Bounds on Risk-averse Mixed-integer Multi-stage Stochastic Programming Problems with Mean-CVaR

Ali İrfan Mahmutoğulları , Özlem Çavuş*, and M. Selim Aktürk

Department of Industrial Engineering, Bilkent University, 06800 Ankara, TURKEY

Abstract

Risk-averse mixed-integer multi-stage stochastic programming forms a class of extremely challenging problems since the problem size grows exponentially with the number of stages, the problem is non-convex due to integrality restrictions and the objective function is a dynamic measure of risk. For this reason, we propose a scenario tree decomposition approach, namely group subproblem approach, to obtain bounds for such problems with an objective of dynamic mean-CVaR risk measure. Our approach does not require any special problem structure such as convexity and linearity, therefore it can be applied to a wide range of problems. We obtain lower bounds by using different convolution of mean-CVaR risk measures and different scenario partition strategies. The upper bounds are obtained through the use of optimal solutions of group subproblems. Using these lower and upper bounds, we propose an algorithm for risk-averse mixed-integer multi-stage stochastic problems with mean-CVaR risk measures. We test the performance of the proposed algorithm on a multi-stage stochastic lot sizing problem and compare different choices of lower bounds and partition strategies. Comparison of the proposed algorithm and the commercial solver revealed that, on the average, the proposed algorithm yields 2.58 times stronger bounds compared to a commercial solver.

Keywords: Stochastic programming; Mixed-integer multi-stage stochastic programming; Dynamic measures of risk; CVaR; Bounding.

1 Introduction

In risk-averse stochastic optimization problems, risk measures are used to assess the risk involved in the decisions made. Due to the structural properties of risk measures, risk-averse models are more challenging than their risk-neutral counterparts. The multi-stage risk-averse stochastic models are even more complicated due to their dynamic nature and excessive amount of decision variables. Both the risk-neutral and risk-averse multi-stage stochastic problems are non-convex when some of the decision variables are required to be integer valued. Therefore, the solution methods suggested for convex multi-stage stochastic problems cannot be used to solve these problems.

In this study, we consider risk-averse mixed-integer multi-stage stochastic problems with dynamic objective functions defined via mean conditional value-at-risk (mean-CVaR). Mean-CVaR is a coherent measure of risk that has been used in the literature extensively. Coherent measures of risk and their axiomatic properties are introduced in the pioneering paper by Artzner et al. [1999]. Later, the theory

*Corresponding author: ozlem.cavus@bilkent.edu.tr

of coherent risk measures are extended by Rockafellar and Uryasev [2002], Ruszczynski and Shapiro [2006a,b], and references therein.

In a multi-stage decision horizon, risk involved in a stream of random outcomes is considered. Therefore, dynamic coherent risk measures are introduced to quantify the risk in multi-period models [see Ruszczynski and Shapiro [2006a,b], Artzner et al. [2007], Pflug and Römisch [2007], Kovacevic and Pflug [2009] and references therein].

For the multi-stage stochastic optimization problems with dynamic measures of risk, some exact solution techniques are suggested under the assumption that the decision variables are continuous. These techniques, such as stochastic dual dynamic programming (SDDP) (see [Pereira and Pinto, 1991, Shapiro, 2011, Shapiro et al., 2013, Philpott et al., 2013]) and Lagrangian relaxation of nonanticipativity constraints [Collado et al., 2012], rely on the convex structure of the problem, therefore, they cannot be used to find an exact solution when some of the decision variables are integer valued. On the other hand, these methods can be used to obtain lower bounds on the optimal value of multi-stage stochastic integer problems. Bonnans et al. [2012] propose an extension of SDDP method for the risk neutral problems with integer variables by relaxing the integrality requirements in the backward steps of the algorithm. Later, Bruno et al. [2016] extend this approach to risk-averse integer problems. Similarly, Schultz [2003] use Lagrangian relaxation of nonanticipativity constraints to obtain lower bounds within a branch-and-bound procedure for risk-neutral multi-stage problems with integer variables. However, these approaches rely on some restrictive assumptions. SDDP method requires stagewise independency of random process and the branch-and-bound procedure requires complete recourse assumptions. Therefore, they cannot be applicable to a wide range of problems.

A recent stream of research proposes an alternative way of obtaining bounds for mixed-integer multi-stage stochastic problems via a scenario tree decomposition. In that approach, the sample space is partitioned into subspaces called as groups, and the problem is solved for the scenarios in a group instead of the original sample space. These smaller problems are called as group subproblems. Sandıkçı et al. [2013] propose a group subproblem approach for risk-neutral mixed-integer two-stage stochastic problems. They show that the expected value of the optimal values of group subproblems gives a lower bound on the optimal value of the original problem. Later, this approach is extended to the risk-neutral multi-stage problems by Sandıkçı and Özaltın [2014], Maggioni et al. [2016] and Zenarosa et al. [2014]. Recently, Maggioni and Pflug [2016] apply group subproblem approach to risk-averse mixed-integer multi-stage stochastic problems where the objective is a concave utility function applied to the cumulative cost in each scenario.

In this study, we propose a scenario tree decomposition algorithm for risk-averse mixed-integer multi-stage stochastic problems with various dynamic objective functions defined via mean-CVaR. The proposed scenario tree decomposition algorithm is based on group subproblem approach. The algorithm is used to find lower and upper bounds on the optimal value of risk-averse mixed-integer multi-stage stochastic problems with different dynamic objective functions defined via mean-CVaR. We propose infinitely many valid lower bounds on mean-CVaR risk measure that can be used within the frame of the algorithm. We also investigate the effect of scenario partitioning strategies on the quality of the selected lower bound by considering different partitioning strategies based on the structure of the scenario tree and disparateness of scenario realizations.

As outlined earlier, our approach does not require any special structural property such as convexity and linearity of feasible set. Moreover, it does not require complete recourse or stagewise independence assumptions. Therefore, it can be applied to a wide range of problems. As an example, computational experiments are conducted on a multi-stage lot sizing problem with different choices of bounds and scenario tree partitions. The experiments reveal that the obtained bounds are tight and require reasonable CPU times. Another set of computational experiments reveal that our approach yields 2.58 times stronger bounds than solving the problem with IBM ILOG CPLEX. On the other hand, CPLEX requires more than 5.45 times of CPU time to obtain the same optimality gaps of our approach.

The presentation of the paper is as follows: In Section 2, we present problem definition and some preliminaries. Section 3 includes our main results on obtaining different lower bounds on mean-CVaR risk measure via a group subproblem approach. We consider the application of these bounds to a risk-averse mixed-integer multi-stage stochastic problem with different dynamic objectives defined via mean-CVaR. We also suggest a method to obtain an upper bound. The computational study conducted on a multi-stage lot sizing problem and related discussions are presented in Section 4. Section 5 is devoted to concluding remarks and future research directions.

2 Risk-averse Mixed-integer Multi-stage Stochastic Problems with Dynamic Mean-CVaR Objective

We consider a multi-stage discrete decision horizon where the decisions at stage $t \in \{1, \ldots, T\}$ are made based on the available information up to that stage. At first stage, the problem parameters are deterministic, and hence, first stage decision vector x_1 is deterministic as well. At stage $t \in \{2, \ldots, T\}$, some or all problem parameters are random. We use ξ_t to denote the random vector of problem parameters at stage t. The random decision vector x_t at stage t is a function of decisions in previous stages $x_1, \ldots, x_{t-1}$ and random parameters $\xi_2, \ldots, \xi_t$ of stages up to t.

Let Ω be a sample space of all possible scenarios. We assume that the sample space is finite. An element ω of Ω is called as a scenario. Realization of a scenario ω corresponds to a realization of random parameters in stages $2, \ldots, T$. Let $\xi_2(\omega), \ldots, \xi_T(\omega)$ be the realization of these random parameters under scenario $\omega \in \Omega$.

Our main interest is a risk-averse mixed-integer multi-stage stochastic problem with the objective of dynamic risk measure $\varrho_{1,T}(\cdot)$ over the horizon $1, \ldots, T$. The problem can be defined as:

$$\min_{x \in \mathcal{X}} \varrho_{1,T}(f_1(x_1), f_2(x_2, \xi_2), \ldots, f_T(x_T, \xi_T)), \tag{1}$$

where $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2(x_1, \xi_2) \times \ldots \times \mathcal{X}_T(x_{T-1}, \xi_T)$ is the abstract representation of possibly non-linear feasibility sets. $\mathcal{X}_1 \subseteq \mathbb{R}^{n_1} \times \mathbb{Z}^{m_1}$ is a mixed-integer deterministic set and $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \mathbb{Z}^{m_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t} \times \mathbb{Z}^{m_t}$ for $t = 2, \ldots, T$ are mixed-integer point-to-set mappings. The cost in the first stage is deterministic and given by a possibly non-linear function $f_1 : \mathbb{R}^{n_1} \times \mathbb{Z}^{m_1} \to \mathbb{R}$. The cost functions $f_t : \mathbb{R}^{n_t} \times \mathbb{Z}^{m_t} \times \Omega \to \mathbb{R}$, $t = 2, \ldots, T$ are random and may be non-linear.

Classical solution methods such as SDDP and Lagrangian relaxation of nonanticipativity constraints cannot be used to solve problem (1) due to integrality restrictions of some decision variables. Therefore, our focus is to obtain bounds on (1) where the objective function $\varrho_{1,T}(\cdot)$ is a dynamic risk measure defined via mean-CVaR risk measures.

Now, we present some necessary concepts and notation on coherent, conditional and dynamic risk measures to exploit the structure of (1) with mean-CVaR objective.

2.1 Coherent Measures of Risk

Let Ω be a sample space equipped with a sigma algebra $\mathscr{F}$. Also let $\mathcal{Z}$ be the space of all $\mathscr{F}$-measurable random variables with respect to sample space Ω and probability distribution P. $Z, W \in \mathcal{Z}$ represent uncertain outcomes for which lower realizations are preferable; i.e. they represent a random cost. Let Z_ω be the value that the random variable Z takes under scenario $\omega \in \Omega$. As defined in Artzner et al. [1999], a function $\rho : \mathcal{Z} \to \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ is called a coherent measure of risk if it satisfies:

(A1) Convexity: $\rho(\alpha Z + (1-\alpha)W) \leq \alpha\rho(Z) + (1-\alpha)\rho(W)$ for all $Z, W \in \mathcal{Z}$ and $\alpha \in [0,1]$,

(A2) Monotonicity: $Z \succeq W$ implies $\rho(Z) \geq \rho(W)$ for all $Z, W \in \mathcal{Z}$,

(A3) Translational Equivariance: $\rho(Z+t) = \rho(Z) + t$ for all $t \in \mathbb{R}$ and $Z \in \mathcal{Z}$,

(A4) Positive Homogeneity: $\rho(tZ) = t\rho(Z)$ for all $t > 0$ and $Z \in \mathcal{Z}$,

where $Z \succeq W$ indicates pointwise partial ordering such that $Z_\omega \geq W_\omega$ for a.e. $\omega \in \Omega$ and $\mathbb{R}$ is the set of real numbers.

Henceforth, we assume that cardinality of Ω is finite and $\mathscr{F}$ is the set of all events defined on Ω. Then, the probability of scenario $\omega \in \Omega$ can be specified as $p_\omega > 0$. In this case, elements of both $\mathcal{Z}$ and its dual space $\mathcal{Z}^*$ can be represented as elements of $\mathbb{R}^N$, that is $\mathcal{Z} = \mathcal{Z}^* = \mathbb{R}^N$ where $N = |\Omega|$.

Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ be the set of all possible scenarios. Also, let Z_ω and μ_ω be the values that $Z \in \mathcal{Z}$ and $\mu \in \mathcal{Z}^*$ takes under scenario $\omega \in \Omega$, respectively. We define the scalar product $\langle \cdot, \cdot \rangle$ as follows:

$$\langle \mu, Z \rangle := \sum_{i=1}^{N} p_{\omega_i} \mu_{\omega_i} Z_{\omega_i}.$$

Following fact is known as dual representation of coherent measure of risk (see for example Ruszczynski and Shapiro [2006b]): if $\rho(\cdot)$ is a coherent measure of risk, then for every random variable $Z \in \mathcal{Z}$,

$$\rho(Z) = \max_{\mu \in \mathcal{A}} \langle \mu, Z \rangle, \tag{2}$$

where

$$\mathcal{A} \subseteq \left\{ \mu \in \mathbb{R}^N : \langle \mu, \mathbf{1} \rangle = 1 \right\},$$

$\mathcal{A}$ is a compact and convex set and $\mathbf{1} \in \mathbb{R}^N$ with all entries being equal to one. Note that $\langle \mu, \mathbf{1} \rangle = \mathbb{E}[\mu]$ where the expectation is taken with respect to reference probability distribution P. Indeed, $\mathcal{A} = \partial\rho(0)$ where the right hand side denotes the subdifferential of $\rho(\cdot)$ at point zero. We call this set as the dual set of the risk measure $\rho(\cdot)$. A coherent measure of risk can be characterized via its dual set. The reader is referred to Ruszczynski and Shapiro [2006b] for a detailed discussion on the dual representation of coherent measures of risk.

2.2 Conditional and Dynamic Risk Measures

When a multi-stage stochastic process is considered, all realizations of the process form a scenario tree in the finite distribution case. In this section, we follow the notation used by Collado et al. [2012] to represent the scenario tree. Let Ω_t be the set of nodes at stage $t = 1, \ldots, T$. At stage $t = 1$, there is only one node, called as root node and it is represented by v_1. The nodes at stages $t = 2, \ldots, T$ represent elementary events in $\mathscr{F}_t$, that is $\mathscr{F}_t = \sigma(\Omega_t)$.

The set Ω_T corresponds to all possible scenarios, that is $\Omega_T = \Omega$. Each node $v \in \Omega_t, t = 2, \dots, T$ has a unique ancestor at stage $t-1$ and this ancestor node is called as $a(v)$. Also, each node $v \in \Omega_t, t = 1, \dots, T-1$ has a set of children nodes $C(v)$ such that $C(v) = \{u \in \Omega_{t+1} : a(u) = v\}$. The probability measure P can be specified by conditional probabilities

$$p_{vu} := P[u|v], \qquad v \in \Omega_t, u \in C(v), t = 1, \dots, T-1,$$

and probability of a scenario $\omega \in \Omega_T$ can be found as

$$p_\omega = p_{v_1 v_2} p_{v_2 v_3} \cdots p_{v_{t-1}\omega},$$

where $v_1, v_2, \dots, v_{t-1}, \omega$ is the unique path from root node v_1 to node ω.

The notation mentioned above is depicted in Figure 1 for a four-stage scenario tree.

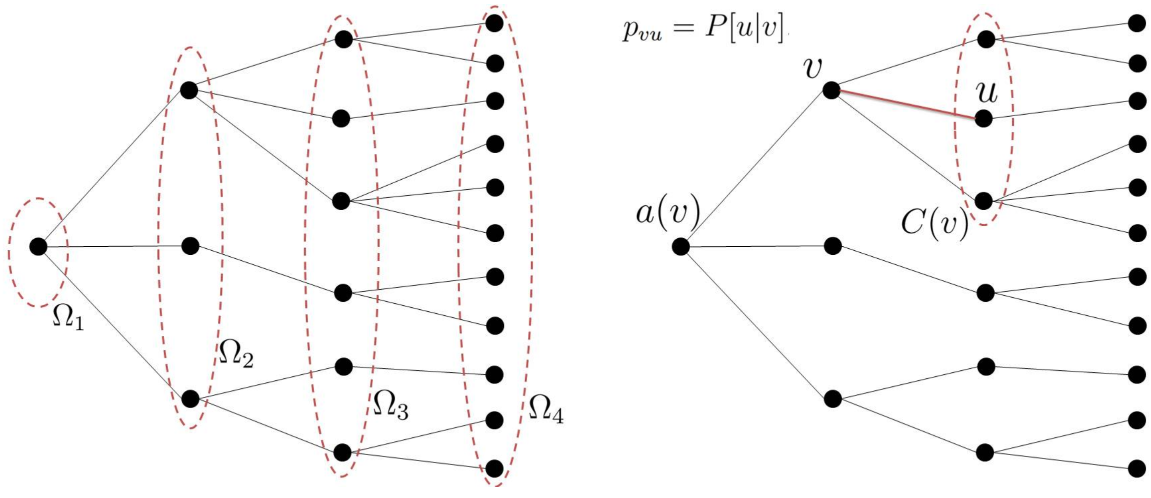


Figure 1: An example of four-stage scenario tree. (a) $\Omega_1, \Omega_2, \Omega_3$ and Ω_4 are the set of nodes at stages 1, 2, 3 and 4, respectively. (b) $C(v)$ is the set of children nodes of node v, $a(v)$ is the ancestor node of node v and p_{vu} is the conditional probability of node u given v.

In our multi-stage problem setting, we use a dynamic risk measure that can be written as a nested structure of one-step conditional risk measures. Before giving the definition of dynamic measure of risk, we first provide the definition of one-step conditional risk measure.

For a multi-stage decision horizon with stages $1, \dots, T$, consider a filtration $\{0, \emptyset\} = \mathscr{F}_1 \subset \mathscr{F}_2 \subset \dots \subset \mathscr{F}_T = \mathscr{F}$ that is an ordered set of sigma algebras representing gradually increasing information through stages. Let $\mathcal{Z}_t$ be the set of $\mathscr{F}_t$-measurable random variables for $t \in \{1, \dots, T\}$. The mapping $\rho_{\mathscr{F}_{t+1}|\mathscr{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ is called a one-step conditional risk measure if it satisfies axioms (A1)-(A4) for corresponding spaces $\mathcal{Z}_t$ and $\mathcal{Z}_{t+1}$ for all $t \in \{1, 2, \dots, T-1\}$.

The risk involved in a sequence of random variables $Z_t \in \mathcal{Z}_t, t \in \{1, \dots, T\}$ adopted to the filtration $\mathscr{F}_t, t \in \{1, \dots, T\}$ can be evaluated by a dynamic measure of risk $\varrho_{1,T}(\cdot)$, that is,

$$\varrho_{1,T}(Z_1, Z_2, \dots, Z_T) = Z_1 + \rho_{\mathscr{F}_2|\mathscr{F}_1}\left(Z_2 + \rho_{\mathscr{F}_3|\mathscr{F}_2}\left(Z_3 + \dots + \rho_{\mathscr{F}_T|\mathscr{F}_{T-1}}(Z_T)\dots\right)\right), \tag{3}$$

where $\rho_{\mathscr{F}_{t+1}|\mathscr{F}_t}(\cdot)$, $t \in \{1, 2, \ldots, T-1\}$ is a one-step conditional risk measure. The structure (3) is presented in Ruszczynski and Shapiro [2006a]. Later, Ruszczyński [2010] shows that the representation (3) can be constructed using monotonicity of conditional risk measures and the concept of time consistency.

Collado et al. [2012] show that the dual representation of coherent risk measures can be extended to dynamic measures of risk. If $\varrho_{1,T}(\cdot)$ is a dynamic risk measure given as in (3), then for every sequence of random variables $\{Z_t \in \mathcal{Z}_t\}_{t=1}^T$,

$$\varrho_{1,T}(Z_1, Z_2, \ldots, Z_T) = \max_{q_T \in \mathcal{Q}_T} \langle q_T, Z_1 + Z_2 + \cdots + Z_T \rangle, \tag{4}$$

where

$$Q_t = \mathcal{A}_{t-1} \circ \cdots \circ \mathcal{A}_2 \circ \mathcal{A}_1, \tag{5}$$

and $\mathcal{A}_t = \partial \rho_{\mathscr{F}_{t+1}|\mathscr{F}_t}(\mathbf{0})$, $t \in \{2, \ldots, T\}$. Here $\mathbf{0} \in \mathbb{R}^{|\Omega_{t+1}|}$, $t \in \{1, 2, \ldots, T-1\}$ is a vector of all zeros. Q_T is a compact and convex set. The operator "$\circ$" defines convolution of probability measures, that is,

$$(\mu_t \circ q_t)(u) = q_t(a(u))\mu_t(a(u), u), \forall u \in \Omega_{t+1},$$

and

$$\mathcal{A}_t \circ \mathcal{Q}_t = \{\mu_t \circ q_t : q_t \in Q_t, \mu_t \in \mathcal{A}_t\},$$

for all $t \in \{1, 2, \ldots, T-1\}$. Note that $a(u)$ is the ancestor node of u.

In this study, we use conditional mean-CVaR as one-step conditional risk measure. Therefore, the next section is devoted to the definition of mean-CVaR.

2.3 CVaR and Mean-CVaR

An important and extensively used example of coherent measures of risk is conditional value-at-Risk (CVaR). CVaR of $Z \in \mathcal{Z}$ at level $\alpha \in [0, 1)$ is defined as (see [Rockafellar and Uryasev, 2002])

$$CVaR_\alpha(Z) := \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{1-\alpha} \mathbb{E}[(Z - \eta)_+] \right\}, \tag{6}$$

where $(a)_+$ is positive part of $a \in \mathbb{R}$, that is, $(a)_+ := \max\{a, 0\}$. The infimum on the right hand side of (6) holds at $VaR_\alpha(Z)$ where $VaR_\alpha(Z) := \inf\{l \in \mathbb{R} : P(Z \leq l) \geq \alpha\}$.

Note that, $CVaR_0(Z) = \mathbb{E}[Z]$ and, $CVaR_\alpha(Z)$ converges to the value of Z in the worst case scenario, i.e. $\max(Z) = \max_{i \in \{1, \ldots, N\}} Z_{\omega_i}$, as $\alpha \uparrow 1$. If P is an atomless distribution, CVaR can also be expressed as:

$$CVaR_\alpha(Z) = \mathbb{E}[Z | Z \geq VaR_\alpha(Z)].$$

In this study, we focus on mean-CVaR, which is a coherent measure of risk. Despite CVaR, mean-CVaR risk measure conveys the expected value information of a random variable, as well.

Given a weight parameter $\epsilon_1 \in [0, 1]$ and a level parameter $\alpha \in [0, 1)$, mean-CVaR of $Z \in \mathcal{Z}$ is defined as

$$\rho(Z) := (1 - \epsilon_1)\mathbb{E}[Z] + \epsilon_1 CVaR_\alpha(Z). \tag{7}$$

As seen in (7), mean-CVaR is a convex combination of expected value of a given random variable Z and CVaR value of this random variable at level α. As ϵ_1 or α increase, $\rho(\cdot)$ gets more risk averse. If

$\epsilon_1 = 0$, then $\rho(Z) = \mathbb{E}[Z]$, similarly $\rho(Z) = CVaR_\alpha(Z)$ when $\epsilon_1 = 1$.

