



## INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Advances in Risk-Averse Optimization

Andrzej Ruszczyński

To cite this entry: Andrzej Ruszczyński. Advances in Risk-Averse Optimization. *In* INFORMS TutORials in Operations Research. Published online: 14 Oct 2014; 168-190.  
<https://doi.org/10.1287/educ.2013.0110>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2013, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Advances in Risk-Averse Optimization

*Andrzej Ruszczyński*

Department of Management Science and Information Systems, Rutgers University, Piscataway, New Jersey 08854, [rusz@business.rutgers.edu](mailto:rusz@business.rutgers.edu)

**Abstract** We discuss main ideas of formalizing risk-averse preferences and using them in optimization problems: utility theories, risk measures, and stochastic order constraints. We also present methods for solving risk-averse optimization problems. Multistage models are discussed as well.

**Keywords** dynamic measures of risk; stochastic order; decomposition methods

---

## 1. Introduction

Modeling and optimization of risk have been objects of intensive research in the last 15 years. Even the classical approaches, such as the expected and dual utility theories, experience revival of interest, but the main activity is associated with new models, involving measures of risk or stochastic order constraints, with increasing interest in dynamic models of risk. We try to discuss the new contributions and their relation to the existing knowledge. Also, because of special relevance for the operations research community, we characterize main ideas of numerical methods for solving risk-averse optimization models.

This article is not a typical review paper, but rather a compressed version of the notes of the author's lectures on risk-averse optimization. We hope that the readers will find it useful not only in their search for relevant literature but also for getting an overview of the area and its most important results.

This paper has two main parts. In §2 we discuss risk-averse preferences—the utility theories, measures of risk, and stochastic order constraints—by providing a concise and modern perspective on these models. In §3 we show how these models of preferences can be used in optimization problems and how these problems can be solved numerically. We include there a long subsection on dynamic risk-averse optimization, which is a relatively new direction of research, with the hope to inspire the interest of other scientists.

## 2. Risk-Averse Preferences

The question of numerical representation of preferences is a classical topic of economics. A *preference relation*  $\preceq$  on a *prospect space*  $\mathcal{Z}$  is a total preorder, that is, a binary relation that is reflexive, transitive, and complete. The corresponding indifference relation  $\sim$  is defined in a usual way:  $Z \sim V$ , if  $Z \preceq V$  and  $V \preceq Z$ . We say that  $Z$  is *strictly preferred* over  $V$  and write it as  $Z \prec V$ , if  $Z \preceq V$ , and  $V \not\preceq Z$ .

In this paper, we are mainly interested in the case when  $\mathcal{Z}$  is a space of random variables, that is, measurable functions  $Z: \Omega \rightarrow \mathbb{R}$ , where  $(\Omega, \mathcal{F}, P)$  is a probability space. If  $\mathcal{Z}$  is a topological space, we call a preference relation  $\preceq$  *continuous*, if for every  $Z \in \mathcal{Z}$  the sets  $\{V \in \mathcal{Z}: V \preceq Z\}$  and  $\{V \in \mathcal{Z}: Z \preceq V\}$  are closed.

A functional  $U: \mathcal{Z} \rightarrow \mathbb{R}$  is a *numerical representation* of the preference relation  $\preceq$  on  $\mathcal{Z}$ , if

$$Z \prec V \iff U(Z) < U(V).$$

It is known that a continuous preorder on a separable and connected topological space  $\mathcal{Z}$  has a continuous numerical representation Eilenberg [21, §6]. Other sufficient conditions have been formulated and analyzed in Debreu [9, 10, Theorem II], Rader [55, Theorem 1], and Jaffray [30].

In our applications, if  $(\Omega, \mathcal{F}, P)$  is a standard probability space, and  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$  with  $p \in [1, \infty]$ , a continuous preference relation does have a continuous numerical representation. We shall interpret the relations  $\triangleleft$  and  $\trianglelefteq$  as “less risky” and “no more risky,” respectively.

The preference relation  $\trianglelefteq$  on  $\mathcal{Z}$  is called *law invariant* if  $Z \succsim W$  implies that  $Z \sim W$ .

## 2.1. Utility Theory

Originally, the utility theory of von Neumann and Morgenstern [73] has been proposed for preference relations on the space of measures on  $\mathbb{R}$ . Here we adapt it to the case when the prospect space is the space  $\mathcal{Z}$  of real random variables.

**2.1.1. Axioms.** Every preference relation  $\preceq$  on the space of probability measures on  $\mathbb{R}$  defines a law invariant preference relation  $\trianglelefteq$  on  $\mathcal{Z}$  as follows:

$$Z \trianglelefteq W \iff \mu_Z \preceq \mu_W.$$

The converse statement is true, if we additionally require that every probability measure  $\mu$  on  $\mathbb{R}$  is a distribution of some  $Z \in \mathcal{Z}$ . This can be guaranteed if  $(\Omega, \mathcal{F}, P)$  is a standard atomless probability space (see Dudley [20, Theorem 11.7.5], Skorohod [71]). In this case, we can consider an operation on random variables in  $\mathcal{Z}$  corresponding to the operation of taking a convex combination of measures on  $\mathbb{R}$ .

For three elements  $Z, V$ , and  $W$  in  $\mathcal{Z}$ , we say that  $W$  is a *lottery* of  $Z$  and  $V$  with probabilities  $\alpha \in (0, 1)$  and  $(1 - \alpha)$ , if an event  $A \in \mathcal{F}$  of probability  $\alpha$  exists, such that the conditional distribution of  $W$ , given  $A$ , is the same as the (unconditional) distribution of  $Z$ , whereas the conditional distribution of  $W$ , given  $\bar{A} = \Omega \setminus A$ , is the same as the unconditional distribution of  $V$ . In this case, the probability measure  $\mu_W$  induced by  $W$  on  $\mathbb{R}$  is the corresponding convex combination of the probability measures  $\mu_Z$  and  $\mu_V$  of  $Z$  and  $V$ , respectively:

$$\mu_W = \alpha\mu_Z + (1 - \alpha)\mu_V.$$

We write the lottery symbolically as

$$W = \alpha Z \oplus (1 - \alpha)V.$$

It should be stressed that only the distribution  $\mu_W$  of the lottery is defined uniquely, not the random variable  $W$  itself. However, if the preference relation  $\trianglelefteq$  on  $\mathcal{Z}$  is law invariant, it makes sense to compare lotteries.

For law-invariant preferences on the space of real random variables, we introduce axioms corresponding to the axioms of the expected utility theory for distributions, due to von Neumann and Morgenstern [73].

**Independence Axiom.** For all  $Z, V, W \in \mathcal{Z}$  one has

$$Z \triangleleft V \implies \alpha Z \oplus (1 - \alpha)W \triangleleft \alpha V \oplus (1 - \alpha)W \quad \forall \alpha \in (0, 1).$$

**Archimedean Axiom.** If  $Z \triangleleft V \triangleleft W$ , then  $\alpha, \beta \in (0, 1)$  exist such that

$$\alpha Z \oplus (1 - \alpha)W \triangleleft V \triangleleft \beta Z \oplus (1 - \beta)W.$$

If a total preorder  $\trianglelefteq$  on  $\mathcal{Z}$  satisfies the independence and Archimedean axioms, then a numerical representation  $U: \mathcal{Z} \rightarrow \mathbb{R}$  of  $\trianglelefteq$  exists, which satisfies for all  $Z, V \in \mathcal{Z}$  and all  $\alpha \in [0, 1]$  the equation

$$U(\alpha Z \oplus (1 - \alpha)V) = \alpha U(Z) + (1 - \alpha)U(V).$$

The original proof of this statement (formulated in the space of probability distributions) is due to von Neumann and Morgenstern [73]; for a modern approach via convex analysis, see Dentcheva and Ruszczyński [18].

**2.1.2. Utility Functions.** To formulate the integral representation of the utility functional  $U(\cdot)$ , we need to introduce the topology on the space  $\mathcal{Z}$  and assume continuity of the preorder  $\preceq$  in this topology. For this purpose we adopt the topology of convergence in distribution. Recall that a sequence of random vectors  $Z_n: \Omega \rightarrow \mathbb{R}$  converges in distribution to a random vector  $Z: \Omega \rightarrow \mathbb{R}$ , if for every continuous and bounded function  $f: \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(Z_n)] = \mathbb{E}[f(Z)].$$

If a total preorder  $\preceq$  on  $\mathcal{Z}$  is law invariant, continuous with respect to convergence in distribution, and satisfies the independence axiom, then a continuous and bounded function  $u: \mathbb{R} \rightarrow \mathbb{R}$  exists, such that the functional

$$U(Z) = \mathbb{E}[u(Z)] = \int_{\Omega} u(Z(\omega))P(d\omega) \tag{1}$$

is a numerical representation of  $\preceq$  on  $\mathcal{Z}$  (see Fishburn [23], Föllmer and Schied [26], Dentcheva and Ruszczyński [18], and references therein).

In this paper we assume that lower values of the outcome  $Z$  are preferred; that is,  $Z$  represents “cost” or “loss.” Then it makes sense to assume that the preorder  $\preceq$  is *monotonic*; that is,  $Z \leq V$  a.s. implies that  $Z \preceq V$ . It is obvious that for a monotonic preorder, the utility function  $u(\cdot)$  in the numerical representation (1) is *nondecreasing*.

**2.1.3. Risk Aversion.** If the space  $\mathcal{Z}$  contains only random variables that are integrable up to power  $p \geq 1$ , i.e.,  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ , then we can prove that the representation (1) is true with a utility function  $u(\cdot)$ , which is not necessarily bounded but has growth rate not exceeding  $p$ , that is,  $|u(z)| \leq C(1 + z^p)$  for some constant  $C$ . In this case the expected value

$$\mathbb{E}[Z] = \int_{\Omega} Z(\omega)P(d\omega)$$

is well defined. We can call a preference relation *weakly risk averse* if  $\mathbb{E}[Z] \preceq Z$  for all  $Z \in \mathcal{Z}$ . In this case, the utility function  $u(\cdot)$  in the numerical representation (1) is *convex*.

