SIAM J. Optim.
Vol. 22, No. 2, pp. 286–312

SAMPLING-BASED DECOMPOSITION METHODS FOR MULTISTAGE STOCHASTIC PROGRAMS BASED ON EXTENDED POLYHEDRAL RISK MEASURES*

VINCENT GUIGUES† AND WERNER RÖMISCH‡

Abstract. We define a risk-averse nonanticipative feasible policy for multistage stochastic programs and propose a methodology to implement it. The approach is based on dynamic programming equations written for a risk-averse formulation of the problem. This formulation relies on a new class of multiperiod risk functionals called extended polyhedral risk measures. Dual representations of such risk functionals are given and used to derive conditions of coherence. In the one-period case, conditions for convexity and consistency with second order stochastic dominance are also provided. The risk-averse dynamic programming equations are specialized considering convex combinations of one-period extended polyhedral risk measures such as spectral risk measures. To implement the proposed policy, the approximation of the risk-averse recourse functions for stochastic linear programs is discussed. In this context, we detail a stochastic dual dynamic programming algorithm which converges to the optimal value of the risk-averse problem.

Key words. convex risk measure, coherent risk measure, stochastic programming, risk-averse optimization, decomposition algorithms, Monte-Carlo sampling, spectral risk measure, CVaR

AMS subject classifications. 90C15, 91B30

DOI. 10.1137/100811696

1. Introduction.

Let us consider a T-stage optimization problem of the form

$$\begin{aligned}(1)\qquad & \inf \mathbb{E}\left[\sum_{t=1}^{T} f_t(x_t,\xi_t)\right]\\ & x_t \in \chi_t(x_{t-1},\xi_t) \text{ a.s.},\ x_t\ \mathcal{F}_t\text{-measurable},\ t=1,\ldots,T,\end{aligned}$$

where $(\xi_t)_{t=1}^T$ is a stochastic process, $\mathcal{F}_t$ is the sigma-algebra $\mathcal{F}_t := \sigma(\xi_j, j \le t)$, and $\chi_t : \mathbb{R}^{N_{t-1,x}} \times \mathbb{R}^{M_t} \rightrightarrows \mathbb{R}^{N_{t,x}}$ are given multifunctions. In this setting, multistage stochastic optimization problems set two challenging questions. The first question refers to modeling: how does one deal with uncertainty in this context? Once a model is chosen, the second question is, how does one design suitable solution methods?

For the first of these questions, we are interested in defining *nonanticipative* policies. This means that the decision we make at any time step should be a function of the available history $\xi_{[t]}$ of the process at this time step. This is a necessary condition for a *policy* to be implementable since a decision has to be made on the basis of the available information. We will focus on models with recourse. More precisely, introducing a recourse function $\mathcal{Q}_{t+1}$ for time step t and given x_{t-1}, the decision x_t is found by solving the problem

$$\begin{aligned}(2)\qquad & \inf_{x_t} f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}(x_t,\xi_{[t]})\\ & x_t \in \chi_t(x_{t-1},\xi_t)\end{aligned}$$

*Received by the editors October 14, 2010; accepted for publication (in revised form) December 23, 2011; published electronically April 4, 2012.
http://www.siam.org/journals/siopt/22-2/81169.html

†IMPA, Instituto de Matemática Pura e Aplicada, 110 Estrada Dona Castorina, Jardim Botanico, Rio de Janeiro, Brazil (vguigues@impa.br).
‡Institute of Mathematics, Humboldt-University Berlin, 10099 Berlin, Germany (romisch@math.hu-berlin.de).