The expression in (7) can equivalently be represented as following linear program for finite probability spaces.

$$\begin{aligned} \rho(Z) = \text{minimize} \quad & (1-\epsilon_1)\sum_{\omega\in\Omega} p_\omega Z_\omega + \epsilon_1 \left(\eta + \frac{1}{1-\alpha}\sum_{\omega\in\Omega} p_\omega \vartheta_\omega\right), \\ \text{subject to} \quad & \vartheta_\omega \geq Z_\omega - \eta, \quad \forall\omega\in\Omega, \\ & \eta \text{ urs}, \vartheta_\omega \geq 0, \quad \forall \omega\in\Omega. \end{aligned}$$

When the sample space is finite, the dual representation (2) holds for mean-CVaR with the set $\mathcal{A}$ represented as (see [Ruszczynski and Shapiro, 2006b]):

$$\mathcal{A} = \left\{\mu \in \mathbb{R}^N : 1-\epsilon_1 \leq \mu_\omega \leq 1+\epsilon_2, \forall\omega\in\Omega \text{ and } \mathbb{E}[\mu] = 1\right\}, \tag{8}$$

where

$$\epsilon_2 := \frac{\alpha}{1-\alpha}\epsilon_1 \geq 0.$$

Similarly, for any $Z \in \mathcal{Z}_{t+1}$, the one-step conditional mean-CVaR risk measure $\rho_{\mathscr{F}_{t+1}|\mathscr{F}_t}(Z)$ with parameters $\epsilon_1^t \in [0,1]$ and $\alpha^t \in [0,1)$ is defined as:

$$\rho_{\mathscr{F}_{t+1}|\mathscr{F}_t}(Z) := (1-\epsilon_1^t)\mathbb{E}[Z|\mathscr{F}_t] + \epsilon_1^t \inf_{\eta\in\mathcal{Z}_t}\left\{\eta + \frac{1}{1-\alpha^t}\mathbb{E}[(Z-\eta)_+|\mathscr{F}_t]\right\}. \tag{9}$$

Its dual set $\mathcal{A}_t$, $t \in \{1,2,\ldots,T-1\}$ is

$$\mathcal{A}_t = \left\{\mu^t \in \mathbb{R}^{|\Omega_{t+1}|} : 1-\epsilon_1^t \leq \mu_\omega^t \leq 1+\epsilon_2^t, \forall\omega\in\Omega_{t+1} \text{ and } \mathbb{E}[\mu^t|\mathscr{F}_t] = \mathbf{1}\right\}, \tag{10}$$

where $\epsilon_2^t = (\alpha^t/(1-\alpha^t))\epsilon_1^t$ and $\mathbf{1} \in \mathbb{R}^{|\Omega_t|}$.

Due to the one-to-one correspondence between a risk measure and its dual set, a mean-CVaR risk measure can either be defined as in (7) with parameters ϵ_1 and α, or via its dual as in (8) with parameters ϵ_1 and ϵ_2.

3 Bounds

The main motivation of this section is to propose lower and upper bounds for problem (1) with an objective of dynamic mean-CVaR. Therefore, using scenario groups, we first propose a continuum of lower bounds for mean-CVaR risk measure. Some possible lower bounds are presented in Section 3.2. The application of these bounds to a risk-averse mixed-integer multi-stage stochastic problems with an objective of (3) is presented in Section 3.3. Extension of the proposed lower bounds to other dynamic mean-CVaR risk measures is discussed in Section 3.4. In Section 3.5, we propose a method for obtaining an upper bound to the problem. The proposed algorithm benefits these results and yields lower and upper bounds for the problem.

3.1 Lower Bounds for Mean-CVaR Risk Measure

Let $\rho(\cdot)$ and $\widetilde{\rho}(\cdot)$ be two coherent measures of risk with dual sets $\mathcal{A}$ and $\widetilde{\mathcal{A}}$, respectively. In Proposition 1, we derive the necessary condition that $\widetilde{\rho}(\cdot)$ gives a lower bound for $\rho(\cdot)$.

Proposition 1: $\widetilde{\rho}(Z) \leq \rho(Z)$ for all $Z \in \mathcal{Z}$ if $\widetilde{\mathcal{A}} \subseteq \mathcal{A}$.

Proof: For any $Z \in \mathcal{Z}$, let $\mu^* \in \widetilde{\mathcal{A}}$ such that maximization in equation (2) is attained at μ^* for $\widetilde{\rho}(Z)$, that is, $\widetilde{\rho}(Z) = \langle \mu^*, Z \rangle$. If $\widetilde{\mathcal{A}} \subseteq \mathcal{A}$, then $\mu^* \in \mathcal{A}$ and $\langle \mu^*, Z \rangle \leq \max_{\mu \in \mathcal{A}} \langle \mu, Z \rangle = \rho(Z)$. Since Z is arbitrary, desired inequality follows. □

Although a similar version of Proposition 1 is presented in Iancu et al. [2015], our purpose is to derive lower bounds for risk averse problems with an objective of dynamic mean-CVaR risk measure by using Proposition 1. Hence, we try to construct a risk measure $\widetilde{\rho}(\cdot)$, or equivalently its dual set $\widetilde{\mathcal{A}}$, in such a way that both computation of lower bound is easy and obtained lower bound is tight.

The risk measure $\widetilde{\rho}(\cdot)$, or equivalently its dual set $\widetilde{\mathcal{A}}$, can be constructed in different ways. When the cardinality of the sample space is large, due to computational concerns, one may think of dealing with subsets of sample space separately and then obtain a lower bound information for $\rho(\cdot)$. For such construction, we need the definition of scenario groups and partition. A subset of scenarios $S \subseteq \Omega$ is called as a group. Let $\mathcal{S} = \{S_j\}_{j=1}^{J}$ be a collection of groups that forms a partition of Ω, that is, $\bigcup_{j=1}^{J} S_j = \Omega$ and $S_j \bigcap Sj' = \emptyset$ for all $j, j' \in \{1, 2, \ldots, J\}$ such that $j \neq j'$. Note that the groups may not be necessarily disjoint (see [Sandıkçı and Özaltın, 2014]), i.e. $S_j \bigcap Sj' \neq \emptyset$, but for the ease of representation, we partition the sample space into disjoint groups. Let $\mathscr{G}$ be a $\sigma-$algebra generated by partition $\mathcal{S}$ where each group $S_j \in \mathcal{S}$ corresponds to an elementary event of $\mathscr{G}$. The probability of an elementary event corresponding to S_j is $p_j = \sum_{\omega \in S_j} p_\omega$ which is the total probability of scenarios in S_j. We also define the adjusted probability of each scenario ω as $p_{j\omega} = p_\omega / p_j$ for all $\omega \in S_j$ and $j \in \{1, 2, \ldots, J\}$. Note that, $\mathscr{G}$ is a sub $\sigma-$algebra of $\mathscr{F}$.

Once a partition of sample space is given, one way to construct $\widetilde{\rho}(\cdot)$ is to define it as a convolution of a coherent risk measure $\rho_{\mathscr{G}} : \mathscr{G} \to \mathbb{R}$ and a one-step conditional risk measure $\rho_{\mathscr{F}|\mathscr{G}} : \mathscr{F} \to \mathscr{G}$, that is, $\widetilde{\rho}(\cdot) = (\rho_{\mathscr{G}} \circ \rho_{\mathscr{F}|\mathscr{G}})(\cdot)$, or equivalently, define its dual set as a convolution of sets $\mathcal{A}_{\mathscr{G}}$ and $\mathcal{A}_{\mathscr{F}|\mathscr{G}}$ such that $\widetilde{\mathcal{A}} = \mathcal{A}_{\mathscr{F}|\mathscr{G}} \circ \mathcal{A}_{\mathscr{G}}$.

The one-step conditional risk measure $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ is defined by conditioning to each elementary event of $\mathscr{G}$. Let $\rho_{S_j} : \sigma(S_j) \to \mathbb{R}$ be a coherent risk measure for the elementary event j of $\mathscr{G}$. Then, $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ can be represented in terms of $\rho_{S_j}(\cdot), j \in \{1, 2 \ldots, J\}$, that is, $\left[\rho_{\mathscr{F}|\mathscr{G}}(\cdot)\right]_j = \rho_{S_j}(\cdot)$. Figure 2 depicts aforementioned notation for a given partition of scenario tree with five scenarios.

For the remainder of the paper, we will focus on mean-CVaR risk measure. Hence, we will use $\rho(\cdot)$ to refer a mean-CVaR risk measure as in (7) and $\rho_{\mathscr{F}_{t+1}|\mathscr{F}_t}(\cdot)$, $t \in \{1, 2 \ldots, T-1\}$ to refer a one-step conditional mean-CVaR risk measure as in (9).

For mean-CVaR case, $\widetilde{\rho}(\cdot)$ or equivalently its dual set $\widetilde{\mathcal{A}}$, can be explicitly stated. Let parameters of $\rho_{\mathscr{G}}$ be $\epsilon_1^1 \in [0, 1]$ and $\epsilon_2^1 \geq 0$, and parameters of $\rho_{\mathscr{F}|\mathscr{G}}$ be $\epsilon_1^2 \in [0, 1]$ and $\epsilon_2^2 \geq 0$. Consider the convolution $\widetilde{\rho} = \rho_{\mathscr{G}} \circ \rho_{\mathscr{F}|\mathscr{G}} : \mathscr{F} \to \mathbb{R}$ and its dual set

$$\begin{aligned}\widetilde{\mathcal{A}} = \mathcal{A}_{\mathscr{F}|\mathscr{G}} \circ \mathcal{A}_{\mathscr{G}} &= \{\mu \in \mathbb{R}^N : \mu = \mu^1 \circ \mu^2, \mu_1 \in \mathcal{A}_{\mathscr{G}}, \mu_2 \in \mathcal{A}_{\mathscr{F}|\mathscr{G}}\} \\ &= \{\mu \in \mathbb{R}^N : \mu = \mu^1 \circ \mu^2, 1 - \epsilon_1^1 \leq \mu_j^1 \leq 1 + \epsilon_2^1, \forall j \in 1, 2 \ldots, J \text{ and } \mathbb{E}[\mu^1] = 1, \\ &\qquad 1 - \epsilon_1^2 \leq \mu_\omega^2 \leq 1 + \epsilon_2^2, \forall \omega \in \Omega \text{ and } \mathbb{E}[\mu^2 | \mathscr{G}] = \mathbf{1}\}, \quad (11)\end{aligned}$$

where $\mathbf{1} \in \mathbb{R}^J$. Construction of the set $\widetilde{\mathcal{A}}$ for the example in Figure 2 can be seen in Appendix A.

Now, we are ready to prove that a lower bound for mean-CVaR risk measure $\rho(\cdot)$ can be obtained by convolution of $\mathcal{A}_{\mathscr{G}}(\cdot)$ and $\mathcal{A}_{\mathscr{F}|\mathscr{G}}(\cdot)$.