Observe that the convexity of  $u(\cdot)$  implies the convexity of the numerical representation  $U(\cdot)$  in  $\mathcal{Z}$ . For every  $\alpha \in [0, 1]$ , we have

$$\begin{aligned} U(\alpha Z + (1 - \alpha)V) &= \int_{\Omega} u(\alpha Z(\omega) + (1 - \alpha)V(\omega))P(d\omega) \\ &\leq \alpha \int_{\Omega} u(Z(\omega))P(d\omega) + (1 - \alpha) \int_{\Omega} u(V(\omega))P(d\omega) \\ &= \alpha U(Z) + (1 - \alpha)U(V). \end{aligned} \tag{2}$$

## 2.2. Dual Utility Theory

**2.2.1. Axioms.** Again, we are considering a law invariant preorder  $\preceq$  in the space  $\mathcal{Z}$  of real random variables on a standard and atomless probability space  $(\Omega, \mathcal{F}, P)$ . However, we make different assumptions than those of the expected utility theory. Instead of considering lotteries, we consider convex combinations of comonotonic random variables. Recall that real random variables  $Z_i, i = 1, \dots, n$ , are *comonotonic*, if

$$(Z_i(\omega) - Z_i(\omega'))(Z_j(\omega) - Z_j(\omega')) \geq 0$$

for all  $\omega, \omega' \in \Omega$  and all  $i, j = 1, \dots, n$ .

The following axioms were formulated in Yaari [74], when the theory of dual utility was axiomatized.

**Dual Independence Axiom.** For all comonotonic random variables  $Z, V,$  and  $W$  in  $\mathcal{Z}$ , one has

$$Z \triangleleft V \implies \alpha Z + (1 - \alpha)W \triangleleft \alpha V + (1 - \alpha)W \quad \forall \alpha \in (0, 1).$$

**Dual Archimedean Axiom.** For all comonotonic random variables  $Z, V,$  and  $W$  in  $\mathcal{Z}$ , satisfying the relations  $Z \triangleleft V \triangleleft W$ , numbers  $\alpha, \beta \in (0, 1)$  exist such that

$$\alpha Z + (1 - \alpha)W \triangleleft V \triangleleft \beta Z + (1 - \beta)W.$$

In addition to that, we assume that the preorder  $\trianglelefteq$  is *monotonic*.

Then a numerical representation  $U: \mathcal{Z} \rightarrow \mathbb{R}$  of  $\trianglelefteq$  exists, which satisfies, for all comonotonic  $Z, V \in \mathcal{Z}$  and all  $\alpha, \beta \in \mathbb{R}_+$ , the equation

$$U(\alpha Z + \beta V) = \alpha U(Z) + \beta U(V).$$

Moreover,  $U(\cdot)$  is a *certainty equivalent*; that is,

$$U(Z)\mathbb{1} \sim Z \quad \text{for all } Z \in \mathcal{Z}. \tag{3}$$

In the last equation, the symbol  $\mathbb{1}$  denotes the sure outcome of 1.

**2.2.2. Rank-Dependent Utility Functions.** If, additionally,  $\mathcal{Z}$  is the space of *bounded* random variables equipped with the the norm topology of the space  $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ , and the preorder  $\trianglelefteq$  is continuous, then a bounded, nondecreasing, and continuous function  $w: [0, 1] \rightarrow \mathbb{R}_+$  exists, such that

$$U(Z) = \int_0^1 F_Z^{-1}(\beta) dw(\beta), \quad Z \in \mathcal{Z}. \tag{4}$$

In the formula above,  $F_Z^{-1}(\cdot)$  is the inverse distribution function of  $Z$ ; that is,

$$F_Z^{-1}(\alpha) = \inf\{\eta: F_Z(\eta) \geq \alpha\}, \tag{5}$$

where

$$F_Z(\eta) = P\{Z \leq \eta\}, \quad \eta \in \mathbb{R}, \tag{6}$$

is the usual distribution function. The function  $w(\cdot)$  in (4) is called the *rank-dependent utility function* (see Dentcheva and Ruszczyński [18] and the references therein).

**2.2.3. Risk Aversion.** To discuss risk aversion, we assume again that the space  $\mathcal{Z}$  contains only integrable random variables. Then for every  $Z \in \mathcal{Z}$  and every  $\sigma$ -subalgebra  $\mathcal{G}$  of  $\mathcal{F}$ , the conditional expectation  $\mathbb{E}[Z | \mathcal{G}]: \Omega \rightarrow \mathbb{R}$  is well defined.

A preference relation  $\trianglelefteq$  on  $\mathcal{Z}$  is called *risk averse* if  $\mathbb{E}[Z | \mathcal{G}] \trianglelefteq Z$ , for every  $Z \in \mathcal{Z}$  and every  $\sigma$ -subalgebra  $\mathcal{G} \subseteq \mathcal{F}$ . In this case the integral representation of  $\trianglelefteq$  is true with a *nondecreasing and concave* function  $w: [0, 1] \rightarrow [0, 1]$  such that  $w(0) = 0$  and  $w(1) = 1$ .

Again, the convexity of  $w(\cdot)$  implies the convexity of the numerical representation  $U(\cdot)$  in  $\mathcal{Z}$ . For every  $\alpha \in [0, 1]$  we have

$$\begin{aligned} U(\alpha Z + (1 - \alpha)V) &= \int_0^1 F_{\alpha Z + (1 - \alpha)V}^{-1}(\beta) dw(\beta) \\ &\leq \alpha \int_0^1 F_Z^{-1}(\beta) dw(\beta) + (1 - \alpha) \int_0^1 F_V^{-1}(\beta) dw(\beta) \\ &= \alpha U(Z) + (1 - \alpha)U(V). \end{aligned} \tag{7}$$

This inequality can be verified by integration by parts, with the use of convexity of  $w(\cdot)$ .

Finally, an equivalent expression for the functional (4) is the following Choquet integral representation (Schmeidler [68]):

$$U(Z) = - \int_{-\infty}^0 w(F_Z(\eta)) d\eta + \int_0^{\infty} [1 - w(F_Z(\eta))] d\eta. \tag{8}$$

Again, it can be derived via integration by parts.

## 2.3. Measures of Risk

**2.3.1. Axioms.** In the theory of measures of risk, the preference relation  $\triangleleft$  is specified directly through its numerical representation  $\rho: \mathcal{Z} \rightarrow \mathbb{R}$ . The prospect space  $\mathcal{Z}$  is the space of  $p$ -integrable random variables  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ , with  $p \in [1, \infty)$ .

A *coherent measure of risk* is a function  $\rho: \mathcal{Z} \rightarrow \mathbb{R}$  satisfying the following axioms.

**Convexity.**  $\rho(\alpha Z + (1 - \alpha)V) \leq \alpha\rho(Z) + (1 - \alpha)\rho(V)$ , for all  $Z, V \in \mathcal{Z}$  and all  $\alpha \in [0, 1]$ .

**Monotonicity.** If  $V, Z \in \mathcal{Z}$  and  $V \leq Z$ , then  $\rho(V) \leq \rho(Z)$ .

**Translation Equivariance.** If  $c \in \mathbb{R}$  and  $Z \in \mathcal{Z}$ , then  $\rho(c\mathbb{1} + Z) = c + \rho(Z)$ .

**Positive Homogeneity.** If  $\gamma \geq 0$  and  $Z \in \mathcal{Z}$ , then  $\rho(\gamma Z) = \gamma\rho(Z)$ .

Inequalities are understood component-wise in all of these conditions. The theory of coherent measures of risk was initiated in Artzner et al. [2]. Interestingly, an earlier work, Kijima and Ohnishi [32], missed the crucial monotonicity axiom. A related theory of *deviation measures* also puts less stress on monotonicity (Krokhmal et al. [37], Rockafellar et al. [60]).

A functional satisfying the first three conditions is called a *convex measure of risk* (Föllmer and Schied [25]).

The key condition is the axiom of convexity. It is the same as the convexity property derived in (2) and (7) for utility theories under risk aversion. It means that combining risky positions does not increase risk, but may decrease it, due to diversification (hedging). The monotonicity axiom is obvious and the same as in the utility theories. If we interpret  $\rho(Z)$  as the certainty equivalent of  $Z$ , then the translation axiom is clear: all sure losses contribute directly to the certainty equivalent. The positive homogeneity axiom means that by combining identical risky positions we cannot achieve diversification. Also, change of the system of units (e.g., from dollars to pesos) cannot affect the real value of the certainty equivalent.

The translation and homogeneity axiom indeed imply that  $\rho(Z)\mathbb{1} \sim Z$ ; that is,  $\rho(Z)$  is a certainty equivalent. Indeed, using homogeneity and translation, we obtain

$$\rho(\rho(Z)\mathbb{1}) = \rho(Z)\rho(\mathbb{1}) = \rho(Z)\rho(\mathbb{1} + 0) = \rho(Z)(1 + \rho(0)) = \rho(Z).$$

One can also directly verify that the dual utility functional (4) with a convex rank-dependent utility function is a special case of a coherent measure of risk. However, the von Neumann–Morgenstern utility functionals (1) are not coherent measures of risk, unless the function  $u(\cdot)$  is the identity, in which case (1) reduces to the expected value.

**2.3.2. Dual Representation.** Coherent measures of risk have many remarkable properties. The most fundamental is the *dual representation*. If  $p \in [1, \infty)$  and  $\rho(\cdot)$  is finite valued on  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ , then a convex closed set  $\mathcal{A}$  of probability measures on  $(\Omega, \mathcal{F})$  exists such that

$$\rho(Z) = \max_{\mu \in \mathcal{A}} \mathbb{E}_\mu[Z] = \max_{\mu \in \mathcal{A}} \int_{\Omega} Z(\omega) \mu(d\omega). \quad (9)$$

Moreover, the measures  $\mu \in \mathcal{A}$  have  $q$ -integrable densities  $d\mu/dP$  with respect to  $P$ , where  $1/p + 1/q = 1$ . The set  $\mathcal{A}$  is the subdifferential of  $\rho(\cdot)$  at 0. Representation (9) is a special case of conjugate duality in convex analysis (Rockafellar [57]). In the risk measure context, it was initially proved in a finite-dimensional case in Artzner et al. [2] and was later refined in several papers (see, e.g., Frittelli and Rosazza Gianin [27], Ruszczynski and Shapiro [67]).