at time step t. In this problem, we have assumed that ξ_t is available at time step t and thus $\xi_{[t]}$ gathers all the realizations of ξ_j up to time step t. The policy depends crucially on the choice of the recourse function $\mathcal{Q}_{t+1}$ used in (2). Given x_0 and the information $\xi_{[1]}$, a non-risk-averse model uses the recourse functions defined by

$$(3) \qquad \mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}) = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf\limits_{x_t} f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) \\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right)$$

for $t = 1, \ldots, T$, with $\mathcal{Q}_{T+1} \equiv 0$. These dynamic programming (DP) equations are associated to the non-risk-averse model

$$(4) \qquad \begin{array}{l} \inf \mathbb{E}\left[\displaystyle\sum_{t=1}^{T} f_t(x_t(\xi_{[t]}), \xi_t)\right] \\ x_t(\xi_{[t]}) \in \chi_t(x_{t-1}(\xi_{[t-1]}), \xi_t), \ t = 1, \ldots, T. \end{array}$$

For the second of these questions, most of the efforts so far have been placed on solution methods that approximate the recourse functions (3) in the case of multistage stochastic linear programs. In this paper, we contribute to these two questions as follows.

From the modeling point of view, we define risk-averse recourse functions. For this purpose, a common approach (Ruszczyński and Shapiro [RS06a], [RS06b]) is based on a risk-averse nested formulation of the problem using conditional (coherent) risk measures. In this situation, it is in general difficult, even for simple risk measures such as the *conditional value-at-risk* (CVaR) (Rockafellar and Uryasev [RU02]), to determine a risk-averse problem (using a risk measure that has a physical interpretation) whose stagewise decomposition is given by these DP equations. However, such an interpretation is important. This is why we define instead a risk-averse problem for (1) that is then decomposed by stages to obtain DP equations. A similar idea appears in the recent book by Shapiro, Dentcheva, and Ruszczyński [SDR09, Chapter 6, p. 326], where a convex combination of the expectation and of the CVaR of the final wealth is used for a portfolio selection problem. Instead, we control partial costs (the sum of the costs up to the current time step) and use a new class of risk measures that is suitable for decomposing the risk-averse problem by stages. This class of multiperiod risk measures called *extended polyhedral risk measures* has three appealing properties. First, the class is large: it contains the polyhedral risk measures (Eichhorn and Römisch [ER05]); in the one-period case some special cases include the optimized certainty equivalent (Ben-Tal and Teboulle [BTT07]), some spectral risk measures (Acerbi [Ace02]), and the CVaR. More generally, conditions for such functionals to be coherent or convex are provided. Second, as stated above, it allows us to define DP equations for our risk-averse problem. Finally, these equations are suitable for proposing convergent solution methods for a class of stochastic linear programs.

Regarding algorithmic issues, exact decomposition algorithms such as the nested decomposition (ND) algorithm have shown their superiority to direct solution methods for obtaining approximations of the recourse functions. Each iteration of these algorithms computes upper and lower bounds on the optimal mean cost. If an optimal solution to the problem exists, the algorithm finds an optimal solution after a finite number of iterations. These exact algorithms build at each iteration and each node of the scenario tree a cut for the recourse functions. These cuts form an outer linearization of these recourse functions.

There are two important variants of the ND algorithm: a variant that adds quadratic proximal terms in the objective functions of the master problems and a variant that uses multicuts (Ruszczyński [Rus86]).

The purpose of the first variant is to discourage the solution from moving too far from the best solution found so far, and this can significantly accelerate the convergence of the method even if the master problems are quadratic programs with this approach. The proximal term penalties are positive and can be dynamically modified in the course of the algorithm.

In the ND algorithm, for a given node in the scenario tree and a given input state x_{t-1} at t, the subproblems associated to all the realizations in stage $t+1$ are solved to obtain their optimal simplex multipliers. These multipliers are then aggregated to obtain a single cut for each node in each iteration. In the multicut variant, there are as many cuts as descendant realizations that are built at each iteration. More information is thus passed from the children nodes to their immediate ancestor by sending disaggregate cuts. The size of the master programs increases, but we expect fewer iterations (see Birge and Louveaux [BL88]).

