathbb{R}$ . Then,*

$$|\Delta|g \leq \zeta^*|\Delta|f.$$

**Proof** Follows directly from the proof of Proposition 3 in [16]. □

**Theorem 2** *Consider the risk-averse TU integer recourse function  $Q$  and its convex approximation  $\tilde{Q}_\alpha$ . Assume that the underlying risk measure  $\rho$  is law invariant and coherent and let  $\mathcal{M}$  be the set of probability measures on  $[0, 1)$  from its Kusuoka representation. Then, for every  $f \in \mathcal{H}^m$ , we have*

$$\|Q - \tilde{Q}_\alpha\|_\infty \leq \sum_{i=1}^m \bar{\lambda}_i^* H \left( \sup_{\mu \in \mathcal{M}} \{ \phi_\mu(1) \} \mathbb{E} [ |\Delta| f_i(\cdot | h_{-i}) ] \right),$$

where for every  $i = 1, \dots, m$ , we have  $\bar{\lambda}_i^* := \mathbb{E}^q [ \max_{k \in K^q} \{ \lambda_i^k \} ]$ , where  $\lambda^k$ ,  $k \in K^q$ ,  $q \in \mathfrak{E}^q$ , are the dual vertices from Lemma 1; and where the function  $H : \mathbb{R}_+ \rightarrow \mathbb{R}$  is defined by

$$H(t) = \begin{cases} t/8, & \text{if } 0 \leq t \leq 4, \\ 1 - t/2, & \text{if } t > 4. \end{cases} \tag{6}$$

**Proof** Let  $x \in \mathbb{R}^{n_1}$  be given. By the Kusuoka representation of  $\rho$  we have  $Q(x) = \sup_{\mu \in \mathcal{M}} \rho_\mu(v(\xi, x))$ , where  $\rho_\mu$  denotes the spectral risk measure corresponding to  $\mu \in \mathcal{M}$ . Let  $\mu^x \in \mathcal{M}$  be optimal in the supremum above (if the maximum is not attained, we can instead consider a sequence  $\{\mu_n\}_{n=1}^\infty$  converging to the optimum). We know that  $\rho_{\mu^x}(v(\xi, x))$  can be represented as an expected value under a distorted probability distribution  $\mathbb{P}^x$ , where  $\phi_{\mu^x}(p)$  reflects the distortion of the probabilities at the  $p$ -quantile of the distribution of  $v(\xi, x)$ . Recall that  $\phi_\mu$  is non-decreasing. Let  $f^x$  denote the joint pdf of  $h$  under  $\mathbb{P}^x$ . Let  $i = 1, \dots, m$ , be given and let  $f_i^x(\cdot; t_{-i})$  denote  $f^x(t)$  as a function of  $t_i$ . Due to the TU structure of the recourse matrix  $W$ , we know that  $v(\xi, x)$  is non-decreasing as a function of  $h_i$ . Hence, we can assume without loss of generality that there exists a non-negative, non-decreasing function  $\zeta_i^x : \mathbb{R} \rightarrow \mathbb{R}$ , satisfying  $\zeta_i^x(t_i) \leq \phi_{\mu^x}(1)$  for all  $t_i \in \mathbb{R}$ , such that  $f_i^x(t_i; t_{-i}) = \zeta_i^x(t_i) f_i(t_i; t_{-i})$ ,  $t_i \in \mathbb{R}$ .

Now, we can write  $Q(x) = \rho_{\mu^x}(v(\xi, x)) = \mathbb{E}^{\mathbb{P}^x} [ v(\xi, x) ]$  and by feasibility of  $\mu_x$  and  $\mathbb{P}^x$ ,  $\tilde{Q}_\alpha(x) \leq \rho_{\mu^x}(\tilde{v}_\alpha(\xi, x)) \leq \mathbb{E}^{\mathbb{P}^x} [ \tilde{v}_\alpha(\xi, x) ]$ . It follows that

$$Q(x) - \tilde{Q}_\alpha(x) \leq \mathbb{E}^{\mathbb{P}^x} [ v(\xi, x) ] - \mathbb{E}^{\mathbb{P}^x} [ \tilde{v}_\alpha(\xi, x) ]$$

$$\leq \sum_{i=1}^m \bar{\lambda}_i^* H \left( \mathbb{E}^{\mathbb{P}^x} [|\Delta| f_i^x(\cdot|h_{-i})] \right), \tag{7}$$

where the second inequality follows from applying the error bound for risk-neutral TU integer recourse models from Proposition 4 in [22].

It remains to write the expected value in the error bound above in terms of the original distribution  $\mathbb{P}$  and pdf  $f$ . Define  $\mathcal{T}_{-i} := \{t_{-i} \in \mathbb{R}^{m-1} \mid f_{-i}(t_{-i}) > 0\}$  as the set on which  $f_{-i}$  is positive. Similarly, define  $\mathcal{T}_{-i}^x := \{t_{-i} \in \mathbb{R}^{m-1} \mid f_{-i}^x(t_{-i}) > 0\}$  and observe that  $\mathcal{T}_{-i}^x \subseteq \mathcal{T}_{-i}$ . Using the definition of the conditional density function  $f_i(\cdot|h_{-i})$  we have  $\mathbb{E}^{\mathbb{P}} [|\Delta| f_i(\cdot|h_{-i})] = \int_{\mathcal{T}_{-i}} |\Delta| \left( \frac{f_i(\cdot; t_{-i})}{f_{-i}(t_{-i})} \right) f_{-i}(t_{-i}) dt_{-i} = \int_{\mathcal{T}_{-i}} |\Delta| f_i(\cdot; t_{-i}) dt_{-i}$ . Similarly, we find  $\mathbb{E}^{\mathbb{P}^x} [|\Delta| f_i^x(\cdot|h_{-i})] = \int_{\mathcal{T}_{-i}^x} |\Delta| f_i^x(\cdot; t_{-i}) dt_{-i}$ . Using these equalities and the fact that  $|\Delta| f_i^x(\cdot; t_{-i}) \leq \phi_{\mu^x}(1) |\Delta| f_i(\cdot; t_{-i})$  by Lemma 4, we obtain

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^x} [|\Delta| f_i^x(\cdot|h_{-i})] &= \int_{\mathcal{T}_{-i}^x} |\Delta| f_i^x(\cdot; t_{-i}) dt_{-i} \leq \int_{\mathcal{T}_{-i}} |\Delta| f_i^x(\cdot; t_{-i}) dt_{-i} \\ &\leq \int_{\mathcal{T}_{-i}} \phi_{\mu^x}(1) |\Delta| f_i(\cdot; t_{-i}) dt_{-i} \leq \sup_{\mu \in \mathcal{M}} \phi_{\mu}(1) \mathbb{E}^{\mathbb{P}} [|\Delta| f_i(\cdot|h_{-i})]. \end{aligned}$$

Substituting this inequality into (7) and using an analogous analysis for an upper bound on the reverse difference  $\tilde{Q}_{\alpha}(x) - Q(x)$  yields the result.  $\square$

Comparing the TU integer error bound from Theorem 2 with the general error bound from Theorem 1, we see two differences. First, instead of merely proving the existence of a constant  $C$ , we derive the constants  $\bar{\lambda}^*, i = 1, \dots, m$ . Hence, the error bound from Theorem 2 is not merely asymptotic in nature, but can be used to derive an actual performance guarantee in finite cases. Second, the error bound from Theorem 2 depends on the risk measure  $\rho$  in a more generous way, as the factor  $\sup_{\mu \in \mathcal{M}} \phi_{\mu}(1)$  enters inside the function  $H$ , which is bounded from above by one. Hence, even if  $\sup_{\mu \in \mathcal{M}} \phi_{\mu}(1)$  is very large, the error bound remains relatively small. This provides evidence for our conjecture that the dependence of the error bound from Theorem 1 on the risk measure  $\rho$  might be (much) too conservative in some cases.

### 4.2 Simple integer recourse with an exponential distribution

We now consider the special case of risk-averse one-dimensional simple integer recourse (SIR) models [31] with an exponentially distributed right-hand side variable  $h$ . We use this very special case as an example for which we can derive an error bound that does not depend on the risk measure  $\rho$  at all. This provides further evidence that in some settings the approximation error  $\|Q - \tilde{Q}_{\alpha}\|_{\infty}$  may depend on the risk measure  $\rho$  more favorably than the error bound from Theorem 1 suggests.

Specifically, we consider the model from (1), with the assumptions that  $n_1 = m = 1$ ;  $q = (1, 0)$  and  $T = [1]$  are fixed;  $Y = \mathbb{Z}_+ \times \mathbb{R}_+$ ; and  $W = [1 \ -1]$ . In other words,

we assume that the value function can be written as

$$v(h, x) = \min_{y_1 \in \mathbb{Z}_+} \{y_1 \mid y_1 \geq h - x\} = \lceil h - x \rceil^+, \quad h \in \mathbb{R}^m, x \in \mathbb{R},$$

where  $\lceil s \rceil^+ := \max\{\lceil s \rceil, 0\}$ ,  $s \in \mathbb{R}$ . Moreover, we assume that  $h$  follows an exponential distribution with a rate parameter of  $\gamma$ . We show that we can derive a *uniform* error bound that holds for *all* spectral risk measures  $\rho$  that satisfy some regularity conditions.