Representation (9) is the foundation of many numerical methods for optimizing measures of risk.

**2.3.3. Examples.** Let us consider several examples of measures of risk.

1. The *mean-semideviation* risk measure is defined following Ogryczak and Ruszczyński [46, 47]:

$$\rho(Z) = \mathbb{E}[Z] + \varkappa \|(Z - \mathbb{E}[Z])_+\|_p, \quad \varkappa \in [0, 1]. \tag{10}$$

We customary write  $Z - c$  for  $Z - c\mathbb{1}$ ; it never leads to misunderstanding. It is a coherent measure of risk, and its subdifferential appearing in the dual representation (9) has the following form (Ruszczyński and Shapiro [67]):

$$\mathcal{A} = \left\{ \mu: \frac{d\mu}{dP} = 1 + h - \int_{\Omega} h(\omega) dP, \|h\|_q \leq \varkappa, h \geq 0 \right\}.$$

2. The *average (conditional) value at risk (AVaR)* at level  $\alpha \in (0, 1)$  is defined as follows (Rockafellar and Uryasev [58, 59]):

$$\text{AVaR}_{\alpha}(Z) = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha} \mathbb{E}[(Z - t)_+] \right\} = \frac{1}{\alpha} \int_{1-\alpha}^1 F_Z^{-1}(\beta) d\beta. \tag{11}$$

The equivalence of both formulations can be verified by conjugate duality (Ogryczak and Ruszczyński [47, Theorem 3.1]). The last formula in (11) shows that the average value at risk is also a dual utility functional of form (4), with the following convex rank-dependent utility function:

$$w(\beta) = \begin{cases} 0 & \text{if } 0 \leq \beta \leq 1 - \alpha, \\ \frac{p + \alpha - 1}{\alpha} & \text{if } 1 - \alpha \leq \beta \leq 1. \end{cases}$$

Average value at risk is a coherent measure of risk, and its subdifferential appearing in the dual representation (9) has the following form (Ruszczyński and Shapiro [67]):

$$\mathcal{A} = \left\{ \mu: 0 \leq \frac{d\mu}{dP} \leq \frac{1}{\alpha}, \mu(\Omega) = 1 \right\}.$$

3. Let  $\kappa > 0$  and let  $\mathcal{Z}$  be the space of random variables  $Z$  such that  $\mathbb{E}[e^{\kappa|Z|}] < \infty$ . On the space  $\mathcal{Z}$ , we define the *entropic measure of risk* as follows:

$$\rho(Z) = \frac{1}{\kappa} \ln(\mathbb{E}[e^{\kappa Z}]).$$

It is a convex measure of risk (see Föllmer and Schied [26, §§3.2, 4.9]).

4. Suppose  $u: \mathbb{R} \rightarrow \mathbb{R}$  is nondecreasing and convex, and  $1 \in \partial u(0)$ . We assume that  $|u(t)| \leq Ct^p$  for all  $t \in \mathbb{R}$ , with some  $C > 0$  and  $p \geq 1$ , and we consider the prospect space  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ . The *optimized certainty equivalent*,

$$\rho(Z) = \inf_{t \in \mathbb{R}} \{ t + \mathbb{E}[u(Z - t)] \},$$

is a convex measure of risk (Ben-Tal and Teboulle [5]). The functional  $\varrho(\cdot)$  is not positively homogeneous, unless  $u(\cdot)$  is positively homogeneous. We see that the average value at risk (11) is a special case of the optimized certainty equivalent, with  $u(z) = (1/\alpha)(z)_+$ . Higher-order analogs, with  $u(z) = (1/\alpha)\|(z)_+\|_p^p$ , were considered by Krokmal [36].

**2.3.4. Kusuoka Representation.** Another important property of coherent measures of risk is the Kusuoka representation of law-invariant measures, first introduced in Kusuoka [39]. A measure of risk  $\rho(\cdot)$  is called *law-invariant* (distribution-based) if  $Z \stackrel{\mathcal{L}}{\sim} V$  implies that  $\rho(Z) = \rho(V)$ . If the probability space  $(\Omega, \mathcal{F}, P)$  is nonatomic, then every law-invariant measure of risk on  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ ,  $p \in [1, \infty)$ , can be expressed as follows:

$$\rho(Z) = \sup_{\lambda \in \Lambda} \int_{[0,1]} \text{AVaR}_\alpha(Z) \lambda(d\alpha), \quad Z \in \mathcal{Z}, \tag{12}$$

where  $\Lambda$  is a convex set of probability measures on  $(0, 1]$ . On the space  $\mathcal{L}_\infty(\Omega, \mathcal{F}, P)$ , the measures  $\lambda$  may contain an additional atom at 0, with  $\text{AVaR}_0(Z) = \text{ess sup}_{\omega \in \Omega} Z(\omega)$ .

In the special case, when  $\Lambda$  contains only one element, the measure of risk  $\rho(\cdot)$  is called *spectral*. In fact, spectral risk measures are identical with dual utility functionals (4), which are risk averse, that is, have convex rank-dependent utility functions  $w(\cdot)$ . This can be verified by using the integral representation of  $\text{AVaR}_\alpha(Z)$  provided in (11) and changing the order of integration.

For example, the Kusuoka representation of the first order ( $p = 1$ ) mean–semideviation risk measure (10) has the form

$$\Lambda = \{ \lambda \in \mathcal{P}(0, 1]: (\exists \alpha \in (0, 1)) \lambda = \alpha \delta_\alpha + (1 - \alpha) \delta_1 \},$$

where  $\delta_c$  denotes the unit mass at point  $c$  (see Ogryczak and Ruszczyński [48, Lemma 3.6]). For other examples, see Dentcheva et al. [19], Pichler and Shapiro [52], and Pflug and Römisch [50].

Kusuoka representation is a convenient tool for constructing new risk measures, reflecting modeling needs.

## 2.4. Stochastic Orders

Stochastic orders are *partial* orders on the space of distributions of random variables. Therefore, they are also partial orders on the space  $\mathcal{Z}$  of random variables. As partial orders, they do not have numerical representations in form of utility functionals or risk measures, but they do have *multiutility* representations, with the use of families of utility functions or rank-dependent utility functions.

**2.4.1. Direct Forms.** To present the stochastic orders in the case when smaller outcomes are preferred, it is convenient to consider the *survival function*

$$\bar{F}_Z(\eta) = 1 - F_Z(\eta) = P[Z > \eta].$$

A random variable  $Z$  is *stochastically smaller* than a random variable  $V$  in the first order, which we write as  $Z \preceq_{(1)} V$ , if

$$\bar{F}_Z(\eta) \leq \bar{F}_V(\eta) \quad \text{for all } \eta \in \mathbb{R}.$$

This relation is also known as the *first order stochastic dominance* (Lehmann [40], Quirk and Saposnik [54]).

The second order survival function  $\bar{F}_Z^{(2)}$  is given by the integrals

$$\bar{F}_Z^{(2)}(\eta) = \int_\eta^\infty \bar{F}_Z(\xi) d\xi, \quad \eta \in \mathbb{R},$$

and defines the weak relation of the *increasing convex order*, corresponding to the second-order stochastic dominance (Hadar and Russell [29], Rothschild and Stiglitz [61]); that is,

an integrable random variable  $Z$  is *stochastically smaller* than  $V$  in the increasing convex order, denoted  $V \preceq_{(\text{icx})} S$ , if

$$\bar{F}_Z^{(2)}(\eta) \leq \bar{F}_V^{(2)}(\eta) \quad \text{for all } \eta \in \mathbb{R}.$$

We can express the function  $\bar{F}_Z^{(2)}(\eta)$  as the expected excess: for each target value  $\eta$ , we have

$$\bar{F}_Z^{(2)}(\eta) = \mathbb{E}[(Z - \eta)_+].$$

The function  $\bar{F}_Z^{(2)}(\cdot)$  is convex, nonnegative, and nonincreasing. It is well defined for all random variables  $Z$  with finite expected value. Because of this representation, the second order relation  $Z \preceq_{(\text{icx})} V$  can be equivalently characterized by the following infinite system of inequalities:

$$\mathbb{E}[(Z - \eta)_+] \leq \mathbb{E}[(V - \eta)_+] \quad \text{for all } \eta \in \mathbb{R}. \tag{13}$$

An equivalent characterization of the stochastic order relations using utility functions as in the expected utility theory of von Neumann and Morgenstern (see, e.g., Müller and Stoyan [43]) is also available:

- (i) For any two random variables  $Z, V$ , the relation  $Z \preceq_{(1)} V$  holds if and only if

$$\mathbb{E}[u(Z)] \leq \mathbb{E}[u(V)] \tag{14}$$

for all nondecreasing functions  $u(\cdot)$  defined on  $\mathbb{R}$ .

- (ii) For any two random variables  $Z, V$  with finite expectations, the relation  $Z \preceq_{(\text{icx})} V$  holds if and only if (14) is satisfied for all nondecreasing convex functions  $u(\cdot)$ .

It follows from (ii) that the second order relation (the increasing convex order) is suitable for modeling risk-averse preferences, because it has multiutility representation with risk-averse utility functionals.