Proposition 2: Let $\rho(\cdot)$ be mean-CVaR risk measure with dual set $\mathcal{A}$ and $\widetilde{\rho}(\cdot)$ be defined with

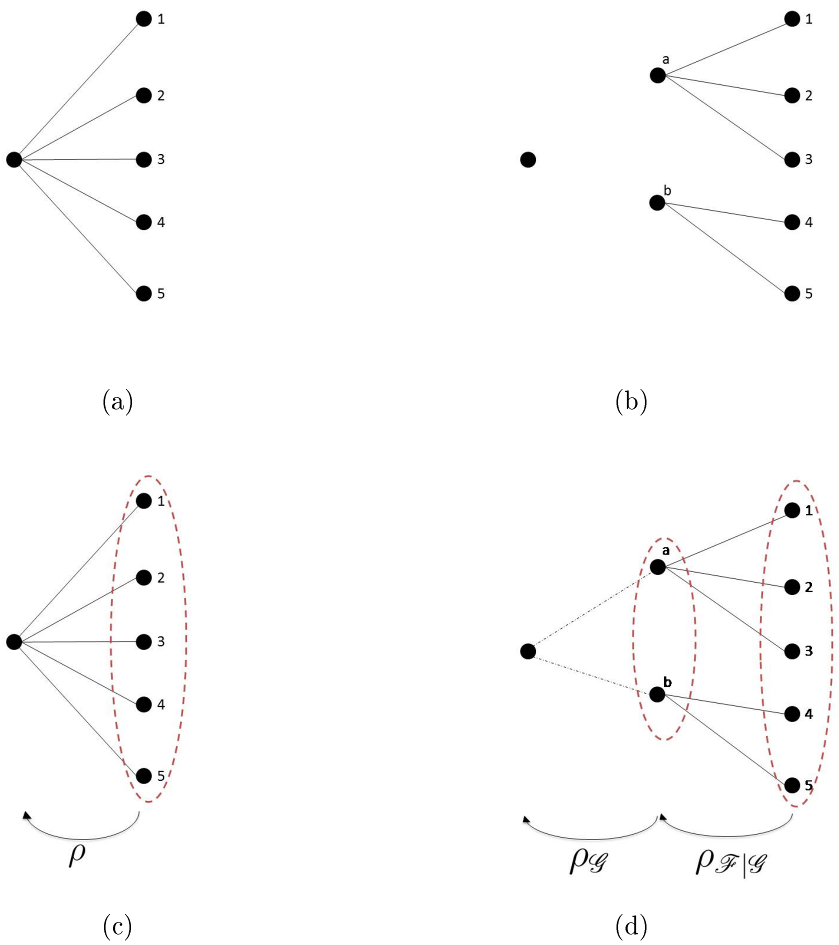


Figure 2: (a) An example partition for a two-stage scenario tree: There are five scenarios $1, 2, 3, 4, 5$ with probabilities p_1, p_2, p_3, p_4, p_5, respectively. (b) $\mathcal{S} = \{S_a, S_b\}$ is a partition of Ω where $S_a = \{1, 2, 3\}$ and $S_b = \{4, 5\}$. Nodes a and b correspond to groups S_a and S_b with probabilities $p_a = p_1 + p_2 + p_3$ and $p_b = p_4 + p_5$, respectively. (c) $\rho : \mathscr{F} \to \mathbb{R}$ is the original risk measure where $\mathscr{F} = \sigma(1, 2, 3, 4, 5)$. (d) $\mathscr{G} = \sigma(a, b)$ is a sub $\sigma-$algebra of $\mathscr{F}$. $\rho_{\mathscr{G}} : \mathscr{G} \to \mathbb{R}$ is a coherent risk measure and $\rho_{\mathscr{F}|\mathscr{G}} : \mathscr{F} \to \mathscr{G}$ is a one-step conditional risk measure that can be represented via $\rho_{S_a} : \sigma(S_a) \to \mathbb{R}$ and $\rho_{S_b} : \sigma(S_b) \to \mathbb{R}$ as $[\rho_{\mathscr{F}|\mathscr{G}}(\cdot)]_a = \rho_{S_a}(\cdot)$ and $[\rho_{\mathscr{F}|\mathscr{G}}(\cdot)]_b = \rho_{S_b}(\cdot)$.

dual set $\widetilde{\mathcal{A}}$ as in (11). Then, $\widetilde{\rho}(Z) \leq \rho(Z)$ for all $Z \in \mathcal{Z}$ if

$$1 - \epsilon_1 \leq (1 - \epsilon_1^1)(1 - \epsilon_1^2) \text{ and } (1 + \epsilon_2^1)(1 + \epsilon_2^2) \leq 1 + \epsilon_2. \tag{12}$$

Proof: Let $\mu \in \widetilde{\mathcal{A}}$. Then there exist $\mu^1 \in \mathcal{A}_{\mathscr{G}}$ and $\mu^2 \in \mathcal{A}_{\mathscr{F}|\mathscr{G}}$ such that $\mu = \mu^1 \circ \mu^2$ with $\mathbb{E}[\mu^1] = 1$ and $\mathbb{E}[\mu^2|\mathscr{G}] = \mathbf{1}$. Properties of conditional expectation implies that $\mathbb{E}[\mu] = \mathbb{E}\left[\mathbb{E}\left[\mu|\mathscr{G}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\mu^1 \circ \mu^2|\mathscr{G}\right]\right] = \mathbb{E}\left[\mu^1 \circ \mathbb{E}\left[\mu^2|\mathscr{G}\right]\right] = \mathbb{E}\left[\mu^1 \circ \mathbf{1}\right] = \mathbb{E}[\mu^1] = 1$.

Moreover, $(1-\epsilon_1^1)(1-\epsilon_1^2) \leq \mu_\omega \leq (1+\epsilon_2^1)(1+\epsilon_2^2)$ for all $\omega \in \Omega$ by (11). If $1 - \epsilon_1 \leq (1-\epsilon_1^1)(1-\epsilon_1^2)$ and $(1+\epsilon_2^1)(1+\epsilon_2^2) \leq 1 + \epsilon_2$, then $1 - \epsilon_1 \leq \mu_\omega \leq 1 + \epsilon_2$, for all $\omega \in \Omega$ which implies, $\mu \in \mathcal{A}$. Since μ is arbitrary, $\widetilde{\mathcal{A}} \subseteq \mathcal{A}$.

Then, by Proposition 1, $\widetilde{\rho}(Z) \leq \rho(Z)$ for all $Z \in \mathcal{Z}$. □

Proposition 2 implies that, under conditions (12), $\widetilde{\rho}(Z) = (\rho_{\mathscr{G}} \circ \rho_{\mathscr{F}|\mathscr{G}})(Z)$ gives a valid lower bound for $\rho(Z)$ for any given partition $\mathcal{S}$ of Ω. If $\rho(\cdot)$ is a conditional mean-CVaR risk measure, Proposition 2 still applies. In this case, the expectations in the proof are replaced with corresponding conditional expectations.

3.2 Possible Lower Bounds

We have shown that a lower bound for $\rho(\cdot)$ can be obtained by convolutions of mean-CVaR risk measures whose parameters satisfy condition (12). Due to Proposition 2, we can generate infinitely many lower bounds. Table 1 presents some special cases of parameters of $\rho_{\mathscr{G}}(\cdot)$ and $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ such that they can be used to obtain a lower bound for $\rho(\cdot)$.

Insert Table 1 here

Bounds 1 and 3 represent the extreme cases where either $\rho_{\mathscr{G}}(\cdot)$ or $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ is the expectation operator. Bound 2 is an intermediate case between Bounds 1 and 3 where both $\rho_{\mathscr{G}}(\cdot)$ and $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ have the same parameters (both represented as $\rho^s(\cdot)$ due to symmetry between them), that is, $\epsilon_1^1 = \epsilon_1^2$, $\epsilon_2^1 = \epsilon_2^2$ and $\alpha^1 = \alpha^2$, and inequalities in (12) hold at equality. Note that risk-aversion of $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ increases and risk-aversion of $\rho_{\mathscr{G}}(\cdot)$ decreases, moving through Bound 1 to 3.

An interesting question is whether one of the possible lower bounds presented above is always preferable among others. Following example reveals that Bounds 1 and 3 are incomparable and Bound 2 is not necessarily the tightest bound among others.

Example 1: Consider a random variable Z with sample space $\Omega = \{\omega_i\}_{i=1}^4$. All four realizations have equal probabilities, that is, $p_{\omega_i} = 1/4$ for all $i \in \{1,2,3,4\}$. The value that Z takes under realization ω_i is $Z_{\omega_i} = i$ for $i \in \{1,2,3,4\}$.

Let $\epsilon_1 = 1$ and $\alpha = 0.5$, then (7) reduces to CVaR value at $\alpha = 0.5$ and then $\rho(Z) = 3.5$.

Two different partition of scenarios are $\mathcal{S} = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$ and $\mathcal{S}' = \{\{\omega_1, \omega_4\}, \{\omega_2, \omega_3\}\}$. Values of Bounds 1, 2, and 3 for partitions $\mathcal{S}$ and $\mathcal{S}'$ are given in Table 2.

Insert Table 2 here

As seen in Table 2, the tightest bounds for partitions $\mathcal{S}$ and $\mathcal{S}'$ are Bounds 1 and 3, respectively. Hence, Bounds 1 and 3 are incomparable. Another observation is the fact that Bound 2 is not necessarily the tightest bound among others.

Although Shapiro et al. [2009] show that Bound 1 is a valid lower bound for any coherent risk measures, the other bounds may not be applicable for all coherent risk measures. The risk measures ρ^s in the definition of Bound 2 cannot be easily represented for general coherent risk measures and Example 2 reveals that Bound 3 is not necessarily a valid lower bound for an arbitrary coherent risk measure.

Example 2: Consider a random variable Z that takes values $Z_{\omega_1} = 100$, $Z_{\omega_2} = 0$, $Z_{\omega_3} = 1$ and $Z_{\omega_4} = 500$ with probabilities 0.3, 0.2, 0.4 and 0.1, respectively. We use the first order mean semi-deviation as a risk measure, that is:

$$\rho(Z) = \mathbb{E}[Z] + \kappa \mathbb{E}\left[(Z - \mathbb{E}[Z])_+\right], \quad \kappa \in [0,1].$$

Let $\kappa = 0.5$. For partition $\mathcal{S} = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$, $\rho(Z) = 104.32$ but $(\mathbb{E} \circ \rho)(Z) = 106.36$. Therefore, Bound 3 is not necessarily a valid lower bound for all coherent risk measures.

3.3 Lower Bound for Optimization Problem

In this section, we extend the lower bound proposed in Proposition 2 to the risk-averse mixed-integer multi-stage stochastic problems with an objective of dynamic mean-CVaR risk measures. Using the structure presented in (3), the problem (1) can be written as

$$\textbf{(P)} \quad \min_{x_1 \in \mathcal{X}_1} \quad f_1(x_1) + \rho(Q(x_1, \xi)), \tag{13}$$

where

$$Q(x_1, \xi) = \min_{x_t \in \mathcal{X}_t, t=2,\ldots,T} \varrho_{2,T} \left(f_2(x_2, \xi_2), \ldots, f_T(x_T, \xi_T) \right), \tag{14}$$

$\xi = \{\xi_t\}_{t=2}^T$ and $\rho(\cdot)$ is a mean-CVaR risk measure with parameters $\epsilon_1 \in [0,1]$ and $\epsilon_2 \geq 0$. Let x_1^* and z^* be an optimal first-stage solution and optimal value of **(P)**, respectively.

Recall the partition $\mathcal{S} = \{S_j\}_{j=1}^J$ of Ω and sigma algebra $\mathscr{G}$ induced by this partition. Then, the j^{th} group subproblem is just problem **(P)** with sample space S_j with adjusted probabilities $p_{j\omega}, \omega \in S_j$. Additionally, the risk measure $\rho(\cdot)$ in (13) is replaced by $\rho_{S_j}(\cdot)$. For $j \in \{1, 2, \ldots, J\}$, let z^j be the optimal value of j^{th} group subproblem. Also let Z_{LB} be a $\mathscr{G}$-measurable random variable that takes value of z^j with probability $p_j = \sum_{\omega \in S_j} p_\omega$.

In Theorem 1, we show that a lower bound for risk-averse mixed-integer multi-stage stochastic problem **(P)** can be obtained by using optimal values of group subproblems.

Theorem 1: Let $\rho_{\mathscr{G}} : \mathscr{G} \to \mathbb{R}$ be a mean-CVaR risk measure with parameters $\epsilon_1^1 \in [0,1]$ and $\epsilon_2^1 \geq 0$ and $\rho_{\mathscr{F}|\mathscr{G}} : \mathscr{F} \to \mathscr{G}$ be a conditional mean-CVaR risk measure with parameters $\epsilon_1^2 \in [0,1]$ and $\epsilon_2^2 \geq 0$ satisfying $1 - \epsilon_1 \leq (1-\epsilon_1^1)(1-\epsilon_1^2)$ and $(1+\epsilon_2^1)(1+\epsilon_2^2) \leq 1 + \epsilon_2$. Then, $z^* \geq \rho_{\mathscr{G}}(Z_{LB})$.