However, in some applications, the number of scenarios may become so large that even these improved variants are difficult to apply since they entail prohibitive computational efforts.

Monte Carlo sampling-based algorithms constitute an interesting alternative in such situations. Higle and Sen [HS96] introduced a stochastic cutting plane method for two-stage stochastic programs and showed its convergence with probability one. Recently, Higle, Rayco, and Sen [HRS10] extended this idea to multistage models by applying a stochastic cutting plane method to the dual problem resulting when dualizing nonanticipativity constraints. Their method is, hence, based on scenario decomposition. A different approach for two-stage problems based on Monte Carlo (importance) sampling within the L-shaped method was introduced by Dantzig and Glynn [DG90] and Infanger [Inf92]. For multistage stochastic linear programs whose number of immediate descendant nodes is small but with many stages, Pereira and Pinto [PP91] proposed sampling in the forward pass of the ND. This sampling-based variant of the ND is the so-called stochastic dual dynamic programming algorithm on which we focus our attention. More precisely, we detail a *stochastic dual dynamic programming* (SDDP) algorithm (Pereira and Pinto [PP91]) to approximate our risk-averse recourse functions, to be used in (2) in place of $\mathcal{Q}_{t+1}$. The computation of the cuts in the backward pass of SDDP are detailed in this risk-averse setting.

Our developments can be easily extended to other sampling-based decomposition methods such as AND and DOASA.

The abridged nested decomposition (AND) algorithm proposed by Birge and Donohue [BD06] is a variant of SDDP that also involves sampling in the forward pass. This algorithm determines in a different manner the sequence of states and scenarios in the forward pass. The numerical simulations in Birge and Donohue [BD06] report lower computational time on average for the AND algorithm in comparison with SDDP.

When the number of immediate descendant nodes is large (possibly infinite) and when the problem has many stages, we also can (or even must) sample in the backward pass. In this case, for a given node on a forward path k, not all the optimal simplex multipliers associated to the descendant subproblems are computed. Only the descendant subproblems associated with some realizations are solved. As explained in the cut calculation algorithm (CCA) in Philpott and Guan [PG08], it is, however, possible in this situation to replace the “missing” multipliers by some coefficients so that the cuts built still lie below the corresponding recourse functions. This gives rise to dynamic outer approximation sampling algorithms (DOASA) described in Philpott and Guan [PG08].

The paper is organized as follows. In the second section, we introduce the class of multiperiod extended polyhedral risk measures and study their properties: dual representations are derived and used to provide criteria for convexity and coherence and, in the one-period case, for convexity and consistency with second order stochastic dominance. In section 3, we derive DP equations for a risk-averse problem defined in terms of extended polyhedral risk measures. We also provide conditions that guarantee the convergence of SDDP in this risk-averse setting. Finally, in section 4, we propose to use SDDP to approximate the risk-averse recourse functions from section 3 for some stochastic linear programs. In particular, formulas for the cuts in the backward pass are given.

We mention that after writing our paper we became aware of two recent and closely related papers: Collado, Papp, and Ruszczyński [CPR], based on scenario decomposition, and Shapiro [Sha11], which suggests using SDDP to approximate risk-averse recourse functions defined from a nested risk-averse formulation of a multistage stochastic program.

We start by setting down some notation:

- For $x \in \mathbb{R}^n$, the vectors x^+ and x^- are defined by $x^+(i) = \max(x(i), 0)$ and $x^-(i) = \max(-x(i), 0)$ for $i = 1, \ldots, n$.
- For a nonempty set $X \subseteq \mathbb{R}^n$, the polar cone X^* is defined by $X^* = \{x^* : \langle x, x^* \rangle \leq 0 \ \forall x \in X\}$, where $\langle \cdot, \cdot \rangle$ is the standard scalar product on $\mathbb{R}^n$.
- e is a column vector of all ones.
- If A is an $m_1 \times n$ matrix and B an $m_2 \times n$ matrix, $(A; B)$ denotes the $(m_1 + m_2) \times n$ matrix $\left(\begin{smallmatrix} A \\ B \end{smallmatrix} \right)$.
- For vectors $x_1, \ldots, x_T \in \mathbb{R}^n$ and $1 \leq t_1 \leq t_2 \leq T$, we denote $(x_{t_1}, \ldots, x_{t_2}) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ by $x_{t_1:t_2}$.
- For $x, y \in \mathbb{R}^n$, the vector $x \circ y \in \mathbb{R}^n$ is defined by $(x \circ y)(i) = x(i)y(i),\ i = 1, \ldots, n$.
- I_n is the $n \times n$ identity matrix, and $0_{m,n}$ is an $m \times n$ matrix of zeros.
- δ_{ij} is the Kronecker delta defined for i, j integers by $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.
- $\mathcal{Q}_{t+1}$ denotes a (generic) recourse function used at time step $t = 1, \ldots, T$, i.e., $\mathcal{Q}_{T+1} \equiv 0$, and if $t < T$, then $\mathcal{Q}_{t+1}(x_t, \xi_{[t]})$ represents a cost over the period $t + 1, \ldots, T$. Various recourse functions at t will be defined using the same notation $\mathcal{Q}_{t+1}$. Which $\mathcal{Q}_{t+1}$ is relevant will be clear from the context.

As is usually done in the stochastic programming literature and to alleviate notation, we use the same notation for a random variable and for a particular realization of this random variable, the context allowing us to know which concept is being referred to.

2. Extended polyhedral risk measures. We consider multiperiod risk functionals ρ whose arguments are sequences of random variables. We confine ourselves to discrete-time processes with a finite time horizon as in Ruszczyński and Shapiro [RS06a]. Such risk functionals have to assess the riskiness of a finite sequence $z_1, \ldots, z_T$ of random variables for some integer $T \geq 2$. To reflect the evolution of information as time goes by, we assume that z_t is measurable with respect to some σ-field $\mathcal{F}_t$, where $\mathcal{F}_1, \ldots, \mathcal{F}_T$ form a filtration, i.e., $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_T = \mathcal{F}$, with $\mathcal{F}_1 = \{\emptyset, \Omega\}$. In this setting, z_1 is deterministic, and a multiperiod risk functional ρ will be seen as a mapping $\rho : \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P}) \to \bar{\mathbb{R}}$ for some $p \in [1, +\infty)$.

Remark 2.1. Throughout the paper, the arguments $(z_1, \ldots, z_T)$ of the risk functionals will be interpreted as accumulated revenues (for which higher values are preferred). More precisely, if $\tilde{z}_t$ is the revenue for time step t, we consider the accumulated revenues $z_t = \sum_{\tau=1}^{t} \tilde{z}_\tau,\ t = 1, \ldots, T$.

For future use, we recall the definition of multiperiod convex risk measures (from Artzner et al. [ADE$^+$], [ADE$^+$07], Föllmer and Schied [FS04]) which are multiperiod risk functionals of special interest when the random variables z_t represent revenues (accumulated or not).

Definition 2.2. *A functional ρ on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ is called a* multiperiod convex risk measure *if conditions* (i)–(iii) *below hold:*

(i) Monotonicity*: if $z_t \le \tilde{z}_t$ a.s, $t = 1, \ldots, T$, then $\rho(z_1, \ldots, z_T) \ge \rho(\tilde{z}_1, \ldots, \tilde{z}_T)$.*

(ii) Translation invariance*: for each $r \in \mathbb{R}$ we have $\rho(z_1 + r, \ldots, z_T + r) = \rho(z_1, \ldots, z_T) - r$.*

(iii) Convexity*: for each $\lambda \in [0, 1]$ and $z, \tilde{z} \in \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ we have $\rho(\lambda z + (1-\lambda)\tilde{z}) \le \lambda\rho(z) + (1-\lambda)\rho(\tilde{z})$.*

It is called a multiperiod coherent risk measure *if in addition condition* (iv) *holds:*

(iv) Positive homogeneity*: for each $\lambda \ge 0$ we have $\rho(\lambda z_1, \ldots, \lambda z_T) = \lambda\rho(z_1, \ldots, z_T)$.*

In the literature, there appear different requirements instead of the translation invariance (ii) above, e.g., Fritelli and Scandalo [FS05] and Pflug and Römisch [PR07].