**2.4.2. Inverse Forms.** The inverse stochastic dominance relations compare the inverse distribution functions (quantile functions) of random variables. First order stochastic dominance can be characterized equivalently as follows:

$$Z \preceq_{(1)} V \iff F_Z^{-1}(p) \leq F_V^{-1}(p) \quad \text{for all } 0 < p < 1.$$

The second order relation  $\preceq_{(\text{icx})}$  (the increasing convex order) is characterized by relations between integrated quantile functions, which are nothing else but the average value at risk given by formula (11):

$$Z \preceq_{(\text{icx})} V \iff \text{AVaR}_\alpha(Z) \leq \text{AVaR}_\alpha(V) \quad \text{for all } 0 \leq \alpha < 1. \tag{15}$$

Following Dentcheva and Ruszczyński [16], we provide an equivalent characterization of stochastic dominance and the increasing convex order using rank-dependent utility functions, which mirrors the expected utility characterization:

- (i) For any two random variables  $Z, V$ , the relation  $Z \preceq_{(1)} V$  holds if and only if

$$\int_0^1 F_Z^{-1}(p) dw(p) \leq \int_0^1 F_V^{-1}(p) dw(p) \tag{16}$$

for all nondecreasing functions  $w(\cdot)$  defined on  $[0, 1]$ .

- (ii) For any two random variables  $Z, V$  with finite expectations, the relation  $Z \preceq_{(\text{icx})} V$  holds if and only if (16) is satisfied for all nondecreasing convex functions  $w(\cdot)$ .

Again, risk-averse preferences are associated with the increasing convex order, because it has a multiutility representation with risk-averse dual utility functionals of form (4). It is also worth noting that the utility functionals of von Neumann and Morgenstern and the dual utility functionals appear in *equivalent* multiutility representations (14) and (16) of stochastic orders.

**2.4.3. Strassen Theorem.** Other useful characterizations of the usual stochastic order and the increasing convex order can be obtained by means of rearrangements of random variables and possible changing of the probability space. We can prove the following results:

(i) The random variable  $Z$  is stochastically smaller than  $V$  ( $Z \preceq_{(1)} V$ ) if random variables  $Z' \mathcal{R} Z$  and  $V' \mathcal{R} V$  exist, such that  $Z' \leq V'$  a.s.

(ii) The random variable  $Z$  is smaller than  $V$  in the increasing convex order ( $Z \preceq_{(\text{icx})} V$ ) if random variables  $Z' \mathcal{R} Z$  and  $V' \mathcal{R} V$  exist, such that  $Z' \leq \mathbb{E}[V' | Z']$ .

The first statement follows by setting  $Z' = F_Z^{-1}(U)$  and  $V' = F_V^{-1}(U)$  for a uniform random variable  $U$ . The second assertion is due to Strassen [72]; for extensions and applications, see Müller and Stoyan [43] and the references therein.

These results can be used in computationally convenient representations of stochastic orders of the first and the second order (Luedtke [41]). Suppose  $Z$  has  $n$  realizations  $z_1, \dots, z_n$  with probabilities  $p_1, \dots, p_n$ , whereas  $V$  has  $m$  realizations  $v_1, \dots, v_m$  with probabilities  $q_1, \dots, q_m$ . Then  $Z \preceq_{(\text{icx})} V$  if and only if a *nonnegative* matrix  $\Pi$  of dimension  $n \times m$  exists, such that

$$\sum_{j=1}^m \pi_{ij} v_j \geq z_i, \quad i = 1, \dots, n, \tag{17}$$

$$\sum_{j=1}^m \pi_{ij} = 1, \quad i = 1, \dots, n, \tag{18}$$

$$\sum_{i=1}^n p_i \pi_{ij} = q_j, \quad j = 1, \dots, m. \tag{19}$$

To characterize the first order relation, we assume that the realizations of  $V$  are ordered; that is,  $v_1 < v_2 < \dots < v_m$ . It is sufficient for applications that we have in mind, because  $V$  will always be a fixed benchmark random variable. Then  $Z \preceq_{(1)} V$  is equivalent to the existence of a *binary* matrix  $\Pi$  such that Equations (17) and (18) and the following system of inequalities are satisfied:

$$\sum_{j=k}^m \sum_{i=1}^n p_i \pi_{ij} \leq \sum_{j=k}^m q_j, \quad k = 1, \dots, m. \tag{20}$$

For the proof and applications, see Luedtke [41].

**2.4.4. Convexification.** On a nonatomic probability spaces  $(\Omega, \mathcal{F}, P)$  or on finite spaces  $\Omega$  with uniform discrete measures, the following convexification property is true (see Dentcheva and Ruszczynski [15, 13]):

$$\{Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P): Z \preceq_{(\text{icx})} V\} = \overline{\text{conv}}\{Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P): Z \preceq_{(1)} V\}. \tag{21}$$

In the case of finite probability space with uniform measure, the closure operation may be omitted. This representation is of importance for optimization models, because it allows us to construct tight convex approximations of sets defined by the first order stochastic dominance relations.

### 3. Risk-Averse Optimization

Let us now consider static optimization models, in which the “cost”  $Z(x, \omega)$  is random and depends on our decision variables  $x$  in the space  $\mathcal{X} = \mathbb{R}^n$ , as well as on an elementary event  $\omega$  in some probability space  $(\Omega, \mathcal{F}, P)$ . In the rest of this section, we shall write  $Z(x)$  for the mapping from the decision space  $\mathcal{X}$  to an appropriate space of random variables  $\mathcal{Z}$ .

### 3.1. Main Models

The problem of “minimizing”  $Z(x)$  with respect to  $x \in X$ , where  $X \subseteq \mathcal{X}$  is a feasible set, does not make much sense, because  $Z(x)$  is random; for some elementary event  $\omega$ , one decision may be better than another, whereas for a different elementary event  $\omega'$ , their relation may be reversed.

Two classes of models deal with this difficulty:

(i) A numerical representation of risk-averse preferences is used as a new objective function.

(ii) The random outcome  $Z(x)$  is related to another random outcome in constraints.

In the first group we have models of optimizing expected utility functionals,

$$\min_{x \in X} \mathbb{E}[u(Z(x))], \tag{22}$$

rank-dependent utility functionals,

$$\min_{x \in X} \int_0^1 F_{Z(x)}^{-1}(\beta) dw(\beta), \tag{23}$$

or coherent risk measures,

$$\min_{x \in X} \rho(Z(x)). \tag{24}$$

Model (22) is a standard problem of optimizing the expected value, with the outcomes transformed in a nonlinear way. Problems of this type are well investigated in stochastic programming (see the monographs Birge and Louveaux [6], Kall and Mayer [31], Shapiro et al. [70], and references therein). The model (23) is a special case of (24); therefore, we shall focus mainly on optimization models using measures of risk.

Models of group (ii) are problems with chance constraints and problems with stochastic orders in constraints. In these models, the objective function may be just the expected value; risk aversion is represented by a properly formulated system of constraints. A typical optimization problem with a stochastic order constraint has the following form:

$$\begin{aligned} \min \quad & \mathbb{E}[Z(x)] \\ \text{s.t.} \quad & Z(x) \preceq_* Y, \\ & x \in X, \end{aligned} \tag{25}$$

with  $\preceq_*$  representing an appropriate stochastic order and  $Y$  being a *benchmark* random variable. This was first introduced in Dentcheva and Ruszczyński [12] and analyzed and extended in Dentcheva and Ruszczyński [14, 15]. It is a generalization of models with chance constraints and integrated chance constraints, which have been well investigated in stochastic programming (see Klein Haneveld [33], Prékopa [53], Shapiro et al. [70], and references therein).

### 3.2. Optimization of Risk Measures

Let us consider model (24) with a finite-valued coherent measure of risk  $\rho(\cdot)$  defined on a space  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ ,  $p \in [1, \infty)$ . If the mapping  $x \mapsto Z(x, \omega)$  is convex for almost all  $\omega \in \Omega$ , and  $Z(x) \in \mathcal{Z}$  for all  $x \in \mathcal{X}$ , then the composition  $\phi(\cdot) = \rho(Z(\cdot))$  is a convex function. If, additionally, the feasible set  $X$  is convex, problem (24) is convex optimization problem. Because coherent measures of risk are nondifferentiable, in general, the theory and methods of nonsmooth optimization are germane for risk-averse models.

**3.2.1. Optimality and Duality.** Suppose a point  $\hat{x}$  is an optimal solution of problem (24). Then a measure  $\hat{\mu} \in \partial\rho(Z(\hat{x}))$  exists such that

$$0 \in \int_{\Omega} \partial_x Z(\hat{x}, \omega) \hat{\mu}(d\omega) + N_X(\hat{x}),$$

with  $N_X(\hat{x})$  denoting the normal cone to  $X$  at  $\hat{x}$ . The integral above is understood as a collection of integrals of measurable selections of the multifunction  $\omega \rightrightarrows \partial_x Z(\hat{x}, \omega)$ . Moreover, the pair  $(\hat{x}, \hat{\mu})$  is a *saddle point* of the function

$$\langle Z(x), \mu \rangle = \int_{\Omega} Z(x, \omega) \mu(d\omega)$$

on the set  $X \times \mathcal{A}$ ; that is,

$$\langle Z(\hat{x}), \mu \rangle \leq \langle Z(\hat{x}), \hat{\mu} \rangle \leq \langle Z(x), \hat{\mu} \rangle \quad \text{for all } (x, \mu) \in X \times \mathcal{A}.$$

This implies that the following *duality relation* is satisfied:

$$\rho(Z(\hat{x})) = \min_{x \in X} \rho(Z(x)) = \min_{x \in X} \max_{\mu \in \mathcal{A}} \langle Z(x), \mu \rangle = \max_{\mu \in \mathcal{A}} \min_{x \in X} \langle Z(x), \mu \rangle = \min_{x \in X} \langle Z(x), \hat{\mu} \rangle. \quad (26)$$

This relation is the basis of efficient cutting plane methods for minimizing risk measures.