Proof: Since x_1^* is an optimal first stage solution of **(P)**, then it is a feasible first stage solution for each group subproblem. By optimality of each group subproblem, we have

$$c_1 x_1^* + \rho_{S_j}(Q(x_1^*, \xi)) \geq z^j, \quad \forall j \in \{1, \ldots, J\}$$

and

$$c_1 x_1^* + \rho_{\mathscr{F}|\mathscr{G}}(Q(x_1^*, \xi)) \succeq Z_{LB}. \tag{15}$$

The values on the both sides of inequality (15) is $\mathscr{G}-$measurable. Since, $\rho_{\mathscr{G}}(\cdot)$ is a coherent risk measure and it satisfies monotonicity axiom (A2), we get

$$\rho_{\mathscr{G}}(c_1 x_1^* + \rho_{\mathscr{F}|\mathscr{G}}(Q(x_1^*, \xi))) \geq \rho_{\mathscr{G}}(Z_{LB}). \tag{16}$$

Note that, $c_1 x_1^*$ is an $\mathscr{F}-$measurable cost. Since $\mathscr{G}$ is a sub $\sigma-$algebra of $\mathscr{F}$, $c_1 x_1^*$ is $\mathscr{G}-$measurable, as well. Applying translational equivariance axiom (A3) to the left hand side of (16), we get

$$\rho_{\mathscr{G}}(\rho_{\mathscr{F}|\mathscr{G}}(c_1 x_1^* + Q(x_1^*, \xi))) \geq \rho_{\mathscr{G}}(Z_{LB}). \tag{17}$$

Since conditions in (12) are satisfied, we can apply Proposition 2 to the left hand side of inequality (17) and obtain:

$$\rho(c_1 x_1^* + Q(x_1^*, \xi)) \geq \rho_{\mathscr{G}}(Z_{LB}).$$

Finally, using translational equivariance axiom (A3), we get

$$z^* = c_1 x_1^* + \rho(Q(x_1^*, \xi)) \geq \rho_{\mathscr{G}}(Z_{LB}). \quad \square$$

Theorem 1 implies that a lower bound on the optimal value of **(P)** can be obtained by solving group subproblems and then applying $\rho_{\mathscr{G}}(\cdot)$ to the optimal values of these group subproblems. Since group subproblems include smaller number of scenarios compared to the original problem, they are computationally less challenging. Moreover, applying $\rho_{\mathscr{G}}(\cdot)$ to the optimal values of group subproblems requires negligible computational effort, since it is only the calculation of value of a risk measure $\rho_{\mathscr{G}}(\cdot)$ for a given random cost.

Although the dynamic risk measure widely used in the literature is the nested structure presented in (3), which is also the focus of this paper, there are other risk measures that can be used to evaluate the risk of a sequence of random variables. We show that our approach can also be applied to the risk-averse mixed-integer multi-stage stochastic problems with different dynamic extensions of mean-CVaR risk measure.

3.4 Extension to Other Dynamic Measures of Risk

Some examples of dynamic risk measures apart from the nested structure in (3) are multiperiod mean-CVaR and sum of mean-CVaR (see Pflug and Römisch [2007], Eichhorn and Römisch [2005], respectively). For a sequence of random variables $Z_t \in \mathcal{Z}_t, t \in \{1, \dots, T\}$ adopted to the filtration $\mathscr{F}_t, t \in \{1, \dots, T\}$, multiperiod mean-CVaR is defined as

$$\rho^m(\{Z_t\}_{t=2}^T) = \sum_{t=2}^{T} \mu_t \mathbb{E}[\rho_{\mathscr{F}_t|\mathscr{F}_{t-1}}(Z_t)], \tag{18}$$

and sum of mean-CVaR is represented as

$$\rho^s(\{Z_t\}_{t=2}^T) = \sum_{t=2}^{T} \mu_t \rho_t(Z_t), \tag{19}$$

with $\sum_{t=2}^T \mu_t = 1, \mu_t \geq 0$ for $t \in \{2, 3, \dots, T\}$.

The multiperiod mean-CVaR risk measure (18) is a time consistent dynamic measure of risk whereas the sum of mean-CVaR (19) is not time consistent. For the definition and discussion on the concept of time consistency, see Kovacevic and Pflug [2009], Shapiro [2009] and Homem-de Mello and Pagnoncelli [2016].

Our approach is also applicable for the case where the risk measure is applied to whole scenario cost as a time inconsistent objective function, that is,

$$\rho^{nd}(\{Z_t\}_{t=1}^T) = \rho(Z_1 + Z_2 + \ldots + Z_T). \tag{20}$$

Although the risk measure (20) can be applied to a sequence of random variables, it is not a dynamic measure of risk.

In the following three propositions, we show that a lower bound for (18), (19) or (20) risk measures can be obtained by scenario grouping. Therefore, our approach is still valid for Problem **(P)** with an objective of one these risk measures.

Consider an arbitrary sequence of random variables $Z_t \in \mathcal{Z}_t, t \in \{1, \dots, T\}$ adopted to the filtration $\mathscr{F}_t, t \in \{1, \dots, T\}$.

Proposition 3: For a multiperiod mean-CVaR risk measure $\rho^m(\cdot)$ as defined in (18), $\mathbb{E} \circ \rho^m(\cdot)$ is a valid lower bound.

Proof: If multiperiod mean-CVaR risk measure (18) is applied to the sequence $Z_t \in \mathcal{Z}_t, t \in \{1, \ldots, T\}$, then

$$\rho^m(\{Z_t\}_{t=2}^T) = \sum_{t=2}^T \mu_t \mathbb{E}\left[\rho_{\mathscr{F}_t|\mathscr{F}_{t-1}}(Z_t)\right].$$

Since $\rho_{\mathscr{F}_t|\mathscr{F}_{t-1}}(\cdot)$ is a conditional mean-CVaR risk measure, the lower bound $\mathbb{E} \circ \rho_{\mathscr{F}_t|\mathscr{F}_{t-1}}(\cdot)$ applies for $t \in \{2, 3, \ldots, T\}$. Then,

$$\rho^m(\{Z_t\}_{t=2}^T) \geq \sum_{t=2}^T \mu_t \mathbb{E}\left[\mathbb{E}\left[\rho_{\mathscr{F}_t|\mathscr{F}_{t-1}}(Z_t)\right]\right].$$

To avoid notational ambiguity, expectation operators are given without reference sigma algebras. Since expectation is a linear operator, we get

$$\rho^m(\{Z_t\}_{t=2}^T) \geq \mathbb{E}\left[\sum_{t=2}^T \mu_t \mathbb{E}[\rho_{\mathscr{F}_t|\mathscr{F}_{t-1}}(Z_t)]\right],$$

or equivalently,

$$\rho^m(\{Z_t\}_{t=2}^T) \geq \mathbb{E}\left[\rho^m(\{Z_t\}_{t=2}^T)\right].$$

Since the sequence $Z_t \in \mathcal{Z}_t, t \in \{1, \ldots, T\}$ is arbitrary, the desired result follows. □

Proposition 4: For a sum of mean-CVaR risk measure $\rho^s(\cdot)$ as defined in (19), $\mathbb{E} \circ \rho^m(\cdot)$ is a valid lower bound.

Proof: If sum of mean-CVaR risk measure (19) is applied to the sequence $Z_t \in \mathcal{Z}_t, t \in \{1, \ldots, T\}$, then

$$\rho^s(\{Z_t\}_{t=2}^T) = \sum_{t=2}^T \mu_t \rho_t(Z_t).$$

Similarly, $\mathbb{E} \circ \rho_t(\cdot)$ applies for $t \in \{2, 3, \ldots, T\}$. Then,

$$\rho^s(\{Z_t\}_{t=2}^T) \geq \sum_{t=2}^T \mu_t \mathbb{E}\left[\rho_t(Z_t)\right],$$

and

$$\rho^s(\{Z_t\}_{t=2}^T) \geq \mathbb{E}\left[\sum_{t=2}^T \mu_t \rho_t(Z_t)\right],$$

or equivalently,

$$\rho^s(\{Z_t\}_{t=2}^T) \geq \mathbb{E}\left[\rho^s(\{Z_t\}_{t=2}^T)\right].$$

Since the sequence $Z_t \in \mathcal{Z}_t, t \in \{1, \ldots, T\}$ is arbitrary, the desired result follows. □

Proposition 5: For the risk measure $\rho^{nd}(\cdot)$ as defined in (20), $\rho_{\mathscr{G}} \circ \rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ is a valid lower bound if parameters of $\rho_{\mathscr{G}}(\cdot)$ and $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ satisfy conditions in (12).

Proof: If the mean-CVaR risk measure (20) is applied to the sequence $Z_t \in \mathcal{Z}_t, t \in \{1, \ldots, T\}$, then

$$\rho^{nd}(\{Z_t\}_{t=1}^T) = \rho(Z_1 + Z_2 + \ldots + Z_T).$$

Since $\rho_{\mathscr{G}}(\cdot)$ and $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ satisfy conditions in (12), their convolution is a valid lower bound on mean-CVaR risk measure $\rho(\cdot)$, that is,

$$\rho^{nd}(\{Z_t\}_{t=1}^T) \geq \rho_{\mathscr{G}} \circ \rho_{\mathscr{F}|\mathscr{G}}(Z_1 + Z_2 + \ldots + Z_T),$$

or equivalently,

$$\rho^{nd}(\{Z_t\}_{t=1}^T) = \rho_{\mathscr{G}} \circ \rho_{\mathscr{F}|\mathscr{G}}(\{Z_t\}_{t=1}^T).$$

Since the sequence $Z_t \in \mathcal{Z}_t, t \in \{1, \dots, T\}$ is arbitrary, the desired result follows. □

As shown above, our proposed lower bound is quite general and can be applied to other dynamic mean-CVaR measures.

3.5 Upper Bound for Optimization Problem

Obtaining an upper bound, or equivalently finding a feasible solution of a minimization problem, is crucial for the instances where an optimal solution is not available. A good quality feasible solution gives the decision maker an action to be taken and measures the quality of obtained lower bound when an optimal solution is not available.

An upper bound for the optimal value of **(P)** can be obtained by using optimal solutions of group subproblems. Once j^{th} group subproblem is solved, an optimal solution of it, namely x^j, is obtained. Let UB_j be the optimal value of **(P)** where (some of) the variables appearing in j^{th} group subproblem are set to x^j. We call this problem as restricted problem. Since some of the problem variables are fixed, solving the restricted problem is easier than the original one and the resulting scenario tree can become decomposable.

If x^j is not feasible for original problem **(P)**, then corresponding upper bound UB_j is set to infinity. The best available upper bound UB is obtained by taking minimum of UB_j values over all $j \in \{1, \dots, J\}$, that is,

$$UB = \min_{j \in \{1,\dots,J\}} UB_j. \tag{21}$$

In Algorithm 1, we present how group subproblem approach can be used to obtain lower and upper bounds for a multi-stage risk-averse mixed-integer problem with mean-CVaR objective.

4 Computational Experiments

In this section, we conduct our numerical experiments on a multi-stage lot sizing problem studied in Guan et al. [2009]. All computational experiments are performed on an Intel(R) Core(TM) i7-4790 CPU@3.60 GHz computer with 8.00 GB of RAM with Java 1.8.0.31 and IBM ILOG CPLEX 12.6. We first introduce risk-averse multi-stage lot sizing problem (RAMLSP) with mean-CVaR risk measure. Then, we compare the results obtained via usage of different scenario partition strategies and lower bound choices. We also compare the proposed algorithm and CPLEX in terms of solution quality and required CPU time.