Consider the recourse function  $Q(x) = \rho(v(\xi, x))$ . Let  $\phi_\mu$  be the risk spectrum associated with  $\rho$  by its Kusuoka representation. Recall that  $\phi_\mu(\beta)$  represents the distortion factor of the  $\beta$ -quantile of the distribution of  $v(h, x)$ . We make the important observation that the value function  $v(h, x) = \lceil h - x \rceil^+$  is non-decreasing in  $h$ . Hence, the  $\beta$ -quantile of the distribution of  $h$  corresponds to the  $\beta$ -quantile of the distribution of  $v(h, x)$ . As a result, we can write  $Q(x)$  as an expected value under a distorted distribution that can be expressed explicitly. We have  $Q(x) = \mathbb{E}^{f_\mu}[v(h, x)]$ ,  $x \in \mathbb{R}$ , where  $f_\mu$  denotes the distorted pdf defined by

$$f_\mu(t) = \phi_\mu(F(t))f(t), \quad t \in \mathbb{R}, \tag{8}$$

where  $F$  and  $f$  are the cdf and pdf, respectively, of  $h$ . Note that  $f_\mu$  does not depend on the value of  $x$ . Similarly, for the  $\alpha$ -approximation we obtain  $\tilde{Q}_\alpha(x) = \mathbb{E}^{f_\mu}[\tilde{Q}_\alpha(\xi, x)]$ .

Using these representations, we can write  $Q$  and  $\tilde{Q}_\alpha$  as a recourse function and its  $\alpha$ -approximation, respectively, of a *risk-neutral* SIR model *under the distorted distribution* represented by  $f_\mu$ . Hence, we can apply an error bound from the literature on convex approximations of risk-neutral MIR models (Theorem 5 in [16]). This yields

$$\|Q - \tilde{Q}_\alpha\|_\infty \leq H(|\Delta|f_\mu), \tag{9}$$

where  $H$  is the function from (6). Interestingly, under the assumption that  $h$  is exponentially distributed, we can derive a uniform upper bound on the total variation  $|\Delta|f_\mu$ , and hence, on  $\|Q - \tilde{Q}_\alpha\|_\infty$  that holds for *all* spectral risk measures that satisfy some regularity conditions.

**Proposition 2** *Consider the recourse function  $Q$  and its  $\alpha$ -approximation in the SIR setting described in Sect. 4.2, where  $h$  follows an exponential distribution with rate parameter  $\lambda$ . Suppose that  $\rho$  is a spectral risk measure with associated probability measure  $\mu$  on  $[0, 1)$  and risk spectrum  $\phi_\mu$ , and that it satisfies the following regularity conditions:*

- (i)  $\phi_\mu(1) < +\infty$ ;
- (ii) *The set  $M$  of points on which  $\mu$  has positive probability mass is finite. Moreover,  $\mu$  has a density  $g$  on  $[0, 1) \setminus M$ ;*
- (iii) *The set  $D := \{\beta \in (0, 1) \mid (1 - \beta)\phi_\mu(\beta) \text{ has a negative derivative at } \beta\}$  can be written as the union of a finite number of intervals.*

Then, we have

$$\|Q - \tilde{Q}_\alpha\|_\infty \leq H(2\lambda), \tag{10}$$

where  $H$  is the function from (6).

**Proof** By (9), it suffices to prove that  $|\Delta|f_\mu \leq 2\lambda$ . Consider the function  $\bar{f}_\mu$  as defined in (8). Using the change of variables  $\beta = F(t)$ , we can express  $f_\mu$  as a function of  $\beta$ . We obtain the function  $\bar{f}_\mu : [0, 1] \rightarrow \mathbb{R}$ , defined by

$$\bar{f}_\mu(\beta) := f_\mu(F^{-1}(\beta)) = \lambda(1 - \beta)\phi_\mu(\beta), \quad \beta \in [0, 1],$$

where we use that the inverse cdf  $F^{-1}$  of  $h$  is given by  $F^{-1}(\beta) = -\frac{1}{\lambda} \log(1 - \beta)$ ,  $\beta \in [0, 1]$ . Since  $F$  is a continuous, invertible cdf, and  $F^{-1}(0) = 0$  and  $F^{-1}(1) = +\infty$ , it follows that  $|\Delta|f_\mu = |\Delta|\bar{f}_\mu((-\infty, 0]) + |\Delta|\bar{f}_\mu([0, +\infty)) = \bar{f}_\mu(0) + |\Delta|\bar{f}_\mu = \bar{f}_\mu(0) + \Delta^+ \bar{f}_\mu + \Delta^- \bar{f}_\mu$ , where we use the fact that  $|\Delta|\bar{f}_\mu = \Delta^+ \bar{f}_\mu + \Delta^- \bar{f}_\mu$ . By definition of  $\bar{f}_\mu$  and condition (i), we have  $\bar{f}_\mu(0) \geq 0 = f_\mu(1)$  and hence,  $\Delta^- \bar{f}_\mu = \Delta^+ \bar{f}_\mu + \bar{f}_\mu(0)$ . It follows that  $|\Delta|f_\mu = 2\Delta^- \bar{f}_\mu$ . Hence, it suffices to find an upper bound on the total decrease  $\Delta^- \bar{f}_\mu$  only.

We will show that  $\bar{f}_\mu$  is decreasing only on the set  $D$  defined in condition (iii), so that we can write  $\Delta^- \bar{f}_\mu = \Delta^- \bar{f}_\mu(D)$ . To prove this, first note that  $\bar{f}_\mu$  is continuous for all  $\beta \in (0, 1] \setminus M$ , has an upward discontinuity at all  $\beta \in M$ , and since  $\bar{f}_\mu$  is non-negative and  $\bar{f}_\mu(0) = 0$ , it is non-decreasing at 0, in the sense that  $\bar{f}_\mu(0) \leq \lim_{s \downarrow 0} \bar{f}_\mu$ . It follows that  $\Delta^- \bar{f}_\mu := \Delta^- \bar{f}_\mu([0, 1]) = \Delta^- \bar{f}_\mu((0, 1] \setminus M)$ . Next, we show that the derivative  $\bar{f}'_\mu(\beta)$  exists at all  $\beta \in (0, 1] \setminus M$ . Let  $\beta \in (0, 1] \setminus M$  be given. From the definition of  $\bar{f}_\mu$  it is clear that the derivative  $\bar{f}'_\mu(\beta)$  exists if and only the derivative  $\phi'_\mu(\beta)$  exists, which is the case for all  $\beta \in (0, 1] \setminus M$  by condition (ii). For such  $\beta$ , it is given by  $\phi'_\mu(\beta) = (1 - \beta)^{-1}g(\beta)$ . It follows that  $\bar{f}'_\mu(\beta)$  exists for all  $\beta \in (0, 1] \setminus M$  and is given by

$$\bar{f}'_\mu(\beta) = \lambda(1 - \beta)\phi'_\mu(\beta) - \lambda\phi_\mu(\beta) = \lambda(1 - \beta)g(\beta) - \lambda\phi_\mu(\beta), \quad \beta \in (0, 1] \setminus M. \tag{11}$$

Consider the set  $D$  from condition (iii) and observe that  $D \cap M = \emptyset$  since  $\phi'_\mu(\beta)$  does not exist for  $\beta \in M$ . Hence, the set  $D \subseteq (0, 1] \setminus M$ , consisting of finitely many intervals by condition (iii), represents the subset of  $[0, 1]$  on which  $\bar{f}_\mu$  is decreasing. We conclude that  $\Delta^- \bar{f}_\mu = \Delta^- \bar{f}_\mu(D)$ , where the right-hand side is well-defined by condition (iii).

Now, to obtain our upper bound, we express  $\Delta^- \bar{f}_\mu(D)$  in terms of an integral over the derivative of  $\bar{f}_\mu$ . It is not hard to prove that  $\bar{f}_\mu$  is absolutely continuous on  $D$ . Hence, by, e.g., Theorem 14 in [32] it follows that  $\Delta^- \bar{f}_\mu(D) = \int_D |\bar{f}'_\mu(\beta)|d\beta$ . Using (11), we obtain

$$\Delta^- \bar{f}_\mu(D) = \int_D |\bar{f}'_\mu(\beta)|d\beta = \int_D (\lambda\phi_\mu(\beta) - \lambda(1 - \beta)g(\beta))d\beta$$



$$\leq \lambda \int_D \phi_\mu(\beta) d\beta \leq \lambda \int_0^1 \phi_\mu(\beta) d\beta = \lambda,$$

where the Riemann integrals are well-defined by condition (iii) and where we use the fact that the risk spectrum  $\phi_\mu$  integrates to one by definition. It follows that  $|\Delta|f_\mu = |\Delta|\tilde{f}_\mu = 2\Delta^- \tilde{f}_\mu \leq 2\lambda$ , which proves the result.  $\square$

### 5 Statistical error bound for SAA

The error bounds from Sects. 3 and 4 depend on the joint density function  $f$  of the right-hand side vector  $h$ . Hence, they are only valid under the assumption that  $h$  is continuously distributed. However, solving recourse models under continuous distributions is generally a very hard task, even if the model is convex. The reason for this is that evaluating the objective function involves computing (multi-dimensional) integrals, which can be numerically challenging.

To overcome this difficulty one often resorts to discretizing the distribution of  $h$ . In particular, a common approach is *sample average approximation (SAA)* [24], in which we draw a sample from the distribution of  $\xi$  (and hence, of  $h$ ) and then solve the model under the empirical distribution corresponding to the sample.

**Definition 7** Consider the recourse function  $Q$  from (1) and its  $\alpha$ -approximation  $\tilde{Q}_\alpha$  from Definition 3, and let  $\Xi^N := (\xi^1, \dots, \xi^N)$  be an i.i.d. sample from the distribution of  $\xi$ . Then, we define the *sample average approximation (SAA)* of  $Q$  and  $\tilde{Q}_\alpha$  by

$$Q^N(x) := \rho^N(v(\xi, x)), \quad \text{and} \quad \tilde{Q}_\alpha^N(x) := \rho^N(\tilde{v}_\alpha(\xi, x)), \quad x \in \mathbb{R}^{n_1},$$

respectively, where  $\rho^N(\cdot)$  denotes the value of the risk measure  $\rho(\cdot)$  evaluated under the empirical distribution corresponding to  $\Xi^N$ .