**3.2.2. Risk-Averse Multicut Method.** To illustrate the application of duality (26), we consider the problem (24) with a finite probability space  $\Omega = \{\omega_1, \omega_2, \dots, \omega_S\}$  with  $p_s = P[\omega_s] > 0$  and  $\sum_{s=1}^S p_s = 1$ . Our presentation follows Miller and Ruszczyński [42], where the method was proposed. We denote

$$Z_s(x) = Z(x, \omega_s), \quad s = 1, \dots, S. \quad (27)$$

Suppose we collected values  $U_s^{(l)} = Z_s(x^{(l)})$  and subgradients  $g_s^{(l)} \in \partial Z_s(x^{(l)})$  at some trial points  $x^{(l)}$  in iterations  $l = 1, \dots, k - 1$ . They can be used to construct lower cutting plane approximations of the functions  $Z_s(x)$  as follows:

$$Z_s(x) \geq Z_s^{(k)}(x) = \max_{1 \leq l \leq k-1} \{U_s^{(l)} + \langle g_s^{(l)}, x - x^{(l)} \rangle\}. \quad (28)$$

The random function  $Z^{(k)}(x)$  with realizations described above is a lower approximation of  $Z(x)$ . In particular, the random vector  $W^{(l)}$  with realizations

$$W_s^{(l)} = Z_s^{(l)}(x^{(l)}), \quad s = 1, \dots, S,$$

is a lower approximation of the true scenario values (27) at  $x^{(l)}$ . We evaluate a subgradient  $\mu^{(l)}$  of the risk measure  $\rho(\cdot)$  at  $W_s^{(l)}$  and write the chain of inequalities:

$$\rho(Z(x)) \geq \rho(Z^{(k)}(x)) \geq \max_{1 \leq l \leq k-1} \langle Z^{(k)}(x), \mu^{(l)} \rangle. \quad (29)$$

The polyhedral model (29) of the composition  $\rho(Z(x))$  is minimized, and the next trial point  $x^{(k)}$  is obtained. The whole algorithm is described in Figure 1. To avoid unboundedness at initial steps, we assume that we know a lower bound  $\eta_{\min}$  on the optimal value of problem (26) and a lower bound  $w_{\min}$  for all scenario values (27).

If the set  $X$  is compact, the method is convergent to the optimal solution of problem (24). The proof is identical to the proof of the standard cutting plane method for convex nonsmooth optimization (Ruszczynski [64, Theorem 7.7]).

Additional complications arise when the functions (27) have domains that do not include the feasible set  $X$ . This is typical for applications to two-stage problems, when each  $Z_s(x)$  is an optimal value of the second-stage problem in scenario  $s$  (see, e.g., Shapiro et al. [70, Chapter 2]). In that case, the polyhedral models (28) of these functions have to encompass *feasibility cuts*, similarly to decomposition methods for risk-neutral two-stage problems (Ruszczynski [63]). The detailed description of the risk-averse multicut method for two-stage problem can be found in Miller and Ruszczyński [42]. A related approach is discussed in Ahmed [1].

FIGURE 1. The risk-averse multicut method.

Step 0. Set  $k = 1$ .  
 Step 1. Find the solution  $(x^{(k)}, \eta^{(k)}, w^{(k)})$  of the *master problem*

$$\begin{aligned} \min_{x, \eta, w} \quad & \eta \\ \text{s.t.} \quad & \eta \geq \langle W, \mu^{(l)} \rangle, \quad l = 1, \dots, k-1, \\ & W_s \geq U_s^{(l)} + \langle g_s^{(l)}, x - x^{(l)} \rangle, \quad l = 1, \dots, k-1, \quad s = 1, \dots, S, \\ & x \in X, \quad \eta \geq \eta_{\min}, \quad W \geq w_{\min}. \end{aligned}$$

Step 2. For  $s = 1, \dots, S$ , calculate  $U_s^{(k)} = Z_s(x^{(k)})$  and  $g_s^{(k)} \in \partial Z_s(x^{(k)})$ .  
 Step 3. Find  $\rho^{(k)} = \max_{\mu \in \mathcal{A}} \langle \mu, W^{(k)} \rangle$  and denote the maximizer by  $\mu^{(k)}$ .  
 Step 4. If  $\rho^{(k)} = \eta^{(k)}$ , then stop. Otherwise, increase  $k$  by 1 and go to Step 1.

### 3.3. Optimization with Stochastic Order Constraints

We now consider the risk-averse model (25) with the usual stochastic order  $\preceq_{(1)}$  and with the increasing convex order  $\preceq_{(\text{icx})}$ . We focus on the latter, because it is most convenient for modeling risk-averse preferences. Models with first order constraints are nonconvex, in general, and frequently involve combinatorial considerations (Dentcheva and Ruszczyński [15], Luedtke [41], Noyan et al. [45], Noyan and Ruszczyński [44]). Also, owing to (21), models with first order constraints can be very well approximated by models with second order constraints.

#### 3.3.1. Increasing Convex Order Constraints.

Consider the problem

$$\begin{aligned} \min \quad & \mathbb{E}[Z(x)] \\ \text{s.t.} \quad & Z(x) \preceq_{(\text{icx})} Y, \\ & x \in X, \end{aligned} \tag{30}$$

in which the random function  $Z(\cdot)$  is convex (a.s.), and the set  $X$  is convex and closed. This model is a “cost” version of the model with second order stochastic dominance constraints of Dentcheva and Ruszczyński [12, 14].

Under suitable regularity conditions, we can prove that  $\hat{x}$  is an optimal solution of this problem, if and only if a convex nondecreasing utility function  $\hat{u}(\cdot)$  exists, such that  $\hat{x}$  is also a solution of the following *expected utility problem*:

$$\min_{x \in X} \mathbb{E}[Z(x) + \hat{u}(Z(X)) - \hat{u}(Y)]. \tag{31}$$

The function above plays the role of a Lagrangian, with the utility function  $\hat{u}(\cdot)$  playing the role of the Lagrange multiplier. Corresponding duality relations can be developed as well (see Dentcheva and Ruszczyński [12, 14]).

In a parallel analysis of Dentcheva and Ruszczyński [16], we also show that  $\hat{x}$  is an optimal solution of problem (30), if and only if a convex nondecreasing rank-dependent utility function  $\hat{w}(\cdot)$  exists, such that  $\hat{x}$  is also a solution of the following *rank-dependent utility problem*:

$$\min_{x \in X} \int_0^1 F_{Z(x)}^{-1}(\beta) d(1 + \hat{w})(\beta) - \int_0^1 F_Y^{-1}(\beta) d\hat{w}(\beta). \tag{32}$$

Again, the function in the problem above plays the role similar to that of a Lagrangian, and the rank-dependent utility function  $\hat{w}(\cdot)$  being the analogue of a Lagrange multiplier. Duality theory based on this “Lagrangian” can be developed as well.

Both utility theories, the expected utility theory of von Neumann and Morgenstern and the dual utility theory of Yaari, provide dual objects for stochastic order constraints. For problems with first order constraints, local optimality conditions can be derived, with

the use of the Lagrangian functions (31) and (32), but with utility functions  $u(\cdot)$  and rank-dependent utility functions  $w(\cdot)$ , which are not necessarily convex (Dentcheva and Ruszczyński [15, 16]).

**3.3.2. Nonlinear Optimization Reformulations.** Suppose  $Z(x)$  has finitely many realizations  $Z_i(x)$  attained with probabilities  $p_i, i = 1, \dots, n$ , and  $Y$  has finitely many realizations  $y_j$  attained with probabilities  $q_j, j = 1, \dots, m$ . Because  $V$  is fixed, we may assume that its realizations are arranged in such a way that  $y_1 < y_2 < \dots < y_m$ . In this case, the relation  $Z \preceq_{(\text{icx})} Y$  is equivalent to a finite number of inequalities (Dentcheva and Ruszczyński [12]):

$$\mathbb{E}[(Z(x) - y_j)_+] \leq \mathbb{E}[(Y - y_j)_+] = \sum_{\nu=j+1}^m q_\nu(y_\nu - y_j), \quad j = 1, \dots, m. \tag{33}$$

This allows us to reformulate problem (30) as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^n p_i Z_i(x) \\ \text{s.t.} \quad & \sum_{i=1}^n p_i s_{ij} \leq \sum_{\nu=j+1}^m q_\nu(y_\nu - y_j), \quad j = 1, \dots, m, \\ & Z_i(x) - s_{ij} \leq y_j, \quad i = 1, \dots, n, j = 1, \dots, m, \\ & s_{ij} \geq 0, \quad i = 1, \dots, n, j = 1, \dots, m, \\ & x \in X. \end{aligned} \tag{34}$$

The difficulty of this formulation is that it requires  $n \times m$  new variables  $s_{ij}$  to represent  $(Z_i(x) - y_j)_+$ .

Another reformulation is based on the relations (17)–(19):

$$\begin{aligned} \min \quad & \sum_{i=1}^n p_i Z_i(x) \\ \text{s.t.} \quad & Z_i(x) \leq \sum_{j=1}^m \pi_{ij} y_j, \quad i = 1, \dots, n, \\ & \sum_{j=1}^m \pi_{ij} = 1, \quad i = 1, \dots, n, \\ & \sum_{i=1}^n p_i \pi_{ij} = q_j, \quad j = 1, \dots, m, \\ & \pi_{ij} \geq 0, \quad i = 1, \dots, n, j = 1, \dots, m, \\ & x \in X. \end{aligned} \tag{35}$$

Again,  $n \times m$  variables  $\pi_{ij}$  are introduced in this formulation.

**3.3.3. Event Cut Formulations.** A significant improvement can be achieved by representing constraints (33) by the following large system of inequalities (see Klein Haneveld and van der Vlerk [34], Rudolf and Ruszczyński [62]):

$$\sum_{i \in I} p_i (Z_i(x) - y_j) \leq \mathbb{E}[(Y - y_j)_+] = \sum_{\nu=j+1}^m q_\nu(y_\nu - y_j), \quad j = 1, \dots, m, \quad \forall I \subseteq \{1, \dots, n\}. \tag{36}$$

Downloaded from informs.org by [216.73.216.242] on 05 June 2026, at 08:28. For personal use only, all rights reserved.