4.1 Risk-averse Multi-stage Lot Sizing Problem with Mean-CVaR

The objective of RAMLSP is to minimize the dynamic risk measure over T periods defined via mean-CVaR risk measures subject to demand satisfaction and capacity constraints. RAMLSP-T-r represents a RAMLSP instance with T stages in which random components can take r different values at each stage. Therefore, total number of scenarios in an RAMLSP-T-r instance is r^{T-1}. We generate random test instances as in Guan et al. [2009]. The same setting of the parameters is also used by Sandıkçı and

Algorithm 1 Lower and upper bounds for **(P)**

Input: A risk-averse mixed-integer multi-stage stochastic problem **(P)** and a partition $\mathcal{S} = \{S_j\}_{j=1}^J$ of sample space Ω.
Initialize: $LB \leftarrow -\infty$ and $UB \leftarrow +\infty$

```
Lower Bounding:
```
for all $j \in \{1, 2, \dots, J\}$ **do**
  Solve the j^{th} group subproblem.
  $x^j \leftarrow$ an optimal solution of j^{th} group subproblem
  $z^j \leftarrow$ optimal value of j^{th} group subproblem
end for
Let Z_{LB} be a random variable that takes value z^j with probability $p_j = \sum_{\omega \in S_j} p_\omega$
$LB \leftarrow \rho_{\mathscr{G}}(Z_{LB})$

```
Upper Bounding:
```
for all $j \in \{1, 2, \dots, J\}$ **do**
  $UB_j \leftarrow$ the optimal value of the original problem with the additional constraint where (some of) the variables appearing in j^{th} group subproblem are set to x^j.
end for
$UB \leftarrow \min_{j \in \{1,2\dots,J\}} UB_j$

Return: LB and UB

Özaltın [2014], that is, $h_{tu} \sim U[0, 10]$, $\alpha_{tu} \sim U[3.2, 4.8]E[h]$, $\beta_{tu} \sim U[320, 480]E[h]$, $d_{tu} \sim U[0, 100]$ and $M_{tu} \sim U[40T, 60T]$, where $U[a, b]$ represents discrete uniform distribution between a and b.

Using the scenario tree representation given in Section 2.2, RAMLSP can be stated as follows:

$$\textbf{(RAMLSP)} \quad \text{minimize} \quad Z_1 + \rho_{\mathscr{F}_2|\mathscr{F}_1}\left(Z_2 + \rho_{\mathscr{F}_3|\mathscr{F}_2}\left(Z_3 + \ldots + \rho_{\mathscr{F}_T|\mathscr{F}_{T-1}}(Z_T)\ldots\right)\right), \tag{22}$$

$$\text{subject to} \quad Z_{tu} = \alpha_{tu}x_{tu} + \beta_{tu}y_{tu} + h_{tu}s_{tu}, \quad \forall t = 1, \dots, T \text{ and } u \in \Omega_t, \tag{23}$$

$$s_{(t-1)a(u)} + x_{tu} = d_{tu} + s_{tu}, \quad \forall t = 1, \dots, T \text{ and } u \in \Omega_t, \tag{24}$$

$$x_{tu} \le M_{tu}y_{tu}, \quad \forall t = 1, \dots, T \text{ and } u \in \Omega_t, \tag{25}$$

$$x_{tu}, s_{tu} \ge 0 \text{ and integer}, y_{tu} \in \{0, 1\} \quad \forall t = 1, \dots, T \text{ and } u \in \Omega_t, \tag{26}$$

$$s_{a(v_1)} = 0.$$

Here x_{tu} is the production level, y_{tu} is the setup indicator and s_{tu} is the inventory level variables at node $u \in \Omega_t$ in period $t = 1, \dots, T$. $\alpha_{tu}, \beta_{tu}, h_{tu}, d_{tu}$ and M_{tu} denote unit production cost, setup cost, inventory holding cost, demand and production capacity parameters, respectively. Z_1 is the total of deterministic production, setup and inventory holding costs incurred in the first stage. Similarly, Z_{tu} is the cost incurred at node $u \in \Omega_t$ at stage $t = 2, \dots, T$. Z_t represents the random variable that takes values of $Z_{tu}, u \in \Omega_t$ with respective probabilities. The objective (22) is the dynamic risk value over the planning horizon. Constraints (23) calculate the cost incurred at each node of the scenario tree. Constraints (24) and (25) are inventory balance and capacity constraints. Constraints (26) are domain constraints. Unlike Guan et al. [2009] and Sandıkçı and Özaltın [2014], we assume that production and inventory levels are required to be integer valued. Although this assumption increases the problem complexity, we have a more realistic representation to evaluate the performance of the algorithm.

For the computational experiments, we use three different values of weight parameter $\epsilon_1 \in \{0.8, 0.5, 0.3\}$ and level parameter $\alpha \in \{0.9, 0.8, 0.7\}$ of mean-CVaR. Therefore, we have nine different risk aversion settings.

4.2 Choices of Scenario Partitions and Lower Bounds

As seen in Example 1, the value of each lower bound highly depends on chosen scenario partition. We consider four possible scenario partitions obtained by different scenario grouping strategies, namely *index1*, *index2*, *similar* and *different*. Partitions *index1* and *index2* are based on scenario tree structure. In partition *index1*, the last stage nodes sharing the same ancestor node are placed in the same group. On the other hand, *index2* is obtained by placing the last stage nodes with different ancestor nodes in the same group.

If a priori information on the cost of each single scenario under the optimal solution is available, the groups can also be obtained with respect to similarity and diversity of individual scenarios. However, this information is not available before solving the original problem. Therefore, the deterministic version of the original problem can be solved for each scenario separately, and these optimal values can be used to approximate the cost of each scenario under an optimal solution of original problem. In partition *similar*, the partition is obtained by placing the scenarios with close approximate costs in a group. On the other hand, partition *different* is obtained by placing the scenarios with distant approximate costs in a group. Note that, for partitions *similar* and *different*, an additional computational effort is required to obtain approximate costs.

Example 3: Figure 3 depicts the scenario tree for a RAMLSP-3-4 instance where numbers near the scenarios indicate the cost of each individual scenario. The scenarios can be ordered as $\omega_9, \omega_4, \omega_{11}, \omega_6, \omega_7, \omega_{16}, \omega_{13}, \omega_1, \omega_{10}, \omega_2, \omega_8, \omega_3, \omega_{12}, \omega_{14}, \omega_{15}, \omega_5$ where the individual scenario costs decrease moving through from ω_9 to ω_5. Table 3 presents different scenario partition strategies for this scenario tree.

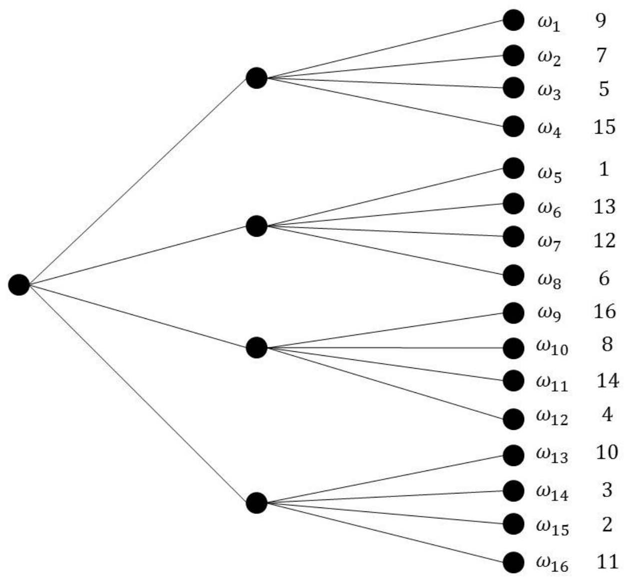


Figure 3: An example of three-stage scenario tree with 16 scenarios.

Insert Table 3 here

In order to observe the quality of bounds obtained by different scenario partitions and lower bound choices, the proposed algorithm is applied to five RAMLSP-3-30 instances generated via different random seeds. Total number of scenarios is 900. We consider the number of groups as $J = 2, 4$ and 10, and hence each group subproblem includes 450, 225 and 90 scenarios, for the respective value of J. While obtaining upper bounds, optimal production decisions of group subproblems are fixed in the restricted problems. For partitions *similar* and *different*, an approximate cost of each scenario is required. The CPU time needed to obtain the approximate costs is also included in the running time of the algorithm. In order to measure the quality of lower and upper bounds, an optimality gap information $Gap(\%) = 100\,((UB - LB)/UB)$ is used. All running times are presented in seconds. The results are presented in Table 4, where the gap and time values are the average values of five randomly generated instances.

Insert Table 4 here

The bolded entries in Table 4 correspond to the smallest average optimality gap values among all lower bound choices, partition strategies and number of group values. It is clear that $\mathbb{E} \circ \rho$ is the best lower bound choice. For all instances, none of the other two lower bound choices yields a better optimality gap value than $\mathbb{E} \circ \rho$. This is a consequence of the fact that group subproblems with original dynamic risk measure reflect the risk aversion behaviour of the original problem better. The smallest average optimality gap value 0.22% is obtained with partition *different*, lower bound choice $\mathbb{E} \circ \rho$ and $J = 2$.

The lower bound choice $\mathbb{E} \circ \rho$ and the partition strategy *different* is the most promising combination among all bound and partition combinations. Therefore, further computational experiments are conducted on the instances with more stages under this setting. We also conduct a set of computational experiments to compare the performance of the proposed algorithm with CPLEX in terms of optimality gap and solution time. For this comparison, using the linearization of mean-CVaR presented in Section 2.3, we solve the linearized version of RAMLSP with CPLEX.

4.3 Computational Study Results for More Stages

In the upper bounding phase of the proposed algorithm, the restricted problem is solved for each group. When the number of groups J in a partition is large, the upper bounding phase requires long CPU times. Therefore, one may solve the restricted problem for only a subset of groups. Another computational enhancement for the upper bounding phase is running the restricted problems with a prespecified time limit and report the objective value of current incumbent solution as UB_j. Since, the optimal value of the restricted problem is an upper bound for the original problem, the objective of any incumbent solution is also a valid upper bound.

We solve RAMLSP-3-64, RAMLSP-4-8 and RAMLSP-5-4 problems with 3, 4 and 5 stages, respectively, and for each risk setting, we generate five instances using different random seeds. The algorithm is applied with lower bound choice $\mathbb{E} \circ \rho$ and the partition *different* where number of groups, J, takes values of 4, 8, 16 and 32. The number of restricted problems to be solved is $\lceil J/5 \rceil$, which are selected randomly. The time limit for each restricted problem is set to 10 seconds. The results are presented in Table 5.

Insert Table 5 here

As seen in Table 5, increasing the number of groups in the partition may not always yield CPU time saving. As J increases, the average optimality gap increases, on the other hand, the CPU time may not always decrease. Specifically, when J is increased to 32 from 16, the CPU time increases in all of the instances. As the number of groups J increases, the subproblems get smaller in size. However the number of group subproblems and the restricted problems to solve increases. Therefore, increasing the number of groups may not always result in a decrease in the running time of the algorithm.

An interesting question is the comparison of the proposed algorithm with CPLEX in terms of optimality gap and CPU time. To make a fair comparison, we use RAMLSP-3-64 instances where CPLEX is run as long as it reaches to the optimality gap or the CPU time of the proposed algorithm.

When CPLEX is allowed to run with one hour of time limit, it cannot solve none of the instances optimally. Table 6 presents the comparison of the proposed algorithm with CPLEX under the setting of $J = 4$.

Insert Table 6 here

In Table 6, the column "*Gap_ CPLEX*" corresponds to the optimality gap value reported by CPLEX when is is allowed to run as long as the running time of the proposed algorithm. Moreover, the values in the column "*Looseness*" is measured as the ratio of CPLEX gap to the gap obtained by the proposed algorithm. When CPLEX is allowed to run as long as the solution time of the proposed algorithm, the algorithm yields 2.58 times stronger bounds on the average. For example, when $\alpha = 0.7$ and $\epsilon_1 = 0.8$, our algorithm terminates with an optimality gap of 0.54% within 51.8 seconds. CPLEX stops with an optimality gap of 2.47% within the same time limit, that is, the bounds obtained by CPLEX is 4.58 times looser than the bounds obtained by our algorithm.