Clearly, the total variation-based error bounds from Sect. 3 and 4 do not apply to the model with the (discrete) SAA distribution. Nevertheless, we show that the error bound based on the original continuous distribution persists *in expectation* under the SAA distribution. Only the standard bias of the empirical risk measure  $\rho^N(\cdot)$  needs to be added to obtain a statistical performance guarantee on the quality of the convex approximation under the SAA distribution.

**Theorem 3** Consider the setting of Theorem 1, with and let  $Q^N$  and  $\tilde{Q}_\alpha^N$  denote the sample average approximations from Definition 7. If  $\rho$  is a spectral risk measure, then there exists a constant  $C > 0$  such that for every  $f \in \mathcal{H}^m$  and  $x \in \mathbb{R}^{n_1}$ ,

$$\left| \mathbb{E} \left[ Q^N(x) - \tilde{Q}_\alpha^N(x) \right] \right| \leq \sup_{\mu \in \mathcal{M}} \phi_\mu(1) \cdot C \cdot \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta|f_i(\cdot|h_{-i})] + B^N(x), \quad (12)$$

where  $B^N(x) := \max \left\{ Q(x) - \mathbb{E}[Q^N(x)], \tilde{Q}_\alpha(x) - \mathbb{E}[\tilde{Q}_\alpha^N(x)] \right\}$ .

**Proof** We know from [33] that  $\text{CVaR}_\beta^N$  is negatively biased, i.e., that  $\mathbb{E}[\text{CVaR}_\beta^N(v(\xi, x))] \leq \text{CVaR}_\beta(v(\xi, x))$ . Combining this with the Kusuoka representation of the spectral risk measure  $\rho$  it immediately follows that  $\rho^N$  is negatively biased, too. Hence,  $\mathbb{E}[\rho^N(\tilde{v}_\alpha(\xi, x))] \leq \rho(\tilde{v}_\alpha(\xi, x))$  and  $\mathbb{E}[\rho^N(v(\xi, x))] \leq \rho(v(\xi, x))$ , and thus  $B^N(x) \geq 0$  for every  $x \in \mathbb{R}^{\mu_1}$ . Using this it follows that for every  $x \in \mathbb{R}^{\mu_1}$ ,

$$\begin{aligned} \mathbb{E}[Q^N(x) - \tilde{Q}_\alpha^N(x)] &= \mathbb{E}[\rho^N(v(\xi, x))] - \mathbb{E}[\rho^N(\tilde{v}_\alpha(\xi, x))] \\ &\leq \rho(v(\xi, x)) - \mathbb{E}[\rho^N(\tilde{v}_\alpha(\xi, x))] \\ &= \rho(v(\xi, x)) - \rho(\tilde{v}_\alpha(\xi, x)) \\ &\quad + \left( \rho(\tilde{v}_\alpha(\xi, x)) - \mathbb{E}[\rho^N(\tilde{v}_\alpha(\xi, x))] \right) \\ &\leq \left| \rho(v(\xi, x)) - \rho(\tilde{v}_\alpha(\xi, x)) \right| + B^N(x), \end{aligned}$$

and analogously,

$$\mathbb{E}[\tilde{Q}_\alpha^N(x) - Q^N(x)] \leq \left| \rho(v(\xi, x)) - \rho(\tilde{v}_\alpha(\xi, x)) \right| + B^N(x).$$

The desired result follows from applying Theorem 1 to  $|\rho(v(\xi, x)) - \rho(\tilde{v}_\alpha(\xi, x))|$ .  $\square$

Theorem 3 provides a theoretical justification for using convex approximations in an SAA framework. This extends our error bound from the more theoretical setting with continuous distributions to the more practical setting of SAA. Compared with Theorem 1, Theorem 3 contains an additional term  $B^N(x)$ , which represents the bias of the empirical risk measure  $\rho^N$ , which typically decreases in the sample size  $N$ . Moreover, the bias  $B^N(x)$  equals zero in the risk-neutral case, i.e., when  $\rho = \mathbb{E}$ .

Interestingly, the proof of Theorem 3 carries over to settings far beyond the setting considered in this paper. In particular, many error bounds from the literature on convex approximations of MIR models can be extended to statistical error bounds for the corresponding SAA model in a similar way. Even though the setting in Theorem 3 is the most general setting considered in the literature, this is still an important observation. The reason is that many of the error bounds from the literature have been tailored to special cases, leading to (much) tighter error bounds than the more general bounds in this paper. For instance, the line of proof of Theorem 3 may be applied to the error bounds from [18] to obtain tight statistical error bounds based on information about higher-order derivatives of the density function  $f$ . In short, Theorem 3 can be seen as an extension of the entire literature on error bounds for convex approximations of MIR models to the setting of SAA.

## 6 Loose Benders decomposition algorithm

In this section we construct an algorithm based on the  $\alpha$ -approximation from Definition 3. Such an algorithm is necessary since solving the  $\alpha$ -approximation model directly is typically computationally intractable. The reason for this is the fact that the  $\alpha$ -approximation  $\tilde{Q}_\alpha$  is defined through an enumeration of all periodic functions  $\psi^k$  corresponding to the dual feasible basis matrices  $B^k$  of the LP-relaxation  $v_{LP}$ , of which there are exponentially many.

We propose an extension of the Loose Benders' decomposition algorithm (LBDA) from [17] to the risk-averse case, which is based on Benders' decomposition in an SAA-framework. The algorithm overcomes the computational issue regarding the exponential number of periodic functions  $\psi^k$  by iteratively computing only a few function values for a single function  $\psi^k$  corresponding to an index  $k$  that is close to optimal at the current value of  $x$ . The potential suboptimality of  $k$  yields so-called "loose cuts", which is reflected in the name of the algorithm. Thus, the algorithm iteratively approximates the  $\alpha$ -approximation, while avoiding computation of *all* functions  $\psi^k$ . Nevertheless, we will derive an error bound that guarantees that the quality of the approximation remains good.

### 6.1 Algorithm construction

The LBDA finds an approximate solution to the original risk-averse MIR model (1) by considering the  $\alpha$ -approximation of the SAA model, defined by

$$\tilde{\eta}_\alpha := \min_{x \in X} \{c^T x + \tilde{Q}_\alpha^S(x)\}. \tag{13}$$

The LBDA uses a Benders' decomposition framework to find a solution to this approximation model. In fact, it finds an *approximate* solution to this model by using so-called *loose optimality cuts* rather than exact cuts. We extend this framework by incorporating risk aversion in the LBDA.

#### 6.1.1 Linearizing the risk-averse objective

The main difficulty in extending the LBDA to the risk-averse case is the non-linearity of the objective function in (13) as a result of the risk measure  $\rho$ . To solve this issue, we develop a method to construct a piecewise linear outer approximation of the objective, using a single sorting procedure. The main building block is the result below, which shows that we can compute the (spectral) risk associated with a discrete uniform random variable by sorting its support.

**Lemma 5** *Let  $\rho$  be a spectral risk measure and let  $Z$  be a discrete uniform random variable on  $\{z^1, \dots, z^S\}$ . Define*

$$w^s := \int_{\frac{s-1}{S}}^{\frac{s}{S}} \phi_\mu(p) dp, \quad s = 1, \dots, S.$$

where  $\mu$  is the probability measure on  $[0, 1)$  from the Kusuoka representation of  $\rho$ . Moreover, let  $\Pi$  be the collection of all permutation functions of the set  $\{1, \dots, S\}$ , i.e.,  $\Pi$  is the collection of all one-to-one functions  $\pi : \{1, \dots, S\} \rightarrow \{1, \dots, S\}$ . Then,

$$\rho(Z) = \max_{\pi \in \Pi} \sum_{s=1}^S w^s z^{\pi(s)},$$

Moreover, any  $\pi^* \in \Pi$  for which  $z_{\pi^*(1)} \leq z_{\pi^*(2)} \leq \dots \leq z_{\pi^*(S)}$  holds (i.e., that sorts the set  $\{z^1, \dots, z^S\}$ ) is optimal.

**Proof** By [23], we can represent  $\rho$  in terms of its associated risk spectrum  $\phi_\mu$ :

$$\rho(Z) = \int_0^1 \phi_\mu(p) F_Z^{-1}(p) dp,$$

where  $F_Z^{-1}(p) := \inf\{z \in \mathbb{R} \mid F_Z(z) \geq p\}$ . Using the fact that  $Z$  is uniformly distributed on  $\{z^1, \dots, z^S\}$ , we have

$$F_Z^{-1}(p) = \begin{cases} -\infty & \text{if } p = 0, \\ z^{\pi^*(s)} & \text{if } \frac{s-1}{S} < p \leq \frac{s}{S}, \quad s = 1, \dots, S, \end{cases}$$

where  $\pi^* \in \Pi$  is such that  $z^{\pi^*(1)} \leq z^{\pi^*(2)} \leq \dots \leq z^{\pi^*(S)}$ , i.e.,  $\pi^*$  sorts the set  $\{z^1, \dots, z^S\}$ . It follows that

$$\rho(Z) = \sum_{s=1}^S \int_{\frac{s-1}{S}}^{\frac{s}{S}} \phi_\mu(p) z^{\pi^*(s)} dp = \sum_{s=1}^S w^s z^{\pi^*(s)}.$$

Since  $\phi_\mu$  is non-negative and non-decreasing, we have  $0 \leq w^1 \leq w^2 \leq \dots \leq w^S$ . This implies that  $\rho(Z) = \sum_{s=1}^S w^s z^{\pi^*(s)} \geq \sum_{s=1}^S w^s z^{\pi(s)}$  for all  $\pi \in \Pi$ . It follows that

$$\rho(Z) = \max_{\pi \in \Pi} \left\{ \sum_{s=1}^S w^s z^{\pi(s)} \right\},$$

with an optimal solution  $\pi^*$ . □

Applying Lemma 5 to our approximating recourse function  $\tilde{Q}_\alpha^S(\bar{x})$  for some fixed  $\bar{x}$ , we can write

$$\tilde{Q}_\alpha^S(\bar{x}) = \max_{\pi \in \Pi} \sum_{s=1}^S w^s \tilde{v}_\alpha(\xi^{\pi(s)}, \bar{x}),$$

and an optimal solution is found by any  $\pi^* \in \Pi$  that sorts the values  $\tilde{v}_\alpha(\xi^s, \bar{x})$ ,  $s = 1, \dots, S$ . Writing  $\tilde{\pi}$  for the inverse of  $\pi^*$ , we obtain a piecewise linear outer approximation of  $\tilde{Q}_\alpha^S$ :

$$\tilde{Q}_\alpha^S(x) \geq \sum_{s=1}^S w^{\tilde{\pi}(s)} \tilde{v}_\alpha(\xi^s, x), \tag{14}$$

where the inequality follows since  $\pi^*$  is feasible but not necessarily optimal for any  $x$ . So Lemma 5 provides a procedure to derive a piecewise linear outer approximation of the approximating recourse function  $\tilde{Q}_\alpha^S(x)$ . We use this sorting-based linearization to extend the LBDA to the risk-averse case.