The inequality (36) for a specific set  $I$  is called a *direct event cut*. The idea is not use these cuts all at once, but rather to use a small subset of them. In iteration  $k$  of the method, we solve the following approximate problem, involving only a subset of event cuts:

$$\begin{aligned} \min \quad & \sum_{i=1}^n p_i Z_i(x) \\ \text{s.t.} \quad & \sum_{i \in I^{(l)}} p_i (Z_i(x) - y_{j^{(l)}}) \leq \sum_{k=j^{(l)}+1}^m q_k (y_k - y_{j^{(l)}}), \quad l = 1, \dots, k-1, \\ & x \in X. \end{aligned} \tag{37}$$

After the solution  $x^{(k)}$  is found, we check the inequalities (33) at  $x^{(k)}$ . If they are all satisfied, the method stops. Otherwise, we choose  $y_{j^{(k)}}$  to be the realization  $y_j$  of  $Y$  for which (33) is violated, set

$$I^{(k)} = \{i: Z_i(x^{(k)}) > y_{j^{(k)}}\},$$

add a new event cut for  $I^{(k)}$ , increase  $k$  by 1, and continue.

An alternative and even more efficient approach can be obtained from representation (15). In the case of discrete distributions, the relation  $Z \preceq_{(\text{icx})} Y$  is equivalent to the following system (Dentcheva and Ruszczyński [17], Dentcheva and Martinez [11]):

$$\sum_{i \in I} p_i Z_i(x) \leq \int_{1-\sum_{i \in I} p_i}^1 F_Y^{-1}(\beta) d\beta \quad \forall I \subseteq \{1, \dots, n\}. \tag{38}$$

Observe that  $F_Y^{-1}(\cdot)$  is piecewise constant, and the right-hand side of (38) can be easily calculated for any set  $I$ . We call inequality (36) for a specific set  $I$  an *inverse event cut*.

Using this characterization, we may proceed in a way analogous to (37). At iteration  $k$  we solve the following relaxed problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^n p_i Z_i(x) \\ \text{s.t.} \quad & \sum_{i \in I^{(l)}} p_i Z_i(x) \leq \int_{1-\sum_{i \in I^{(l)}} p_i}^1 F_Y^{-1}(\beta) d\beta, \quad l = 1, \dots, k-1, \\ & x \in X. \end{aligned} \tag{39}$$

After the solution  $x^{(k)}$  is found, we check the inequalities (38) at  $x^{(k)}$ . If they are all satisfied, the method stops. Otherwise, we choose an event  $I^{(k)}$  for which (33) is violated, add a new inverse event cut for  $I^{(k)}$ , increase  $k$  by 1, and continue. Convergence of this method for general distributions is proved in Dentcheva and Ruszczyński [17]. A related approach by Fábíán et al. [22], which is based on a similar idea for AVaR constraints of Künzi-Bay and Mayer [38], requires  $m = n$  and equal probabilities  $p_i = 1/n$  of the scenarios.

### 3.4. Multistage Risk-Averse Optimization

**3.4.1. Problem Formulation.** We start from the formulation of a *linear* multistage stochastic optimization problem; a convex problem may be modeled in a similar way, but with additional complications.

Let  $(\Omega, \mathcal{F}, P)$  be a probability space with a sigma-algebra  $\mathcal{F}$  and probability measure  $P$ . Consider a filtration  $\{\emptyset, \Omega\} = \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_T = \mathcal{F}$ . A random vector  $x = (x_1, \dots, x_T)$ , where each  $x_t$  has values in  $\mathbb{R}^{n_t}$ ,  $t = 1, \dots, T$ , is called a *policy*. If each  $x_t$  is  $\mathcal{F}_t$ -measurable,



For a dynamic risk measure  $\{\rho_{t,T}\}_{t=1}^T$ , we can define *one-step conditional risk measures*  $\rho_t: \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$ ,  $t = 1, \dots, T-1$  as follows:

$$\rho_t(Z_{t+1}) = \varrho_{t,T}(0, Z_{t+1}, 0, \dots, 0).$$

We can derive the following structure of a time-consistent dynamic risk measure. If a dynamic risk measure  $\{\rho_{t,T}\}_{t=1}^T$ , satisfies, for all  $Z \in \mathcal{Z}$  and all  $t = 1, \dots, T$ , the conditions

$$\varrho_{t,T}(Z_t, Z_{t+1}, \dots, Z_T) = Z_t + \varrho_{t,T}(0, Z_{t+1}, \dots, Z_T), \tag{46}$$

$$\varrho_{t,T}(0, \dots, 0) = 0, \tag{47}$$

then it is time consistent if and only if for all  $1 \leq t \leq T$  and all  $Z \in \mathcal{Z}_{1,T}$  the following identity is true:

$$\varrho_{t,T}(Z_t, \dots, Z_T) = Z_t + \rho_t(Z_{t+1} + \rho_{t+1}(Z_{t+2} + \dots + \rho_{T-2}(Z_{T-1} + \rho_{T-1}(Z_T)) \dots)). \tag{48}$$

Condition (46) is a form of the *translation property*, discussed in various settings in Artzner et al. [3], Frittelli and Scandolo [28], and Pflug and Römisch [50]. Our version is weaker, because  $Z_t$  is  $\mathcal{F}_t$ -measurable.

It follows that a time-consistent dynamic risk measure is completely defined by one-step conditional risk measures  $\rho_t$ ,  $t = 1, \dots, T-1$ . For  $t = 1$ , formula (48) defines a risk measure of the *entire* sequence  $Z \in \mathcal{Z}_{1,T}$  (with a deterministic  $Z_1$ ).

**3.4.3. Coherent Conditional Measures of Risk.** Further restrictions are obtained by assuming the property of *coherency*, which extends to the conditional case the definition of a coherent measure of risk from §2.3. Let  $1 \leq t \leq T-1$ . A *coherent one-step conditional risk measure* is a function  $\rho_t: \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$  satisfying the following axioms.

**Convexity.**  $\rho_t(\alpha Z_{t+1} + (1-\alpha)V_{t+1}) \leq \alpha\rho_t(Z_{t+1}) + (1-\alpha)\rho_t(V_{t+1})$  for all  $Z_{t+1}, V_{t+1} \in \mathcal{Z}_{t+1}$  and all  $\alpha \in [0, 1]$ .

**Monotonicity.** If  $V_{t+1}, Z_{t+1} \in \mathcal{Z}_{t+1}$  and  $V_{t+1} \leq Z_{t+1}$ , then  $\rho_t(V_{t+1}) \leq \rho_t(Z_{t+1})$ .

**Translation Equivariance.** If  $V_t \in \mathcal{Z}_t$  and  $Z_{t+1} \in \mathcal{Z}_{t+1}$ , then  $\rho_t(V_t + Z_{t+1}) = V_t + \rho_t(Z_{t+1})$ .

**Positive Homogeneity.** If  $\gamma \geq 0$  and  $Z_{t+1} \in \mathcal{Z}_{t+1}$ , then  $\rho_t(\gamma Z_{t+1}) = \gamma\rho_t(Z_{t+1})$ .

An example of coherent one-step conditional risk measure is the *conditional mean-upper semideviation* model defined by

$$\rho_t(Z) = \mathbb{E}[Z | \mathcal{F}_t] + \kappa_t \mathbb{E}[(Z - \mathbb{E}[Z | \mathcal{F}_t])_+ | \mathcal{F}_t], \tag{49}$$

with an  $\mathcal{F}_t$ -measurable  $\kappa_t \in [0, 1]$ . See Shapiro et al. [70, p. 277] for the details showing that formula (49) defines a coherent conditional risk measure and for other examples of conditional risk measures.

Using coherent conditional risk measures in (48) leads to a model with a coherent dynamic measure of risk.

**3.4.4. Scenario Trees.** For practical tractability, we assume that all sigma-algebras are finite and all vector spaces are finite-dimensional. Under such assumptions, possible realizations of data form a *scenario tree*. It has nodes organized in levels that correspond to stages  $1, \dots, T$ . At level  $t = 1$  we have only one *root node*  $n = 1$ . Nodes at levels  $t = 2, \dots, T$  correspond to elementary events in  $\mathcal{F}_t$ . Each node  $n$  at level  $t = 2, \dots, T$  is connected to a unique node  $a(n)$  at level  $t-1$ , called the *ancestor node*, which corresponds to the elementary event in  $\mathcal{F}_{t-1}$  that contains the event associated with  $n$ . Thus, every node  $n$  at levels  $t = 1, \dots, T-1$  is connected to a set  $C(n)$  of *children nodes* at level  $t+1$ , which correspond

to the elementary events in  $\mathcal{F}_{t+1}$  included in the event corresponding to  $n$ . We denote by  $\Omega_t$  the set of all nodes at stage  $t = 1, \dots, T$ , and  $\mathbb{N} = \bigcup_{t=1}^T \Omega_t$ . We have the following relations:

$$\Omega_{t+1} = \bigcup_{n \in \Omega_t} C(n), \quad C(n) = \{m \in \Omega_{t+1} : n = a(m)\}.$$

The sets  $C(n)$  are disjoint:  $C(n) \cap C(n') = \emptyset$  if  $n \neq n'$ . A *scenario* is a path  $s$  from the root to a node at the last stage  $T$ . By construction, there is one-to-one correspondence between the scenarios and the set  $\Omega_T = \Omega$ . Let  $\mathcal{S}(n)$  be the set of scenarios passing through node  $n$ . These sets satisfy the recursive relation

$$\begin{aligned} \mathcal{S}(n) &= \{n\}, \quad n \in \Omega_T, \\ \mathcal{S}(n) &= \bigcup_{m \in C(n)} \mathcal{S}(m), \quad n \in \Omega_t, t = T - 1, \dots, 1. \end{aligned}$$

As the nodes of the tree correspond to events defining nested partitions of  $\Omega$ , the measure  $P$  can be specified by conditional probabilities:

$$p_{nm} = P[m | n], \quad n \in \Omega_t, m \in C(n), t = 1, \dots, T - 1.$$

Every node  $n$  at level  $t$  has a *history*: the path  $(n_1, \dots, n_{t-1}, n)$  from the root to  $n$ . The probability of the node  $n$  is thus the product of the corresponding conditional probabilities

$$p_n = p_{n_1 n_2} p_{n_2 n_3} \dots p_{n_{t-1} n}. \tag{50}$$

In particular, when  $t = T$ , formula (50) describes the probability of a scenario  $s \in \Omega_T$ .