In Table 6, the column "*Time_ CPLEX*" corresponds to the amount of seconds CPLEX took to reduce its gap to the level of the gap obtained by the proposed algorithm. Also, the values in the column "*Delay*" is measured as the ratio of amount of time CPLEX took to reduce its gap to the level of the gap obtained by the proposed algorithm to the running time of the algorithm. CPLEX requires 5.45 times longer running time to achieve the optimality gap of the proposed algorithm, on the average. For $\alpha = 0.9$ and $\epsilon_1 = 0.5$, CPLEX requires 2366.7 seconds to achieve the optimality gap of the proposed algorithm, that means CPLEX needs to spend more than 17 times running time in order to reach the optimality gap of the proposed algorithm. These results show that the proposed algorithm outperforms CPLEX with respect to both optimality gap and running time.

5 Conclusion

In this paper, we propose a group subproblem approach for risk-averse mixed-integer multi-stage stochastic problems with different dynamic risk measures defined by mean-CVaR. To the best of our knowledge, this is the first study where group subproblem approach is applied to a risk-averse problem with an objective of a dynamic risk measure. We show that infinitely many lower bounds on the optimal value of the problem can be obtained using different convolution of mean-CVaR risk measures. An upper bound is obtained through the use of optimal solutions of group subproblems, as well. The results are tested by a computational study on a multi-stage lot sizing problem. The effect of partition strategies and lower bound choices on the optimality gap of the proposed algorithm is investigated. It is revealed that, on the average, the optimality gap of the proposed algorithm is 2.58 times stronger than the optimality gap of CPLEX within the same running time. By solving the original problem with CPLEX, the optimality gaps of our algorithm can be achieved with additional running time more than a factor of five.

In the lower bounding phase of the proposed algorithm, the group subproblems can be assigned to different threads of a computer and solved in parallel. Similarly, parallel computing can be used to solve the restricted problems of the upper bounding phase. The parallel implementation of the proposed algorithm may possibly decrease the running time significantly.

Another possible extension of the study is to find better scenario partition strategies. Finding optimal grouping strategies is still an interesting research direction.

References

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and bellman's principle. *Annals of Operations Research*, 152(1):5–22, 2007.

J. F. Bonnans, Z. Cen, and T. Christel. Energy contracts management by stochastic programming techniques. *Annals of Operations Research*, 200(1):199–222, 2012.

S. Bruno, S. Ahmed, A. Shapiro, and A. Street. Risk neutral and risk averse approaches to multistage renewable investment planning under uncertainty. *European Journal of Operational Research*, 250 (3):979–989, 2016.

R. A. Collado, D. Papp, and A. Ruszczyński. Scenario decomposition of risk-averse multistage stochastic programming problems. *Annals of Operations Research*, 200(1):147–170, 2012.

A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16(1):69–95, 2005.

Y. Guan, S. Ahmed, and G. L. Nemhauser. Cutting planes for multistage stochastic integer programs. *Operations research*, 57(2):287–298, 2009.

T. Homem-de Mello and B. K. Pagnoncelli. Risk aversion in multistage stochastic programming: A modeling and algorithmic perspective. *European Journal of Operational Research*, 249(1):188–199, 2016.

D. A. Iancu, M. Petrik, and D. Subramanian. Tight approximations of dynamic risk measures. *Mathematics of Operations Research*, 40(3):655–682, 2015.

R. Kovacevic and G. C. Pflug. Time consistency and information monotonicity of multiperiod acceptability functionals. *Advanced financial modelling*, 8:347, 2009.

F. Maggioni and G. C. Pflug. Bounds and approximations for multistage stochastic programs. *SIAM Journal on Optimization*, 26(1):831–855, 2016.

F. Maggioni, E. Allevi, and M. Bertocchi. Monotonic bounds in multistage mixed-integer stochastic programming. *Computational Management Science*, 13(3):423–457, 2016.

M. Pereira and L. M. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52(1-3):359–375, 1991.

G. C. Pflug and W. Römisch. *Modeling, measuring and managing risk*, volume 190. World Scientific, 2007.

A. Philpott, V. de Matos, and E. Finardi. On solving multistage stochastic programs with coherent risk measures. *Operations Research*, 61(4):957–970, 2013.

R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

A. Ruszczyński. Risk-averse dynamic programming for markov decision processes. *Mathematical programming*, 125(2):235–261, 2010.

A. Ruszczynski and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31 (3):544–561, 2006a.

A. Ruszczynski and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31(3):433–452, 2006b.

B. Sandıkçı and O. Y. Özaltın. A scalable bounding method for multi-stage stochastic integer programs, 2014. Available on Optimization Online, submitted for publication.

B. Sandıkçı, N. Kong, and A. J. Schaefer. A hierarchy of bounds for stochastic mixed-integer programs. *Mathematical Programming*, 138(1-2):253–272, 2013.

R. Schultz. Stochastic programming with integer variables. *Mathematical Programming*, 97(1-2):285–309, 2003.

A. Shapiro. On a time consistency concept in risk averse multistage stochastic programming. *Operations Research Letters*, 37(3):143–147, 2009.

A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63–72, 2011.

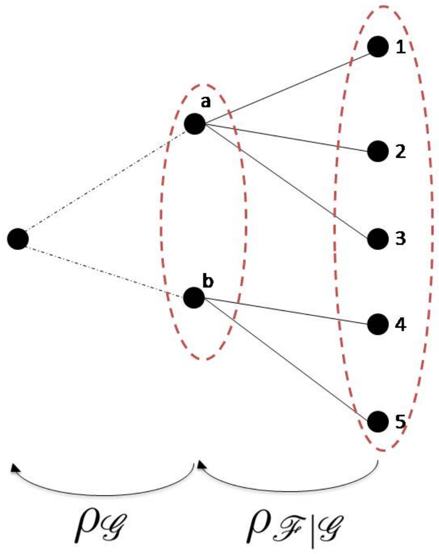
A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*, volume 9. SIAM, 2009.

A. Shapiro, W. Tekaya, J. P. da Costa, and M. P. Soares. Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research*, 224(2):375–391, 2013.

G. L. Zenarosa, O. A. Prokopyev, and A. J. Schaefer. Scenario-tree decomposition: Bounds for multistage stochastic mixed-integer programs, 2014. Available on Optimization Online, submitted for publication.

Appendix A

Construction of set $\widetilde{\mathcal{A}}$ for the example in Figure 2.



$$\mathcal{A}_{\mathscr{G}} = \Big\{ (\mu_a, \mu_b) \in \mathbb{R}^2 : 1 - \epsilon_1^1 \leq \mu_a \leq 1 + \epsilon_2^1, 1 - \epsilon_1^1 \leq \mu_b \leq 1 + \epsilon_2^1, (p_1 + p_2 + p_3)\mu_a + (p_4 + p_5)\mu_b = 1 \Big\}.$$

$$\mathcal{A}_{\mathscr{F}|\mathscr{G}} = \Big\{ (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) \in \mathbb{R}^5 : 1 - \epsilon_1^2 \leq \mu_1 \leq 1 + \epsilon_2^2, 1 - \epsilon_1^2 \leq \mu_2 \leq 1 + \epsilon_2^2, 1 - \epsilon_1^2 \leq \mu_3 \leq 1 + \epsilon_2^2, \frac{p_1}{p_1 + p_2 + p_3}\mu_1 + \frac{p_2}{p_1 + p_2 + p_3}\mu_2 + \frac{p_3}{p_1 + p_2 + p_3}\mu_3 = 1, 1 - \epsilon_1^2 \leq \mu_4 \leq 1 + \epsilon_2^2, 1 - \epsilon_1^2 \leq \mu_5 \leq 1 + \epsilon_2^2, \frac{p_4}{p_4 + p_5}\mu_4 + \frac{p_5}{p_4 + p_5}\mu_5 = 1 \Big\}.$$

$$\widetilde{\mathcal{A}} = \mathcal{A}_{\mathscr{F}|\mathscr{G}} \circ \mathcal{A}_{\mathscr{G}} = \Big\{ (\mu_a\mu_1, \mu_a\mu_2, \mu_b\mu_3, \mu_b\mu_4, \mu_b\mu_5) \in \mathbb{R}^5 : (\mu_a, \mu_b) \in \mathcal{A}_{\mathscr{G}}, (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) \in \mathcal{A}_{\mathscr{F}|\mathscr{G}} \Big\}.$$

		$\rho_{\mathscr{G}}$			$\rho_{\mathscr{F}\|\mathscr{G}}$		
	$\rho_{\mathscr{G}} \circ \rho_{\mathscr{F}\|\mathscr{G}}$	ϵ_1^1	ϵ_2^1	α^1	ϵ_1^2	ϵ_2^2	α^2
Bound 1	$\rho \circ \mathbb{E}$	ϵ_1	ϵ_2	α	0	0	0
Bound 2	$\rho^s \circ \rho^s$	$1-\sqrt{1-\epsilon_1}$	$\sqrt{1+\epsilon_2}-1$	$\frac{\sqrt{1+\epsilon_2}-1}{\sqrt{1+\epsilon_2}-\sqrt{1-\epsilon_1}}$	$1-\sqrt{1-\epsilon_1}$	$\sqrt{1+\epsilon_2}-1$	$\frac{\sqrt{1+\epsilon_2}-1}{\sqrt{1+\epsilon_2}-\sqrt{1-\epsilon_1}}$
Bound 3	$\mathbb{E} \circ \rho$	0	0	0	ϵ_1	ϵ_2	α

Table 1: Possible choices of $\rho_{\mathscr{G}}(\cdot)$ and $\rho_{\mathscr{F}|\mathscr{G}}(\cdot)$ that can be used to obtain lower bound on mean-CVaR risk measure $\rho(\cdot)$.

	$\mathcal{S}$	$\mathcal{S}'$
Bound 1	3.5	2.5
Bound 2	3.12	3
Bound 3	3	3.5

Table 2: Values of Bounds 1,2 and 3 for Example 1.

Partition	S_1	S_2	S_3	S_4
index1	$\{\omega_1, \omega_2, \omega_3, \omega_4\}$	$\{\omega_5, \omega_6, \omega_7, \omega_8\}$	$\{\omega_9, \omega_{10}, \omega_{11}, \omega_{12}\}$	$\{\omega_{13}, \omega_{14}, \omega_{15}, \omega_{16}\}$
index2	$\{\omega_1, \omega_5, \omega_9, \omega_{13}\}$	$\{\omega_2, \omega_6, \omega_{10}, \omega_{14}\}$	$\{\omega_3, \omega_7, \omega_{11}, \omega_{15}\}$	$\{\omega_4, \omega_8, \omega_{12}, \omega_{16}\}$
similar	$\{\omega_9, \omega_4, \omega_{11}, \omega_6\}$	$\{\omega_7, \omega_{16}, \omega_{13}, \omega_1\}$	$\{\omega_{10}, \omega_2, \omega_8, \omega_3\}$	$\{\omega_{12}, \omega_{14}, \omega_{15}, \omega_5\}$
different	$\{\omega_9, \omega_7, \omega_{10}, \omega_{12}\}$	$\{\omega_4, \omega_{16}, \omega_2, \omega_{14}\}$	$\{\omega_{11}, \omega_{13}, \omega_8, \omega_{15}\}$	$\{\omega_6, \omega_1, \omega_3, \omega_5\}$

Table 3: Different scenario partitions $\mathcal{S} = \{S_1, S_2, S_3, S_4\}$ for the example scenario tree in Figure 3.