### 6.1.2 Algorithm description

We now present our LBDA in the risk-averse setting. In Algorithm 1 we provide pseudo-code for the algorithm. We explain each step below; for more elaborate explanations of some steps we refer to [17].

---

#### Algorithm 1 Risk-averse loose Benders' decomposition algorithm (LBDA)

---

- 1: **Input** Parameters:  $A, b, c, W$ ; distribution of  $\xi$ ; spectral risk measure  $\rho$ ; shift parameter  $\alpha$ ; lower bound  $L$  on  $\tilde{Q}_\alpha^S$ ; sample size  $S$ ; tolerance  $\varepsilon$
  - 2: **Output** Near-optimal solution  $\tilde{x}_\alpha$  for (13)
  - 3: **Initialization**
  - 4: Initialize  $\tau = 0$  and  $\tilde{Q}_{\text{OUT}}^\tau \equiv L$
  - 5: Draw a sample  $\{\xi^1, \dots, \xi^S\}$  of size  $S$  from the distribution of  $\xi$
  - 6: Compute  $w^1, \dots, w^S$  corresponding to the risk measure  $\rho$
  - 7: **Iteration step**
  - 8: Solve  $\min_x \{c^T x + \tilde{Q}_{\text{OUT}}^\tau(x) \mid Ax = b, x \in \mathbb{R}_+^{n_1}\}$
  - 9: Denote optimal solution by  $x^\tau$
  - 10: **for**  $s = 1, \dots, S$  **do**
  - 11:     Solve  $v_{\text{LP}}(\xi^s, x^\tau)$
  - 12:     Denote optimal basis matrix index by  $\tilde{k}_s^\tau$
  - 13:     Compute  $\psi^{\tilde{k}_s^\tau}(h^s - \alpha)$
  - 14:     Define  $u_s^\tau := \lambda^{\tilde{k}_s^\tau}(h^s - T^s x^\tau) + \psi^{\tilde{k}_s^\tau}(h^s - \alpha)$
  - 15: **end for**
  - 16: Find  $\pi^\tau \in \Pi$  such that  $u_{\pi^\tau(1)}^\tau \leq \dots \leq u_{\pi^\tau(S)}^\tau$ ; define  $\tilde{\pi}^\tau := (\pi^\tau)^{-1}$
  - 17:  $\beta_{\tau+1} \leftarrow -\sum_{s=1}^S w^{\tilde{\pi}^\tau(s)} \lambda^{\tilde{k}_s^\tau} T^s$
  - 18:  $\delta_{\tau+1} \leftarrow \sum_{s=1}^S w^{\tilde{\pi}^\tau(s)} (\lambda^{\tilde{k}_s^\tau} h^s + \psi^{\tilde{k}_s^\tau}(h^s - \alpha))$
  - 19:  $\tilde{Q}_{\text{OUT}}^{\tau+1}(x) := \max\{\tilde{Q}_{\text{OUT}}^\tau(x), \beta_{\tau+1}x + \delta_{\tau+1}\}$
  - 20: **Stopping criterion**
  - 21: **if**  $\tilde{Q}_{\text{OUT}}^\tau(x^\tau) \geq \tilde{Q}_{\text{OUT}}^{\tau+1}(x^\tau) - \varepsilon$  **then**
  - 22:     **return**  $\tilde{x}_\alpha := x^\tau$
  - 23: **stop**
  - 24: **else**
  - 25:      $\tau \leftarrow \tau + 1$ . Go to line 8
  - 26: **end if**
-

Besides a description of the underlying risk-averse MIR model, the risk-averse LBDA needs a few other inputs. First, it needs a shift parameter  $\alpha$  for the  $\alpha$ -approximation used in the algorithm. Next, it needs a uniform lower bound  $L$  on  $\tilde{Q}_\alpha^S$  (which always exists since we assume  $X$  is bounded). Finally, it needs a sample size  $S$  for the SAA and a tolerance level  $\varepsilon$  for the stopping criterion. The algorithm will output a near-optimal solution  $\tilde{x}_\alpha$  for (13).

In the initialization phase, we set the iteration counter  $\tau$  to zero and initialize an outer approximation function  $\tilde{Q}_{\text{OUT}}^\tau$  to the uniform lower bound  $L$ . Moreover, we draw a sample from the distribution of  $h$  and we compute the values  $w^1, \dots, w^S$  corresponding to  $\rho$ . Note that these values only need to be computed once.

Every iteration of the algorithm consists of two parts. In the first part, we solve a master problem of the form  $\min_x \{c^T x + \tilde{Q}_{\text{OUT}}^\tau(x) \mid Ax = b, x \in \mathbb{R}_+^{n_1}\}$ , yielding a candidate solution  $x^\tau$ . In the second part, we generate a so-called *loose optimality cut* at  $x^\tau$ . This cut generation procedure consists of two steps, which we will discuss subsequently.

In the first step we derive a *loose* optimality cut for every scenario  $s = 1, \dots, S$ . Note that a tight optimality cut is given by  $\tilde{v}_\alpha(\xi^s, x) \geq \lambda^{k_s^\tau}(\xi^s - T^s x^\tau) + \psi^{k_s^\tau}(h^s - \alpha)$ , where  $k_s^\tau$  is an optimal index in the maximization problem defining  $\tilde{v}_\alpha(\xi^s, x^\tau)$ . It turns out, however, that finding  $k_s^\tau$  exactly is computationally challenging. Therefore, van der Laan and Romeijnders [17] propose to approximate  $k_s^\tau$  by  $\tilde{k}_s^\tau$ , the optimal index in the dual representation of the LP-relaxation. We follow this procedure. See [17] for more details. Moreover, note that computing  $\psi^{k_s^\tau}(h^s - \alpha)$  involves solving the Gomory relaxation  $v_{B^k}(h^s - \alpha)$ , as explained in Lemma 2, which is a mixed-integer program.

In the second step we combine the scenario cuts into a single loose optimality cut for the master problem. We use the procedure from Sect. 6.1.1 to combine the scenarios in a linear way, despite the inherent nonlinearity of our risk measure  $\rho$ . In particular, we find  $\pi^\tau \in \Pi$  that sorts the values  $u_1^\tau, \dots, u_S^\tau$ . Defining  $U^\tau$  as the random variable with a uniform distribution on  $u_1^\tau, \dots, u_S^\tau$  and writing  $\tilde{\pi}^\tau$  as the inverse of  $\pi^\tau$ , this means that  $\rho(U^\tau) = \sum_{s=1}^S w^{\tilde{\pi}^\tau(s)}(\lambda^{\tilde{k}_s^\tau} h^s - T^s x^\tau + \psi^{\tilde{k}_s^\tau}(h^s - \alpha))$ . Now, we define the optimality cut  $\tilde{Q}_{\text{OUT}}^\tau(x) \geq \sum_{s=1}^S w^{\tilde{\pi}^\tau(s)}(\lambda^{\tilde{k}_s^\tau} h^s - T^s x + \psi^{\tilde{k}_s^\tau}(h^s - \alpha))$  and add it to our outer approximation  $\tilde{Q}_{\text{OUT}}^{\tau+1}$ . Our optimality cut is indeed feasible for all  $x$ , since  $\pi^\tau$  is optimal at  $x^\tau$  and feasible in the corresponding maximization problem for all other  $x$ .

The algorithm stops whenever the improvement in the outer approximation as a result of the most recently added optimality cut is less than the predefined tolerance level  $\varepsilon$ . The algorithm outputs the most recent solution  $x^\tau$  as the LBDA solution  $\tilde{x}_\alpha$ .

### 6.2 Performance guarantee

We now present a performance guarantee for our risk-averse LBDA. That is, we derive an upper bound on the optimality gap

$$c\tilde{x}_\alpha + Q(\tilde{x}_\alpha) - \eta^*,$$

where  $\tilde{x}_\alpha$  is the algorithm solution. Our performance guarantee is based on the performance guarantee in [17] for the risk-neutral LBDA. The proof for our risk-averse setting is similar, with a few necessary adjustments to account for risk aversion.

**Theorem 4** *Let  $\tilde{x}_\alpha$  be the solution from the LBDA with tolerance level  $\varepsilon > 0$ . Then, there exists a constant  $C > 0$  such that for every  $f \in \mathcal{H}^m$ ,*

$$c^T \tilde{x}_\alpha + Q(\tilde{x}_\alpha) - \eta^* \leq \varepsilon + \phi_\mu(1) \cdot C \cdot \sum_{i=1}^m \mathbb{E}_{h_{-i}} [|\Delta| f_i(\cdot|h_{-i})]$$

with probability one as  $S \rightarrow +\infty$ .