For every node  $n \in \Omega_t$ , an  $\mathcal{F}_t$ -measurable random variable  $Z$  has identical values on all scenarios  $s \in \mathcal{S}(n)$ . It can therefore be equivalently represented as a function of a node at level  $t$ . All data of the problem will be now indexed by nodes of the tree, rather than by time.

**3.4.5. Risk Evaluation on a Tree.** Consider a coherent one-step conditional measure of risk  $\rho_t(\cdot)$ . Its value is  $\mathcal{F}_t$ -measurable, and thus we can consider its representation as a function of a node at level  $t$ . It follows from Ruszczyński and Shapiro [66, Theorem 3.2] that the value of  $\rho_t(Z_{t+1})$  at elementary events associated with node  $n$  depends only on the values of  $Z_{t+1}$  at nodes  $m \in C(n)$ . This feature of a conditional risk measure is called the *local property*. With a slight abuse of notation, we denote the vector of these values by  $Z_{C(n)}$ , and we write the conditional risk measure equivalently as  $\rho_n(Z_{C(n)})$ .

Let us define the random variables

$$V_t = \rho_t(Z_{t+1} + \rho_{t+1}(Z_{t+2} + \dots + \rho_{T-1}(Z_T) \dots)), \quad t = 1, \dots, T. \tag{51}$$

They are  $\mathcal{F}_t$ -measurable, and thus we shall index them by node and write  $V_n$  for the value at node  $n$  on level  $t$ . It follows that the value of a time-consistent coherent measure of risk (48) can be written on the scenario tree in a recursive manner:

$$\begin{aligned} V_n &= Z_n, \quad n \in \Omega_T, \\ V_n &= Z_n + \rho_n(V_{C(n)}), \quad n \in \Omega_t, t = T - 1, \dots, 1. \end{aligned} \tag{52}$$

The recursive character of time-consistent measures of risk can be also used to derive optimality conditions in the form of *dynamic programming equations*. Let us consider a node  $n \in \Omega_t$  and assume that the value of  $x_{a(n)}$  is fixed. Owing to the local property, we can consider the subproblem of minimizing  $V_n$  with respect to  $x_n, x_{C(n)}, \dots, x_{C^{T-t}(n)}$ , subject to the corresponding subset of conditions (40) involving these variables. We call the optimal value of this subproblem the *value function* and denote it by the symbol  $\mathcal{Q}_n(x_{a(n)})$ . Except

for the case of  $t = 1$ , it is a function of  $x_{a(n)}$ . Similarly to (52), we use the symbol  $\mathcal{Q}_{C(n)}(x_n)$  to represent a random variable with values  $\mathcal{Q}_m(x_n)$ ,  $m \in C(n)$ , attained with probabilities  $p_{nm}$ .

Proceeding exactly as in Shapiro et al. [70, §6.7.3], we conclude that the value functions  $\mathcal{Q}_n(\cdot)$  satisfy the *dynamic programming equations*

$$\begin{aligned} \mathcal{Q}_n(x_{a(n)}) &= \min_{x_n} \{ \langle c_n, x_n \rangle : B_n x_{a(n)} + A_n x_n = b_n, x_n \in X_n \}, \quad n \in \Omega_T, \\ \mathcal{Q}_n(x_{a(n)}) &= \min_{x_n} \{ \langle c_n, x_n \rangle + \rho_n(\mathcal{Q}_{C(n)}(x_n)) : B_n x_{a(n)} + A_n x_n = b_n, x_n \in X_n \}, \quad (53) \\ & n \in \Omega_t, \quad t = T - 1, \dots, 1, \end{aligned}$$

with the convention that  $\mathcal{Q}_1$  has no arguments. The value of  $\mathcal{Q}_1$  is the optimal value of the problem, and the optimal decisions in problems (53) constitute the optimal policy. The optimal value functions  $\mathcal{Q}_n(\cdot)$  are convex, because their arguments appear as parameters in the constraints of convex optimization problems.

Equations (53) is the theoretical foundation of computational methods for solving multistage time-consistent risk-averse problems.

**3.4.6. Risk-Averse Nested Decomposition.** We now extend the multicut method proposed in Miller and Ruszczyński [42] to multistage problems (see Figure 1). Very similar approaches are associated with the risk-averse extensions of the *stochastic dual decomposition method* of Pereira and Pinto [49], proposed in Kozmik and Morton [35], Philpott and De Matos [51], and Shapiro [69].

In the nested decomposition method, a subproblem is associated with every node of the scenario tree. The subproblems communicate along the arcs of the tree. A typical subproblem at a node  $n \in \Omega_t$  will get the value of  $x_{a(n)}$  from the ancestor node  $a(n)$  and pass its current solution  $x_n$  to its children nodes  $m \in C(n)$ . The subproblems will also pass the values of Lagrange multipliers associated with the constraints to their ancestor nodes and receive the corresponding multipliers from their children nodes.

Consider a node  $n$  at level  $1 < t < T$  and suppose a new value  $\bar{x}_{a(n)}$  has been received from the ancestor node. For simplicity, we assume that feasibility issues do not occur and that the system

$$B_n \bar{x}_{a(n)} + A_n x_n = b_n, \quad x_n \in X_n$$

has always a solution. At iteration  $k$  at node  $n$ , we solve the following *master problem*:

$$\begin{aligned} \min_{x_n, U, W} \quad & \eta \\ \text{s.t.} \quad & \eta \geq \langle c_n, x_n \rangle + \langle W, \mu_n^{(l)} \rangle, \quad l = 1, \dots, k - 1, \\ & W_m \geq U_m^{(l)} + \langle g_m^{(l)}, x_n - \bar{x}_n^{(l)} \rangle, \quad l = 1, \dots, k - 1, m \in C(n), \quad (54) \\ & B_n \bar{x}_{a(n)} + A_n x_n = b_n, \\ & x_n \in X_n, \quad \eta \geq \eta_{\min}, \quad W \geq W_{\min}. \end{aligned}$$

The iterations  $k$  are numbered locally, at each node  $n$ . The constants  $\eta_{\min}$  and  $W_{\min}$  are uniform lower bounds on the value functions in (53).

Denote by  $U_n^{(k)}$  the optimal value of problem (54), by  $\bar{x}_n^{(k)}$  the optimal solution, and by  $\pi$  the vector of Lagrange multipliers associated with the equality constraints. The solution  $\bar{x}_n^{(k)}$ , if different from previously reported, is passed to the children nodes  $m \in C(n)$ . If the value  $U_n^{(k)}$  is strictly larger than the value previously reported, then it is used together with the multipliers  $\pi$  to construct an *objective cut*

$$V_n \geq U_n^{(k)} + \langle g_n, x_{a(n)} - \bar{x}_{a(n)} \rangle,$$

with the subgradient  $g^n$  defined as follows:

$$g_n = -(B^n)^\top \pi. \tag{55}$$

The objective cut is passed to the ancestor problem  $a(n)$ . Moreover, at any iteration  $l$  we also calculate the probability measure

$$\mu_n^{(l)} = \arg \max_{\mu \in \mathcal{A}_n} \langle \mu, W \rangle, \tag{56}$$

where  $\mathcal{A}_n = \partial \rho_n(0)$  is the set featuring in the dual representation (9) of  $\rho_n(\cdot)$ . After all these updates, the master problem (54) is resolved, and the resulting information (if essentially new) is transmitted to the ancestor and children nodes.

If  $n = 1$  (the root node of the tree), there is no ancestor node. Also, the constraint involving  $\bar{x}_{a(n)}$  in (54) has to be omitted. For  $n \in \Omega_T$  problem (54) simplifies as follows:

$$\min_{x_n} \{ \langle c_n, x_n \rangle : B_n \bar{x}_{a(n)} + A_n x_n = b_n, x_n \in X_n \}. \tag{57}$$

The protocol by which the nodes are processed is not essential for convergence of the method, although it may affect its speed.

The proof of convergence is very similar to proofs of convergence of the nested decomposition of expected-value multistage stochastic linear programming problems (see Birge and Louveaux [6], Ruszczyński [63], and the references therein).

If some of the node problems are infeasible, the method may involve *feasibility cuts*, similarly to Miller and Ruszczyński [42].

**3.4.7. Scenario Decomposition.** The dual representation (9) applied to node  $n$  has the form

$$\rho_n(W) = \max_{q_n \in \mathcal{A}_n} \langle W, q_n \rangle, \tag{58}$$

where  $W$  is a vector associated with nodes  $m \in C(n)$ . Observe that by selecting a measure  $q_n \in \mathcal{A}_n$  we select, in fact, conditional probabilities  $q_{nm}$ ,  $m \in C(n)$ . This results in a probability distribution  $q$  on the scenario set. Indeed, every scenario  $s$  is a path of nodes  $n_1, n_2, \dots, n_T$ , and thus

$$q_s = q_{n_1 n_2} q_{n_2 n_3} \cdots q_{n_{T-1} n_T}.$$

The set  $\mathcal{A}_{1, T}$  of probability distributions constructed in this way is convex and closed, and it can be described by a system of inequalities and equations with the use of the analytical description of the sets  $\mathcal{A}_n$ . This leads to the following equivalent representation of a coherent dynamic measure of risk:

$$\rho(Z_1, \dots, Z_T) = \max_{q \in \mathcal{A}_{1, T}} \langle q, Z_1 + Z_2 + \cdots + Z_T \rangle. \tag{59}$$

Owing to this representation, we can rewrite the problem (42) as follows:

$$\min_{x \in I \cap F} \max_{q \in \mathcal{A}_{1, T}} \langle q, \langle c_1, x_1 \rangle + \langle c_2, x_2 \rangle + \cdots + \langle c_T, x_T \rangle \rangle. \tag{60}$$

This result is the theoretical foundation of a scenario decomposition method for solving time-consistent risk-averse optimization problems, proposed in Collado et al. [8]. Interchanging the “min” and “max” operations (which we can do owing to the convexity of the feasible set), we obtain an equivalent formulation

$$\max_{q \in \mathcal{A}_{1, T}} \min_{x \in I \cap F} \langle q, \langle c_1, x_1 \rangle + \langle c_2, x_2 \rangle + \cdots + \langle c_T, x_T \rangle \rangle. \tag{61}$$

Downloaded from informs.org by [216.73.216.242] on 05 June 2026, at 08:28. For personal use only, all rights reserved.