			J = 2								J = 4								J = 10							
			index1		*index2*		*similar*		*different*		*index1*		*index2*		*similar*		*different*		*index1*		*index2*		*similar*		*different*	
LB choice	α	ϵ_1	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time
$\rho \circ \mathbb{E}$	0.9	0.8	10.86%	7.1	12.45%	16.5	9.34%	11.1	12.54%	13.4	9.45%	2.0	13.07%	6.9	7.71%	2.1	13.71%	6.4	4.26%	7.8	18.54%	9.7	5.10%	6.6	16.99%	6.0
		0.5	7.64%	24.5	8.81%	24.8	7.92%	29.6	8.67%	23.9	6.85%	3.1	9.36%	8.6	7.97%	2.5	9.78%	8.5	3.63%	9.7	11.93%	8.6	7.01%	7.2	12.73%	5.4
		0.3	5.32%	18.2	5.79%	52.6	6.56%	95.3	5.69%	48.2	4.91%	4.1	6.33%	13.4	8.29%	3.7	6.73%	11.7	3.01%	11.4	8.34%	7.9	8.32%	7.3	8.28%	7.3
	0.8	0.8	9.61%	13.2	10.82%	21.9	7.16%	14.2	10.97%	24.0	8.14%	3.3	11.69%	9.6	5.32%	3.7	11.58%	10.1	3.46%	8.2	15.55%	10.3	6.89%	8.9	14.43%	5.5
		0.5	6.57%	22.0	7.20%	55.0	6.28%	116.0	7.40%	31.9	5.81%	4.0	7.80%	11.2	6.05%	3.3	8.15%	11.4	3.00%	9.8	9.89%	8.8	7.83%	7.2	10.51%	5.7
		0.3	4.44%	26.6	4.75%	88.6	5.46%	118.9	4.88%	99.2	4.12%	5.0	5.21%	14.9	7.01%	4.8	5.56%	13.8	2.59%	12.2	6.74%	8.0	8.68%	7.5	6.71%	7.5
	0.7	0.8	8.86%	21.2	10.11%	74.9	5.35%	24.2	10.10%	44.4	7.60%	4.0	10.49%	12.0	5.33%	4.5	10.51%	12.6	3.85%	7.7	13.09%	9.1	7.82%	7.6	12.70%	7.1
		0.5	5.90%	31.5	6.56%	73.7	5.11%	74.2	6.62%	89.7	5.24%	4.3	6.98%	16.3	6.31%	4.2	7.35%	14.9	3.08%	12.1	8.22%	8.5	8.45%	7.3	9.17%	7.5
		0.3	3.95%	36.4	4.27%	114.0	4.73%	144.6	4.29%	139.9	3.70%	5.6	4.50%	17.0	7.08%	5.4	4.80%	17.2	2.48%	14.5	5.50%	8.5	8.91%	7.8	5.81%	10.6
$\rho^s \circ \rho^s$	0.9	0.8	4.52%	7.2	5.52%	14.8	7.57%	8.3	5.10%	9.6	4.38%	1.6	5.99%	6.1	9.83%	1.8	6.03%	6.3	4.20%	6.7	8.95%	4.4	14.27%	6.8	8.76%	5.2
		0.5	3.99%	10.6	4.43%	18.5	7.80%	16.6	4.23%	15.5	3.86%	2.3	4.91%	6.9	10.97%	1.7	5.13%	6.2	4.04%	9.9	7.18%	4.6	13.95%	7.3	7.67%	4.1
		0.3	3.08%	20.4	3.14%	21.5	6.57%	66.6	3.02%	20.1	3.04%	4.0	3.56%	7.4	10.47%	3.2	3.79%	10.6	3.38%	10.5	5.31%	6.0	12.83%	7.4	5.18%	4.3
	0.8	0.8	3.67%	12.5	4.45%	18.9	6.16%	17.3	4.40%	19.9	3.66%	4.1	5.06%	8.5	10.85%	2.2	5.06%	8.9	3.35%	8.5	7.58%	5.5	14.81%	8.7	7.41%	3.8
		0.5	3.30%	12.9	3.52%	22.0	6.68%	34.7	3.40%	59.3	3.12%	3.7	4.01%	10.2	9.91%	2.8	4.09%	9.6	3.09%	9.2	5.94%	5.4	13.86%	7.4	5.98%	4.8
		0.3	2.46%	21.1	2.49%	60.9	5.83%	166.9	2.41%	44.2	2.44%	4.2	2.85%	12.3	9.28%	4.1	2.98%	12.8	2.67%	10.3	4.26%	7.0	12.41%	7.6	4.07%	5.6
	0.7	0.8	3.37%	21.0	4.40%	19.8	5.10%	22.0	4.22%	28.6	3.39%	4.5	4.68%	11.1	10.73%	4.7	4.90%	8.5	3.80%	7.8	6.18%	5.1	15.10%	7.7	6.53%	3.8
		0.5	2.94%	25.1	3.26%	46.0	6.09%	52.5	3.10%	47.1	2.82%	6.1	3.62%	14.7	10.17%	3.4	3.63%	13.7	3.06%	10.2	4.72%	5.0	13.80%	7.3	5.12%	4.9
		0.3	2.30%	21.9	2.19%	73.9	5.57%	169.8	2.10%	95.8	2.26%	7.1	2.49%	15.2	9.50%	3.5	2.57%	16.4	2.39%	12.4	3.36%	7.0	12.46%	7.9	3.50%	8.8
$\mathbb{E} \circ \rho$	0.9	0.8	1.15%	6.4	0.65%	5.6	11.78%	6.8	**0.36%**	5.8	1.97%	2.9	1.10%	3.8	20.35%	1.9	1.00%	3.1	6.81%	9.3	3.80%	4.0	25.67%	6.7	4.31%	3.4
		0.5	1.26%	4.8	0.42%	11.9	9.54%	5.6	**0.26%**	9.1	1.82%	2.4	0.82%	4.3	16.15%	1.6	0.94%	4.4	5.09%	7.4	2.99%	3.1	20.73%	7.3	2.92%	3.1
		0.3	1.16%	9.2	0.43%	14.4	7.67%	14.7	**0.29%**	14.1	1.51%	3.1	0.76%	6.8	13.31%	2.0	0.90%	6.8	3.88%	10.7	2.41%	4.1	17.00%	7.4	2.00%	3.3
	0.8	0.8	0.59%	10.0	0.25%	10.2	10.95%	8.9	**0.15%**	10.9	1.23%	4.5	0.59%	6.9	18.86%	2.8	0.56%	7.4	5.17%	10.9	2.54%	4.1	24.29%	8.8	2.10%	3.9
		0.5	0.64%	11.3	0.27%	15.3	8.74%	16.8	**0.12%**	12.3	1.07%	3.2	0.59%	6.2	14.79%	1.9	0.52%	7.0	3.78%	9.2	2.08%	3.6	19.49%	7.3	1.89%	3.2
		0.3	0.60%	16.5	0.27%	17.8	7.13%	58.0	**0.19%**	18.8	1.04%	3.5	0.53%	7.1	12.36%	2.5	0.47%	8.5	2.94%	10.4	1.69%	4.4	15.95%	7.7	1.36%	4.4
	0.7	0.8	0.74%	11.2	0.26%	14.9	10.70%	13.0	**0.23%**	13.2	1.04%	4.3	0.49%	7.6	17.62%	4.5	0.45%	7.8	4.09%	8.0	1.72%	4.2	22.96%	7.8	1.77%	3.7
		0.5	0.63%	16.9	0.26%	14.7	8.47%	22.0	**0.22%**	18.2	0.87%	3.4	0.51%	7.3	14.15%	3.0	0.47%	8.3	3.02%	9.9	1.35%	3.4	18.68%	7.2	1.52%	3.1
		0.3	0.58%	40.8	0.23%	28.2	7.02%	71.0	**0.13%**	29.0	0.93%	5.8	0.47%	10.1	11.91%	4.3	0.40%	11.6	2.37%	12.6	1.25%	5.4	15.47%	7.9	1.27%	7.7

Table 4: Average optimality gap and running time values of the proposed algorithm for five different RAMLSP-3-30 instances with different partition and lower bound choices.

			J = 4		J = 8		J = 16		J = 32	
	α	ϵ_1	Gap	Time	Gap	Time	Gap	Time	Gap	Time
RAMLSP-3-64	0.9	0.8	1.02%	68.9	2.42%	55.8	5.02%	44.6	8.13%	82.0
		0.5	0.92%	132.6	2.07%	66.2	4.27%	60.9	6.99%	86.1
		0.3	0.74%	697.4	1.57%	101.4	3.33%	75.6	5.80%	93.1
	0.8	0.8	0.71%	27.5	1.67%	13.0	3.82%	26.3	7.10%	85.2
		0.5	0.66%	97.4	1.55%	25.4	3.19%	39.3	6.03%	82.4
		0.3	0.60%	376.1	1.20%	49.6	2.53%	55.6	4.76%	86.5
	0.7	0.8	0.54%	51.8	1.30%	14.5	2.82%	37.6	5.67%	82.9
		0.5	0.56%	129.2	1.12%	27.4	2.37%	48.2	4.74%	81.9
		0.3	0.48%	373.1	0.93%	63.7	1.96%	57.6	3.73%	84.9
RAMLSP-4-8	0.9	0.8	5.92%	10.4	6.05%	9.0	9.62%	17.6	11.23%	37.5
		0.5	5.49%	16.9	6.69%	19.3	7.28%	26.0	9.18%	59.1
		0.3	4.06%	21.3	5.87%	27.0	6.45%	37.8	7.76%	66.5
	0.8	0.8	3.59%	13.9	5.03%	11.2	6.69%	18.8	9.62%	50.0
		0.5	3.64%	18.6	5.53%	25.5	6.35%	33.2	8.26%	63.5
		0.3	3.40%	22.3	5.28%	29.3	5.86%	40.9	7.48%	62.0
	0.7	0.8	3.48%	16.0	4.95%	14.5	6.14%	22.0	8.38%	54.3
		0.5	3.19%	21.1	5.31%	24.4	5.92%	33.7	7.95%	61.8
		0.3	3.12%	28.3	4.86%	32.2	5.54%	41.1	7.09%	65.1
RAMLSP-5-4	0.9	0.8	6.55%	8.9	10.25%	4.8	13.24%	11.7	16.09%	29.9
		0.5	5.53%	13.7	9.09%	5.6	11.69%	16.0	13.27%	43.2
		0.3	5.12%	17.1	7.97%	8.0	10.02%	20.0	11.38%	48.2
	0.8	0.8	6.83%	9.9	10.78%	4.7	13.40%	11.3	16.16%	29.6
		0.5	5.70%	13.6	9.48%	6.2	11.81%	14.9	13.40%	43.7
		0.3	5.28%	16.4	8.27%	8.6	10.08%	17.4	11.53%	50.1
	0.7	0.8	6.04%	10.3	9.92%	6.0	13.04%	17.0	14.58%	40.5
		0.5	5.39%	11.7	8.96%	6.7	10.96%	20.9	12.49%	51.4
		0.3	4.93%	15.5	7.94%	9.8	9.53%	22.5	11.00%	56.9

Table 5: Average optimality gap and running time values of the proposed algorithm for five different RAMLSP-3-30, RAMLSP-4-8 and RAMLSP-5-4 instances with partition *different* and lower bound choice $\mathbb{E} \circ \rho$.

		Proposed Algorithm		DEP with CPLEX			
α	ϵ_1	Gap	Time	Gap_CPLEX	Looseness	Time_CPLEX	Delay
0.9	0.8	1.02%	68.9	2.76%	2.71	248.3	3.60
	0.5	0.92%	132.6	3.47%	3.78	2366.7	17.84
	0.3	0.74%	697.4	1.39%	1.89	719.9	1.03
0.8	0.8	0.71%	27.5	2.04%	2.87	241.9	8.79
	0.5	0.66%	97.4	1.48%	2.24	403.9	4.15
	0.3	0.60%	376.1	0.89%	1.47	603.0	1.60
0.7	0.8	0.54%	51.8	2.47%	4.58	246.4	4.75
	0.5	0.56%	129.2	1.13%	2.01	444.2	3.44
	0.3	0.48%	373.1	0.81%	1.70	1445.6	3.87

Table 6: Comparison of optimality gap and running time of the proposed algorithm and CPLEX.