**Proof** See the appendix. □

The performance guarantee from Theorem 4 is equal to the error bound from Theorem 1, up to the tolerance level  $\varepsilon$ . This shows that, as the sample size goes to  $\infty$ , the risk-averse LBDA produces a solution that is equally good as the solution to the  $\alpha$ -approximation model, in terms of the corresponding performance guarantee. In particular, this shows that the *loose* cuts used in the algorithm do not come at a cost in terms of (provable) solution quality. This is in line with the results for the risk-neutral case in [17].

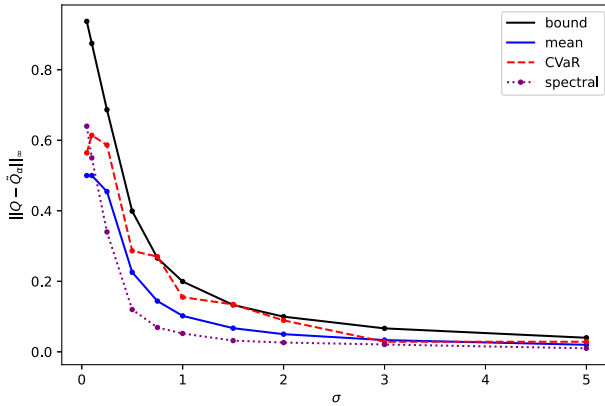
## 7 Numerical experiments

In this section we perform a series of numerical experiments to gain more insight into the performance of our convex approximation approach in risk-averse MIR models. Since the main difference between our paper and the existing results in the literature on convex approximations of risk-neutral MIR models is the introduction of a risk measure  $\rho$ , we will focus primarily on the effect of the risk measure  $\rho$  on different performance measures. For a numerical analysis of convex approximations and the LBDA in a risk-neutral setting, we refer to [13] and [17], respectively.

We will explore our convex approximations approach in a risk-averse one-dimensional SIR setting. We consider this setting for two reasons. First, it allows for *exact* computation of approximation errors, optimal solutions, and optimality gaps. Hence, we avoid having to resort to estimation of performance guarantees, which would make our results more noisy and which would hinder interpretation. Second, the one-dimensional setting makes it easier to plot certain results, again aiding interpretation of the results.

### 7.1 Asymptotic behavior

Before zooming in on the risk measure  $\rho$  we first test whether the asymptotic behavior of the error bounds from Theorem 1 and 2 is similar to the behavior of the actual approximation error  $\|Q - \tilde{Q}_\alpha\|_\infty$ . We consider a one-dimensional SIR model with  $q^- = 0$  and  $q^+ = 1$ , where the right-hand side variable  $h$  follows a normal distribution with mean zero and standard deviation  $\sigma > 0$ . We consider three different



**Fig. 1** Maximum approximation error for a one-dimensional SIR model with  $q^- = 0$  and  $q^+ = 1$ , for three different risk measures, where the underlying distribution is a normal distribution truncated at its mean  $\mu = 0$  with varying values of the standard deviation  $\sigma$ . Also plotted is the error bound from Theorem 5 in [16] corresponding to the risk-neutral case  $\rho = \mathbb{E}$

risk measures: the expected value  $\rho = \mathbb{E}$  (denoted “mean”),  $\rho = \text{CVaR}_{0.8}$  (denoted “CVaR”), and the spectral risk measure defined by the risk spectrum  $\phi_\mu(p) = 2p$ ,  $p \in [0, 1]$  (denoted “spectral”).

For each risk measure, we compute the maximum approximation error  $\|Q - \tilde{Q}_\alpha\|_\infty$  for  $\sigma \in \{0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5\}$ . The results are presented in Fig. 1. For comparison, we also plot the error bound from Theorem 2 corresponding to  $\rho = \mathbb{E}$ . Note that the error bounds for the other risk measures can be obtained by multiplying this bound by the factor  $\phi_\mu(1)$  corresponding to the respective risk measure. This factor equals 5 for CVaR and 2 for spectral.

From Fig. 1 we observe that for all risk measures the maximum approximation error  $\|Q - \tilde{Q}_\alpha\|_\infty$  converges to zero as  $\sigma$  increases. The rate of convergence is comparable to that of the error bound. This confirms that in this setting our asymptotic interpretation of the error bound translates to the actual approximation error.

Interestingly, in contrast with the approximation error in the risk-neutral case, the approximation error for CVaR exhibits an erratic pattern; compare the solid and dashed graph in Fig. 1. To understand this behavior, recall that we can represent  $\text{CVaR}_{0.8}$  as an expected value under an adjusted pdf, which has a sharp peak at the 80th percentile of the original distribution. This peak may be located at a point where the approximation error  $v(h, x) - \tilde{v}_\alpha(h, x)$  is particularly large. Depending on the shape of the rest of the adjusted pdf (which depends on  $\sigma$ ), the effect of this peak may be mitigated by errors in the other direction or not. If so, the corresponding value of  $\|Q - \tilde{Q}_\alpha\|_\infty$  will be relatively small; if not, it will be large. This unpredictable behavior can be seen as the reason why the error bound for CVaR needs to be much higher than the one for mean, even though in most cases the corresponding approximation errors are comparable, as we can see from the graph.

In case of the risk measure “spectral”, we have a similar effect as for CVaR, but now there infinitely many peaks, one for each value of  $\beta$  for which the risk spectrum  $\phi_\mu(\beta)$  is increasing. It is possible that these peaks line up in a way that they magnify



each other’s effect on  $\|Q - \tilde{Q}_\alpha\|_\infty$ , which is reflected in the corresponding error bound. However, in many situations some of the peaks will cancel out the effect of other peaks, resulting in a low value for  $\|Q - \tilde{Q}_\alpha\|_\infty$ . This is indeed what we observe in Fig. 1: the graph for spectral is even *below* that of mean, while its corresponding error bound is more than twice as large. This result suggests that combining multiple CVaRs into a spectral risk measure can lead to small approximation errors. We will analyze this topic in more detail in Sect. 7.2.

### 7.2 Smoothness of the risk spectrum

Recall that the error bounds from Theorem 1 and Theorem 2 depends on the risk measure  $\rho$  only through the associated value  $\sup_{\mu \in \mathcal{M}} \phi_\mu(1)$ . Since we only use a very limited amount of information about the risk measure  $\rho$ , we conjecture that the resulting bound may be overly conservative for some risk measures. In particular, we are interested how, for a given  $\mu \in \mathcal{M}$ , the shape of  $\phi_\mu$  (so not only its maximum value) affects the approximation error.

To investigate this matter, we consider two spectral risk measures with the same associated value  $\phi_\mu(1) = 2$ . These risk measures differ in the “smoothness” of the associated risk spectrum  $\phi_\mu$ . On the one hand we have CVaR<sub>0.5</sub>, whose risk spectrum  $\phi_\mu$  is a step function with one jump at 0.5 (see Fig. 2). In this subsection we will refer to this risk measure simply as “CVaR”. On the other end we have the spectral risk measure associated with the risk spectrum

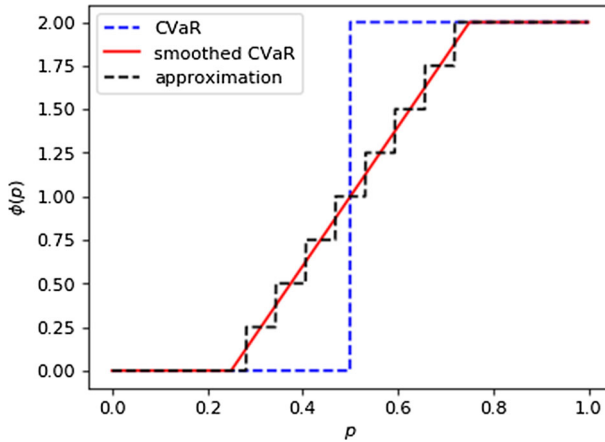
$$\phi(p) = \begin{cases} 0 & \text{if } p < 0.25, \\ 4(p - 0.25) & \text{if } 0.25 \leq p \leq 0.75, \\ 1 & \text{if } p > 0.75. \end{cases}$$

This risk spectrum can be seen as a “smoothed” transformation of the risk spectrum for CVaR. Hence, we refer to the corresponding risk measure as “smoothed CVaR”.

Observe that smoothed CVaR is a convex combination of an infinite amount of CVaRs. For numerical reasons, we will not work with smooth itself, but approximate it with convex combinations of a finite number of CVaRs. For a given  $n \in \mathbb{N}$ , we can approximate smooth CVaR by the risk measure  $\rho = \sum_{k=1}^n \mu_k \text{CVaR } \beta_k$ , where  $\beta_k = \frac{k-1/2}{n}$  and  $\mu_k = \frac{2(1-\beta_k)}{n}$ ,  $k = 1, \dots, n$ . See Fig. 2 for an illustration of the corresponding risk spectrum. In our experiments we will use  $n = 5$ .

For each of the two risk measures, we compute the maximum approximation error  $\|Q - \tilde{Q}_\alpha\|_\infty$  in a one-dimensional SIR setting with  $q^- = 0$  and  $q^+ = 1$ , where  $h$  follows a Weibull distribution with shape parameter  $k$  and scale parameter  $\lambda$ . The reason for choosing the Weibull family is that it exhibits a wide variety of different shapes of the pdf, depending on the parameters, thereby increasing our confidence that the results are representative for more general cases. In Table 1 we present the difference between the maximum approximation errors for these two risk measures.