Observe that the inner minimization problems are standard multistage stochastic programming problems, just with a different probability distribution  $q$ . They can be decomposed in any of the standard ways, and are well understood for risk-neutral problems (see Ruszczyński [63] and the references therein), for example, the scenario decomposition method. The outer maximization problem can be solved by a nonsmooth optimization algorithm (see, e.g., Ruszczyński [64, Chapter 7]). One possibility is the bundle method, employed in Collado et al. [8].

## Acknowledgments

This work was supported by National Science Foundation Award [CMMI-0965689] and Air Force Office of Scientific Research Award [FA-9550-11-1-0164].

## References

- [1] S. Ahmed. Convexity and decomposition of mean-risk stochastic programs. *Mathematical Programming* 106(3):433–446, 2006.
- [2] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance* 9(3):203–228, 1999.
- [3] 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* 152:5–22, 2007.
- [4] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhäuser, Boston, 1990.
- [5] A. Ben-Tal and M. Teboulle. An old new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance* 17(3):449–476, 2007.
- [6] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.
- [7] P. Cheridito, F. Delbaen, and M. Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability* 11(3):57–106, 2006.
- [8] 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.
- [9] G. Debreu. Representation of a preference ordering by a numerical function. C. Davis, A. Thrall, and B. Combs, eds. *Decision Processes*. John Wiley & Sons, New York, 159–165, 1954.
- [10] G. Debreu. Continuity properties of Paretian utility. *International Economic Review* 5:285–293, 1964.
- [11] D. Dentcheva and G. Martinez. Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse. *European Journal of Operational Research* 219(1):1–8, 2012.
- [12] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization* 14(2):548–566, 2003.
- [13] D. Dentcheva and A. Ruszczyński. Convexification of stochastic ordering. *Comptes Rendus de l’Academie Bulgare des Sciences* 57:11–16, 2004.
- [14] D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming* 99(2): 329–350, 2004.
- [15] D. Dentcheva and A. Ruszczyński. Semi-infinite probabilistic optimization: First-order stochastic dominance constraints. *Optimization* 53(5–6):583–601, 2004.
- [16] D. Dentcheva and A. Ruszczyński. Inverse stochastic dominance constraints and rank dependent expected utility theory. *Mathematical Programming* 108(2–3):297–311, 2006.
- [17] D. Dentcheva and A. Ruszczyński. Inverse cutting plane methods for optimization problems with second-order stochastic dominance constraints. *Optimization* 59(3–4):323–338, 2010.
- [18] D. Dentcheva and A. Ruszczyński. Common mathematical foundations of expected utility and dual utility theories. *SIAM Journal on Optimization* 23(1):381–405, 2013.

- [19] D. Dentcheva, S. Penev, and A. Ruszczyński. Kusuoka representation of higher order dual risk measures. *Annals of Operations Research* 181(1):325–335, 2010.
- [20] R. M. Dudley. *Real Analysis and Probability, Cambridge Studies in Advanced Mathematics*, Vol. 74. Revised reprint of the 1989 original, Cambridge University Press, Cambridge, UK, 2002.
- [21] S. Eilenberg. Ordered topological spaces. *American Journal of Mathematics* 63:39–45, 1941.
- [22] C. I. Fábián, G. Mitra, and D. Roman. Processing second-order stochastic dominance models using cutting-plane representations. *Mathematical Programming* 130(1):33–57, 2011.
- [23] P. C. Fishburn. *The Foundations of Expected Utility, Theory and Decision Library*, Vol. 31. D. Reidel Publishing, Dordrecht, The Netherlands, 1982.
- [24] H. Föllmer and I. Penner. Convex risk measures and the dynamics of their penalty functions. *Statistics and Decisions* 24(1):61–96, 2006.
- [25] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics* 6(4):429–447, 2002.
- [26] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*, 2nd ed. De Gruyter Studies in Mathematics, Vol. 27. Walter de Gruyter, Berlin, 2004.
- [27] M. Frittelli and E. Rosazza Gianin. Putting order in risk measures. *Journal of Banking and Finance* 26(7):1473–1486, 2002.
- [28] M. M. Frittelli and G. Scandolo. Risk measures and capital requirements for processes. *Mathematical Finance* 16(4):589–612, 2006.
- [29] J. Hadar and W. Russell. Rules for ordering uncertain prospects. *American Economic Review* 59:25–34, 1969.
- [30] J.-Y. Jaffray. Existence of a continuous utility function: An elementary proof. *Econometrica* 43(5–6):981–983, 1975.
- [31] P. Kall and J. Mayer. *Stochastic Linear Programming*. Springer, New York, 2005.
- [32] M. Kijima and M. Ohnishi. Mean-risk analysis of risk aversion and wealth effects on optimal portfolios with multiple investment opportunities. *Annals of Operations Research* 45:147–163, 1993.
- [33] W. K. Klein Haneveld. On integrated chance constraints. *Stochastic Programming, Lecture Notes in Control and Information Science*, Vol. 76. Springer, Berlin, 194–209, 1986.
- [34] W. K. Klein Haneveld and M. H. van der Vlerk. Integrated chance constraints: Reduced forms and an algorithm. *Computational Management Science* 3(4):245–269, 2006.
- [35] V. Kozmik and D. P. Morton. Risk-averse stochastic dual dynamic programming. *Optimization Online*, E-print, [http://www.optimization.online.org/DB\\_HTML/2013/02/3794.html](http://www.optimization.online.org/DB_HTML/2013/02/3794.html), 2013.
- [36] P. A. Krokhmal. Higher moment coherent risk measures. *Quantitative Finance* 7(4):373–387, 2007.
- [37] P. A. Krokhmal, M. Zabarankin, and S. Uryasev. Modeling and optimization of risk. *Surveys in Operations Research and Management Science* 16:49–66, 2011.
- [38] A. Küenzi-Bay and J. Mayer. Computational aspects of minimizing conditional value-at-risk. *Computational Management Science* 3(1):3–27, 2006.
- [39] S. Kusuoka. On law invariant coherent risk measures. *Advances in Mathematical Economics* 3(1):83–95, 2001.
- [40] E. Lehmann. Ordered families of distributions. *Annals of Mathematical Statistics* 26:399–419, 1955.
- [41] J. Luedtke. New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization* 19(3):1433–1450, 2008.
- [42] N. Miller and A. Ruszczyński. Risk-averse two-stage stochastic linear programming: Modeling and decomposition. *Operations Research* 59(1):125–132, 2011.
- [43] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester, UK, 2002.
- [44] N. Noyan and A. Ruszczyński. Valid inequalities and restrictions for stochastic programming problems with first order stochastic dominance constraints. *Mathematical Programming* 114(2):249–275, 2008.

- [45] N. Noyan, G. Rudolf, and A. Ruszczyński. Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research Letters* 34(6):653–659, 2006.
- [46] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research* 116(1):33–50, 1999.
- [47] W. Ogryczak and A. Ruszczyński. On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming* 89(2):217–232, 2001.
- [48] W. L. Ogryczak and A. Ruszczyński. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* 13(1):60–78, 2002.
- [49] M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming* 52(2):359–375, 1991.
- [50] G. Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, Singapore, 2007.
- [51] A. B. Philpott and V. L. De Matos. Dynamic sampling algorithms for multistage stochastic programs with risk aversion. Technical report, Electric Power Optimization Centre, University of Auckland, Auckland, New Zealand, 2011.
- [52] A. Pichler and A. Shapiro. Uniqueness of Kusuoka representations. *Optimization Online*. E-print, [http://www.optimization-online.org/DB\\_HTML/2012/10/3660.html](http://www.optimization-online.org/DB_HTML/2012/10/3660.html), 2012.
- [53] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.
- [54] J. P. Quirk and R. Saposnik. Admissibility and measurable utility functions. *Review of Economic Studies* 29:140–146, 1962.
- [55] T. Rader. The existence of a utility function to represent preferences. *Review of Economic Studies* 30:229–232, 1963.
- [56] F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and Their Applications* 112(2):185–200, 2004.
- [57] R. T. Rockafellar. *Conjugate Duality and Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, 1974.
- [58] R. T. Rockafellar and S. Uryasev. Optimization of conditional value at risk. *Journal of Risk* 2:21–41, 2000.
- [59] R. T. Rockafellar and S. Uryasev. Conditional value at risk for general loss distributions. *Journal of Banking and Finance* 26:1443–1471, 2002.
- [60] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics* 10(1):51–74, 2006.
- [61] M. Rothschild and J. E. Stiglitz. Increasing risk. I. A definition. *Journal of Economic Theory* 2:225–243, 1970.
- [62] G. Rudolf and A. Ruszczyński. Optimization problems with second order stochastic dominance constraints: Duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization* 19(3):1326–1343, 2008.
- [63] A. Ruszczyński. Decomposition methods. A. Ruszczyński and A. Shapiro, eds. *Stochastic Programming, Handbooks in Operations Research and Management Science*, Vol. 10. Elsevier, Amsterdam, 141–211, 2003.
- [64] A. Ruszczyński. *Nonlinear Optimization*, Vol. 13. Princeton University Press, Princeton, NJ, 2006.
- [65] A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming* 125(2):235–261, 2010.
- [66] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research* 31(3):544–561, 2006.
- [67] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research* 31(3):433–452, 2006.
- [68] D. Schmeidler. Integral representation without additivity. *Proceedings of the American Mathematical Society* 97(2):255–261, 1986.
- [69] A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research* 209(1):63–72, 2011.

- [70] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. SIAM, Philadelphia, 2009.
- [71] A. V. Skorohod. Limit theorems for stochastic processes. *Teoria Veroyatnostey i Primeneniya* 1:289–319, 1956.
- [72] V. Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics* 38:423–439, 1965.
- [73] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1944.
- [74] M. E. Yaari. The dual theory of choice under risk. *Econometrica* 55(1):95–115, 1987.