From Table 1 we see that CVaR has a larger corresponding maximum approximation error than smoothed CVaR in all settings, except for  $k = 0.6, \lambda = 1.25$ . These results suggest that generally speaking, spectral risk measures with a smoother associated risk



**Fig. 2** A plot of the risk spectra corresponding to CVaR, smoothed CVaR, and the approximation of the latter (with  $n = 4$ ), as discussed in Sect. 7.2

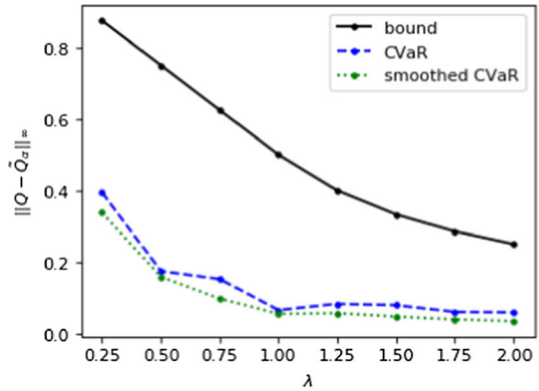
**Table 1** Difference between  $\|Q - \tilde{Q}_\alpha\|_\infty$  for CVaR and for smoothed CVaR, for a one-dim. SIR model with  $q^+ = 1, q^- = 0$ , and a Weibull distribution with various settings for the shape parameter  $k$  and scale parameter  $\lambda$ . Also presented are the average for each row and column

$k$	$\lambda$								avg
	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	
0.5	0.054	0.037	0.028	0.036	0.010	0.000	0.014	0.018	0.025
0.6	0.042	0.045	0.038	0.028	-0.002	0.017	0.027	0.020	0.027
0.7	0.069	0.036	0.047	0.017	0.011	0.026	0.024	0.016	0.031
0.8	0.075	0.016	0.055	0.010	0.019	0.033	0.025	0.018	0.031
0.9	0.065	0.015	0.063	0.006	0.020	0.035	0.021	0.021	0.031
1	0.057	0.016	0.055	0.010	0.026	0.032	0.021	0.024	0.030
1.1	0.053	0.016	0.048	0.015	0.033	0.027	0.024	0.028	0.030
1.2	0.051	0.016	0.041	0.022	0.030	0.031	0.027	0.032	0.031
1.3	0.046	0.017	0.035	0.037	0.037	0.022	0.031	0.034	0.032
1.4	0.050	0.017	0.030	0.037	0.038	0.025	0.034	0.029	0.032
1.5	0.057	0.017	0.026	0.045	0.034	0.014	0.036	0.023	0.032
avg	0.056	0.023	0.042	0.024	0.023	0.024	0.026	0.024	0.030

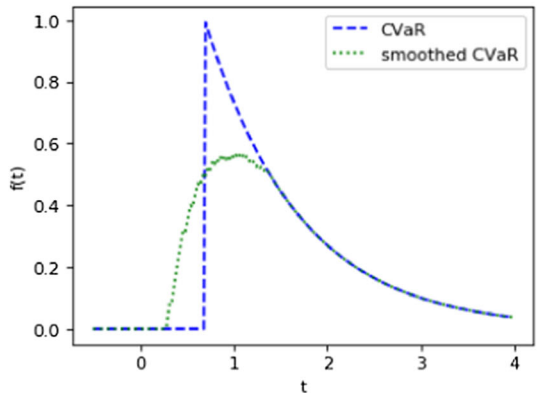
spectrum lead to lower approximation errors. To understand these results, we provide plots of the maximum approximation errors and the corresponding adjusted pdfs for  $k = 1$  and varying values for  $\lambda$  in Fig. 3.

In Fig. 3a we see that the graphs of the maximum approximation error for both CVaR and smoothed CVaR decrease as  $\lambda$  increases, similar to the associated error bound. This is in line with the error bounds from Sect. 3 and 4, as the distribution of  $h$  becomes more dispersed with  $\lambda$ . Also, we see a more erratic behavior for CVaR than for smoothed CVaR, which is also in line with our results in Sect. 7.1.

**Fig. 3** Plots related to a one-dimensional SIR model with  $q^+ = 1, q^- = 0$ , under a Weibull distribution with  $k = 1$  and varying values for  $\lambda$ , for two different underlying risk measures: CVaR and smoothed CVaR. Top: the maximum approximation error as a function of  $\lambda$ . The corresponding error bound from Theorem 2 is also plotted. Bottom: the adjusted pdf for each of the two risk measures, for  $k = 1$  and  $\lambda = 1$



(a) Approximation error



(b) Distorted pdf

More importantly, the graph of CVaR is uniformly above that of smoothed CVaR, as we already learned from Table 1. Figure 3b provides an explanation for this finding. The adjusted pdf corresponding to CVaR has a sharp peak at the 50th percentile of the original distribution. For smoothed CVaR, which is a convex combination of a multitude of CVaRs, many such peaks are aggregated, resulting in a lower, less sharp peak. As a result, the total variation of the adjusted pdf of CVaR is significantly higher than for smoothed CVaR. Hence, based on risk-neutral error bounds, we expect the maximum approximation error for CVaR to be larger, which is indeed what we find.

These results suggest that spectral risk measures with relatively smooth risk spectra lead to relatively low approximation errors. So interestingly, this finding suggests that finding good solutions for models with a more complicated mixed CVaR risk measure might in fact be *easier* than for models with a simple CVaR risk measure. This is especially interesting from a modeling point of view, given the fact that risk preferences might be more accurately reflected by these more complicated mixed CVaR risk measures than by a single CVaR.

**Table 2** Solution times for the numerical experiments in Sect. 7.3 in seconds, rounded to the nearest integer

Instance	Expectation		CVaR		Mean-CVaR		Spectral	
	DEF	LBDA	DEF	LBDA	DEF	LBDA	DEF	LBDA
sslp_10_50_50	3600	68	8	45	3600	61	551	65
sslp_10_50_100	3600	140	338	109	3600	109	3600	119
sslp_10_50_500	3600	552	3600	586	3600	481	NA	545
sslp_10_50_1000	3600	1125	3600	1186	3600	1135	NA	1100
sslp_10_50_2000	3600	2283	3600	2285	3600	2216	NA	2198

### 7.3 Risk-averse LBDA performance

Finally, we investigate the performance of our risk-averse LBDA. A comparison of the performance of the risk-neutral LBDA with alternative solution methods has already been performed in [17], showing that the LBDA is indeed competitive in the risk-neutral case. Here, we focus on investigating the effect of the choice of risk measure  $\rho$  on the performance of the algorithm.

For this purpose, we consider the stochastic server location problem (SSLP) from the stochastic integer programming library SIPLIB [34]. The first- and second-stage problems of the SSLP are pure binary and mixed-binary, respectively, and the largest instance has over 1,000,000 binary decision variables and 120,000 constraints. A full description can be found in [34]. We solve instances with 10 server locations and 50 potential clients, which are denoted by `sslp_10_50_XXXX`, where `XXXX` represents the number of scenarios, which is taken from the set {50, 100, 500, 1000, 2000}. This allows us to test the effect of an increase in the number of scenarios on the performance of the algorithm. We use four different risk measures: (1) the *expectation*, (2) a *CVaR* risk measure with parameter  $\beta = 0.8$ , (3) a *mean-CVaR* risk measure which is an equally weighted average between the former two risk measures, and (4) a *spectral* risk measure defined by the risk spectrum  $\phi_\mu(p) = 2p$ ,  $p \in [0, 1]$ . For every combination of problem instance and risk measure, we run our LBDA algorithm, as well as a deterministic equivalent formulation with a computation time limit of 1 h. The instances are run using Gurobi 10.0.0 on an 11th Gen Intel(R) Core(TM) i5-1145G7 processor running at 1.50 GHz.

The results of the experiments are given in Tables 2 and 3. First, Table 2 shows the computation times of the LBDA and the deterministic equivalent formulation. We observe that the LBDA is able to find a solution within less than the 1-h time limit in all instances. The computation time grows more or less linearly in the number of scenarios. The deterministic equivalent, on the other hand, reaches the 1-h cutoff mark in almost all instances. Moreover, the deterministic equivalent formulation of the instances with 500–2000 scenarios and the spectral risk measure break down due to a memory overflow, as this naive implementation requires various additional variables for every scenario in order to deal with the spectral risk measure. The LBDA, on the other hand, does not have this problem. It finds a solution within a computation time comparable to other risk measures. This illustrates one advantage of our risk-averse

**Table 3** Objective function values for the numerical experiments in Sect. 7.3 (rounded to the nearest integer). For each combination of problem instance and risk measure, we present the objective function value of the solution found by the deterministic equivalent formulation (DEF) and by the LBDA. For instances in which the DEF does not find a solution within the time limit, we report the optimality gap between brackets

Risk measure $\sup_{\mu \in \mathcal{M}} \phi_{\mu}(1)$	Expectation 1		CVaR 5		Mean-CVaR 3		Spectral 2	
	DEF	LBDA	DEF	LBDA	DEF	LBDA	DEF	LBDA
<b>Instance</b>	DEF	LBDA	DEF	LBDA	DEF	LBDA	DEF	LBDA
sslp_10_50_50	-365 (0.11%)	-365	-279	-275	-322 (0.02%)	-314	-333	LBDA -333
sslp_10_50_100	-354 (0.10%)	-354	-264	-249	-305 (0.05%)	-305	-319 (0.02%)	-319
sslp_10_50_500	-334 (4.74%)	-349	-261 (2.17%)	-261	-305 (0.14%)	-305	NA	-315
sslp_10_50_1000	-336 (5.44%)	-352	-190 (47.7%)	-258	-123 (15.4%)	-308	NA	-318
sslp_10_50_2000	-169 (110%)	-347	-53 (43.4%)	-254	-111 (180%)	-301	NA	-312

LBDA over methods based on risk-neutral reformulations: it can easily deal with arbitrary spectral risk measures.

Next, Table 3 shows the objective function values of the solutions obtained by the LBDA and the deterministic equivalent formulation. We observe that for smaller instances, the LBDA finds the optimal solution for the risk measures expectation and spectral, whereas there is a positive optimality gap for the risk measures mean-CVaR and especially CVaR, due to the fact that LBDA only finds an approximate solution. Observe that the latter two have higher values for  $\sup_{\mu \in \mathcal{M}} \phi_{\mu}(1)$  than the former two. Hence, this result is in line with our performance guarantee from Theorem 4. For larger instances, we observe that the performance of the deterministic equivalent formulation deteriorates as the number of scenarios increases, while that of the LBDA remains consistent. This illustrates that the LBDA is well-equipped to solve large-scale problems.

We conclude that the numerical performance of the LBDA is in line with what can be expected from the performance guarantee in Theorem 4 and is particularly prominent in larger problem instances.

## 8 Conclusion

In this paper we develop convex approximations of risk-averse MIR models. We extend existing performance guarantees for these convex approximations to a general setting with an arbitrary law invariant coherent risk measure. The performance guarantee turns out to depend on the maximum distortion factor  $\sup_{\mu \in \mathcal{M}} \phi_{\mu}(1)$  corresponding to the risk measure  $\rho$ . We also show how our performance guarantee extends to an SAA setting and we extend an SAA-based Benders' decomposition algorithm from the literature to our risk-averse setting. Finally, we show in numerical experiments that the performance of the LBDA is in line with our theoretical performance guarantees. Moreover, it is capable of dealing with arbitrary spectral risk measures and is particularly suitable for solving large problem instances.

Future research may be aimed at deriving tighter error bounds than the ones in this paper. We present worst-case error bounds that hold under very general conditions, but our numerical experiments suggest that, at least in some cases, there might be room for improvement. For example, it would be interesting to see whether tighter error bounds can be derived by using more information about the risk measure  $\rho$  than our error bounds use. Another interesting avenue for future research is to investigate whether our convex approximation approach extends to risk-averse settings beyond the law invariant coherent case considered in this paper. Our error bound for the non-coherent risk measure expected excess can be seen as a first step in this direction.

**Acknowledgements** We thank Niels van der Laan and Niels Wouda for their help with the LBDA implementation from [35]. The research of Ward Romeijnders has been supported by Grant 451-17-034 4043 from The Netherlands Organisation for Scientific Research (NWO).

**Data availability** We did not use any externally collected data in our numerical experiments. The settings used in the experiments are of a theoretical nature and they are completely described in the manuscript itself.

We made use of the SSLP instances described by Ntaimo and Sen [34], which can be found online; see [37].

### Declarations

**Conflict of interest** The authors have no conflict of interest to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### Appendix: proof of Theorem 4

In this subsection we prove the performance guarantee on the solution from the risk-averse LBDA from Theorem 4. The proof is analogous to the proof of the performance guarantee for the risk-neutral LBDA in [17]. First, we show that  $\tilde{Q}_\alpha^S$  converges to  $\tilde{Q}_\alpha$  as  $S \rightarrow +\infty$ . To do so, we need to prove a strong law of large numbers (SLLN) for coherent risk measures.

**Lemma 6** *Consider the function  $\tilde{Q}_\alpha$  from Definition 3 and its sample average approximation  $\tilde{Q}_\alpha^S$ . Then,*

$$\sup_{x \in X} |\tilde{Q}_\alpha^S(x) - \tilde{Q}_\alpha(x)| \rightarrow 0,$$

with probability one as  $S \rightarrow +\infty$ .

**Proof** Using the strong law of large numbers for law invariant coherent risk measures from [36], the proof follows analogously to the proof of Lemma 2 in [17], if we can show that  $\tilde{Q}_\alpha$  and  $\tilde{Q}_\alpha^S$  are Lipschitz continuous.

It is not hard to see that  $\tilde{v}_\alpha(\xi, \cdot)$  is Lipschitz continuous (with respect to the  $\ell_1$  norm) with Lipschitz constant  $L := \max_{i=1, \dots, m} \{ \max_{(q, T) \in \Xi^q \times \Xi^T} \{ \max_{k \in K^q} \{ |(q_{B^k}(B^k)^{-1}T)_i| \} \} \}$ , for every  $\xi \in \Xi$ . To prove Lipschitz continuity of  $\tilde{Q}_\alpha = \rho(\tilde{v}_\alpha(\xi, \cdot))$ , we use the following duality representation of the coherent risk measure  $\rho$ : there exists a collection  $\mathcal{Q}$  of probability measures on  $(\Omega, \mathcal{F})$  such that  $\rho(Z) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[Z]$ , for every  $Z \in \mathcal{Z}$  (see, e.g., [23]). Let  $x, x' \in \mathbb{R}^{n_1}$  be given. Then, using Lipschitz continuity of  $\tilde{v}_\alpha(\xi, \cdot)$ , we obtain

$$\begin{aligned} |\tilde{Q}_\alpha(x) - \tilde{Q}_\alpha(x')| &= \left| \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[\tilde{v}_\alpha(\xi, x)] - \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[\tilde{v}_\alpha(\xi, x')] \right| \\ &\leq \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}} \left[ |\tilde{v}_\alpha(\xi, x) - \tilde{v}_\alpha(\xi, x')| \right] \end{aligned}$$

$$\leq \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q \left[ L \cdot \|x - x'\|_1 \right] = L \cdot \|x - x'\|_1,$$

which proves that  $\tilde{Q}_\alpha$  is Lipschitz continuous with Lipschitz constant  $L$ . Similarly,  $\hat{Q}_\alpha^S$  is Lipschitz continuous as well. This completes the proof.  $\square$

Next, in Lemma 7 we consider the loose optimality cuts. To show that these are nearly tight at the current solution  $(x^\tau, \pi^\tau)$ , we define the loose optimality cut-based function  $\hat{Q}_\alpha^S(x) := \rho(\hat{v}_\alpha(\hat{\xi}_S, x))$ ,  $x \in \mathbb{R}^{n_1}$ , where the value function  $\hat{v}_\alpha$  is defined by

$$\hat{v}_\alpha(\xi, x) := q_{B^k(\xi, x)}(B^{k(\xi, x)})^{-1}(h - Tx) + \psi^{k(\xi, x)}(h - \alpha), \quad \xi \in \Xi, x \in \mathbb{R}^{n_1},$$

with  $k(\xi, x) \in \operatorname{argmax}_{k \in K^q} \{q_{B^k}(B^k)^{-1}(h - Tx)\}$ ,  $\xi \in \Xi, x \in \mathbb{R}^{n_1}$ . We provide an upper bound on the approximation error resulting from the loose optimality cuts.

**Lemma 7** *There exists a constant  $C > 0$  such that for every  $f \in \mathcal{H}^m$ ,*

$$\sup_{x \in X} |\tilde{Q}_\alpha^S(x) - \hat{Q}_\alpha^S(x)| \leq \phi_\mu(1) \cdot C \sum_{i=1}^m \mathbb{E}^{\mathbb{P}} [|\Delta| f_i(\cdot | h_{-i})]$$

with probability one as  $S \rightarrow +\infty$ .

**Proof** The proof is along the same lines of the proof of Proposition 1 in [17], but several changes are necessary since we consider a risk-averse setting and since we allow for  $q$  and  $T$  to be random as well.

First, define  $\Delta(\xi, x) := \tilde{v}_\alpha(\xi, x) - \hat{v}_\alpha(\xi, x)$ ,  $\xi \in \Xi, x \in \mathbb{R}^{n_1}$ , and note that  $\Delta(\xi, x) \geq 0$  for all  $\xi \in \Xi$  and  $x \in \mathbb{R}^{n_1}$ , since  $\hat{v}_\alpha$  is a lower bound for  $\tilde{v}_\alpha$ . This implies that  $\tilde{Q}_\alpha^S(x) \geq \hat{Q}_\alpha^S(x)$  for all  $x \in \mathbb{R}^{n_1}$ . Using this fact and applying Lemma 5 to both  $\tilde{Q}_\alpha^S(x)$  and  $\hat{Q}_\alpha^S(x)$ , we obtain

$$\begin{aligned} |\tilde{Q}_\alpha^S(x) - \hat{Q}_\alpha^S(x)| &= \tilde{Q}_\alpha^S(x) - \hat{Q}_\alpha^S(x) \\ &= \max_{\pi \in \Pi} \left\{ \sum_{s=1}^S w^{\pi(s)} \tilde{v}_\alpha(\xi^s, x) \right\} - \max_{\pi \in \Pi} \left\{ \sum_{s=1}^S w^{\pi(s)} \hat{v}_\alpha(\xi^s, x) \right\} \\ &\leq \max_{\pi \in \Pi} \left\{ \sum_{s=1}^S w^{\pi(s)} \tilde{v}_\alpha(\xi^s, x) - \sum_{s=1}^S w^{\pi(s)} \hat{v}_\alpha(\xi^s, x) \right\} \\ &= \max_{\pi \in \Pi} \left\{ \sum_{s=1}^S w^{\pi(s)} \Delta(\xi^s, x) \right\}. \end{aligned} \tag{15}$$

Similar as in the proof of Proposition 1 in [17], we can derive an upper bound  $\bar{\eta}(\xi)$  on  $\Delta(\xi, x)$  that does not depend on  $x$ . For this purpose, define  $\mathcal{M}_q := \cup_{k \in K^q} (\sigma_k + \Lambda^k)$  and  $\mathcal{M}'_q := \{s \in \mathcal{M}_q \mid \mathcal{B}(s; D) \subseteq \mathcal{M}_q\}$ , where  $\mathcal{B}(s; D)$  denotes the closed ball



around  $s$  with radius  $D := \sup_{T \in \Xi^T} \left\{ \sup_{x, x' \in X} \{ \|Tx - Tx'\|_2 \} \right\}$ . Then, it follows from the proof of Proposition 1 in [17] that there exists a constant  $Z > 0$  such that

$$\Delta(\xi, x) \leq \bar{\eta}(\xi) := \begin{cases} 0 & \text{if } h - Tx^* \in \mathcal{M}'_q, \\ Z & \text{if } h - Tx^* \notin \mathcal{M}'_q. \end{cases}$$

Denoting the optimal argument in (15) by  $\tilde{\pi}$ , we obtain

$$|\tilde{Q}_\alpha^S(x) - \hat{Q}_\alpha^S(x)| \leq \sum_{s=1}^S w^{\tilde{\pi}(s)} \Delta(\xi^s, x) \leq (S \cdot w^S) \cdot \left( \frac{1}{S} \cdot \sum_{s=1}^S \bar{\eta}(\xi^s) \right).$$

Here, we have the limits  $\lim_{S \rightarrow +\infty} S \cdot w^S = \lim_{S \rightarrow +\infty} S \cdot \int_{\frac{S-1}{S}}^S \phi_\mu(p) dp = \phi_\mu(1)$ , and (using the SLLN for law invariant coherent risk measures from [36]),  $\lim_{S \rightarrow +\infty} \frac{1}{S} \cdot \sum_{s=1}^S \bar{\eta}(\xi^s) = \mathbb{E}_\mathbb{P}[\bar{\eta}(\xi)] = Z \cdot \mathbb{P}\{h - Tx^* \in \mathcal{M}'_q\}$ . It follows that

$$|\tilde{Q}_\alpha^S(x) - \hat{Q}_\alpha^S(x)| \leq \phi_\mu(1) \cdot Z \cdot \mathbb{P}\{h - Tx^* \in \mathcal{M}'_q\}. \tag{16}$$

By the proof of Theorem 2 in [17], it follows that for any fixed  $\bar{q} \in \Xi^q$ , there exists a constant  $\beta^{\bar{q}} > 0$  such that  $\mathbb{P}\{h - Tx \in \mathcal{M}'_{\bar{q}}\} \leq \beta^{\bar{q}} \sum_{i=1}^m \mathbb{E}_\mathbb{P}[|\Delta|f_i(\cdot | h_{-i})]$ . Hence,

$$\begin{aligned} \mathbb{P}\{h - Tx^* \notin \mathcal{M}'_q\} &= \sum_{\bar{q} \in \Xi^q} \mathbb{P}\{q = \bar{q}\} \cdot \mathbb{P}\{h - Tx^* \notin \mathcal{M}'_{\bar{q}}\} \\ &\leq \sum_{\bar{q} \in \Xi^q} \mathbb{P}\{q = \bar{q}\} \cdot \beta^{\bar{q}} \cdot \sum_{i=1}^m \mathbb{E}_\mathbb{P}[|\Delta|f_i(\cdot | h_{-i})] \\ &= \beta \cdot \sum_{i=1}^m \mathbb{E}_\mathbb{P}[|\Delta|f_i(\cdot | h_{-i})], \end{aligned}$$

where  $\beta := \sum_{\bar{q} \in \Xi^q} \mathbb{P}\{q = \bar{q}\} \cdot \beta^{\bar{q}}$ . Now, defining  $C := \beta \cdot Z$  and combining this with (16), we obtain

$$|\tilde{Q}_\alpha^S(x) - \hat{Q}_\alpha^S(x)| \leq \phi_\mu(1) \cdot C \cdot \sum_{i=1}^m \mathbb{E}_\mathbb{P}[|\Delta|f_i(\cdot | h_{-i})].$$

The result now follows from the fact that the bound above does not depend on  $x$ .  $\square$

We are now ready to prove our main result, providing a performance guarantee for the LBDA solution  $\tilde{x}_\alpha$ .

**Proof of Theorem 4** Using the Lemmas above, the proof is completely analogous to the proof of Theorem 2 in [17].  $\square$

## References

1. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming, 2nd edn. Springer, New York (2011)
2. Nowak, M.P., Römis, W.: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. *Ann. Oper. Res.* **100**(1), 251–272 (2000)
3. Alonso-Ayuso, A., Escudero, L.F., Ortuño, M.T., Pizarro, C.: On a stochastic sequencing and scheduling problem. *Comput. Oper. Res.* **34**(9), 2604–2624 (2007)
4. Lei, H., Laporte, G., Guo, B.: The capacitated vehicle routing problem with stochastic demands and time windows. *Comput. Oper. Res.* **38**(12), 1775–1783 (2011)
5. Ahmed, S.: Convexity and decomposition of mean-risk stochastic programs. *Math. Program.* **106**(3), 433–446 (2006)
6. Miller, N., Ruszczyński, A.: Risk-averse two-stage stochastic linear programming: modeling and decomposition. *Oper. Res.* **59**(1), 125–132 (2011)
7. Noyan, N.: Risk-averse two-stage stochastic programming with an application to disaster management. *Comput. Oper. Res.* **39**(3), 541–559 (2012)
8. Rockafellar, R.T.: Solving stochastic programming problems with risk measures by progressive hedging. *Set-Valued Var. Anal.* **26**, 759–768 (2018)
9. Schultz, R., Tiedemann, S.: Conditional value-at-risk in stochastic programs with mixed-integer recourse. *Math. Program.* **105**(2–3), 365–386 (2006)
10. Soleimani, H., Seyyed-Esfahani, M., Kannan, G.: Incorporating risk measures in closed-loop supply chain network design. *Int. J. Prod. Res.* **52**(6), 1843–1867 (2014)
11. Alem, D., Clark, A., Moreno, A.: Stochastic network models for logistics planning in disaster relief. *Eur. J. Oper. Res.* **255**(1), 187–206 (2016)
12. Klein Haneveld, W.K., Stougie, L., van der Vlerk, M.H.: Simple integer recourse models: convexity and convex approximations. *Math. Program.* **108**(2–3), 435–473 (2006)
13. Romeijn, W., Morton, D.P., van der Vlerk, M.H.: Assessing the quality of convex approximations for two-stage totally unimodular integer recourse models. *INFORMS J. Comput.* **29**(2), 211–231 (2017)
14. Romeijn, W., Schultz, R., van der Vlerk, M.H., Klein Haneveld, W.K.: A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM J. Optim.* **26**(1), 426–447 (2016)
15. Romeijn, W., van der Vlerk, M.H., Klein Haneveld, W.K.: Convex approximations for totally unimodular integer recourse models: a uniform error bound. *SIAM J. Optim.* **25**(1), 130–158 (2015)
16. Romeijn, W., van der Vlerk, M.H., Klein Haneveld, W.K.: Total variation bounds on the expectation of periodic functions with applications to recourse approximations. *Math. Program.* **157**(1), 3–46 (2016)
17. van der Laan, N., Romeijn, W.: A loose Benders decomposition algorithm for approximating two-stage mixed-integer recourse models. *Math. Program.* **190**, 1–34 (2020)
18. van der Laan, N., Romeijn, W., van der Vlerk, M.H.: Higher-order total variation bounds for expectations of periodic functions and simple integer recourse approximations. *CMS* **3**(15), 325–349 (2018)
19. van der Vlerk, M.H.: Convex approximations for complete integer recourse models. *Math. Program.* **99**(2), 297–310 (2004)
20. van der Vlerk, M.H.: Convex approximations for a class of mixed-integer recourse models. *Ann. Oper. Res.* **177**(1), 139–150 (2010)
21. van Beesten, E.R., Romeijn, W.: Parametric error bounds for convex approximations of two-stage mixed-integer recourse models with a random second-stage cost vector. *Oper. Res. Lett.* **50**(5), 541–547 (2022)
22. van Beesten, E.R., Romeijn, W.: Convex approximations for two-stage mixed-integer mean-risk recourse models with conditional value-at-risk. *Math. Program.* **181**, 473–507 (2020)
23. Shapiro, A.: On Kusuoka representation of law invariant risk measures. *Math. Oper. Res.* **38**(1), 142–152 (2013)
24. Kleywegt, A.J., Shapiro, A., Homem-de-mello, T.: The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* **12**(2), 479–502 (2002)
25. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. *Math. Finance* **9**(3), 203–228 (1999)

26. Acerbi, C.: Spectral measures of risk: a coherent representation of subjective risk aversion. *J. Bank. Finance* **26**(7), 1505–1518 (2002)
27. van Beesten, E.R., Romeijnders, W., Morton, D.P.: Pragmatic distributionally robust optimization for simple integer recourse models. arXiv preprint [arXiv:2206.13250](https://arxiv.org/abs/2206.13250) (2022)
28. Kusuoka, S.: On law invariant coherent risk measures. In: Kusuoka, S., Maruyama, T. (eds.) *Advances in Mathematical Economics*, vol. 3, pp. 83–95. Springer, Tokyo (2001)
29. van Beesten, E.R.: Pragmatic convex approaches for risk-averse and distributionally robust mixed-integer recourse models. PhD thesis, University of Groningen (2022)
30. Pflug, G.C., Römisch, W.: *Modeling, Measuring and Managing Risk*. World Scientific, Singapore (2007)
31. Louveaux, F.V., van der Vlerk, M.H.: Stochastic programming with simple integer recourse. *Math. Program.* **61**(1), 301–325 (1993)
32. Varberg, D.E.: On absolutely continuous functions. *Am. Math. Mon.* **72**(8), 831–841 (1965)
33. Russo, R.P., Shyamalkumar, N.D.: Bounds for the bias of the empirical CTE. *Insurance Math. Econom.* **47**(3), 352–357 (2010)
34. Ntaimo, L., Sen, S.: The million-variable “march” for stochastic combinatorial optimization. *J. Global Optim.* **32**, 385–400 (2005)
35. van der Laan, N., Romeijnders, W., Wouda, N.: LBDA+: A loose Benders decomposition algorithm for approximating two-stage mixed-integer recourse models. *DataverseNL* (2020). <https://doi.org/10.34894/45RPGb>
36. Shapiro, A.: Consistency of sample estimates of risk averse stochastic programs. *J. Appl. Probab.* **50**(2), 533–541 (2013)
37. Ntaimo, L., Sen, S.: The SSLP test problems. <https://www2.isye.gatech.edu/sahmed/siplib/sslp/sslp.html>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.