Convex duality can be used to obtain dual representations of multiperiod convex risk measures. Next, we cite such a representation that uses the set $\mathcal{D}_T$ of generalized density functions where

$$\mathcal{D}_T := \left\{ \lambda \in \times_{t=1}^{T} L_1(\Omega, \mathcal{F}_t, \mathbb{P}) \ :\ \lambda_t \ge 0 \text{ a.s.}, t = 1, \ldots, T,\ \sum_{t=1}^{T} \mathbb{E}[\lambda_t] = 1 \right\}.$$

Theorem 2.3. *Let $\rho : \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P}) \to \bar{\mathbb{R}}$ and assume that ρ is proper (i.e., ρ is finite on the nonempty set* $\operatorname{dom} \rho = \{z : \rho(z) < \infty\}$*) and lower semicontinuous. Then ρ is a multiperiod convex risk measure if and only if it admits the representation*

$$\rho(z) = \sup \left\{ \mathbb{E}\left(-\sum_{t=1}^{T} \lambda_t z_t \right) - \rho^*(\lambda) \ :\ \lambda \in \mathcal{P}_\rho \right\} \tag{5}$$

for some convex closed subset $\mathcal{P}_\rho \subseteq \mathcal{D}_T$ of the space $\times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P})$ $(\frac{1}{p} + \frac{1}{q} = 1)$ on which the conjugate ρ^* of ρ is given too. The functional ρ is coherent if and only if the conjugate ρ^* in (5) *is the indicator function of $\mathcal{P}_\rho$.*

A proof of the above theorem can be found in, e.g., Ruszczyński and Shapiro [RS06b]. We are now in a position to define the class of multiperiod extended polyhedral risk measures.

Definition 2.4. *A risk measure ρ on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ is called* multiperiod extended polyhedral *if there exist matrices $A_t, B_{t,\tau}$, vectors a_t, c_t, and functions $h_t(z) = (h_{t,1}(z), \ldots, h_{t,n_{t,2}}(z))^\top$ for given functions $h_{t,1}, \ldots, h_{t,n_{t,2}} : L_p(\Omega, \mathcal{F}_t, \mathbb{P}) \to L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P})$ with $1 \le p' \le p$ such that*

$$\rho(z_1, \ldots, z_T) = \begin{cases} \inf\ \mathbb{E}[\sum_{t=1}^{T} c_t^\top y_t] \\ y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}),\ t = 1, \ldots, T, \\ A_t y_t \le a_t \ a.s., t = 1, \ldots, T, \\ \sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} = h_t(z_t)\ a.s.,\ t = 2, \ldots, T. \end{cases} \tag{6}$$

Another less general extension of polyhedral risk measures is due to Eichhorn [Eic07]. Like a multiperiod polyhedral risk measure (Eichhorn and Römisch [ER05]), a multiperiod extended polyhedral risk measure is given as the optimal value of a T-stage linear stochastic program where the arguments of the risk measure appear on the right-hand side of the dynamic constraints. Multiperiod polyhedral risk measures constitute a particular case with $a_t = 0,\ t = 2, \ldots, T,\ B_{t,\tau}$ row vectors, and $h_t(z_t) = h_{t,1}(z_t) = z_t$ (i.e., $n_{t,2} = 1$).

We mention that multiperiod extended polyhedral risk measures satisfy two additional properties that were recently discussed in the literature: *information monotonicity* (see Kovacevic and Pflug [KP09]) and *time consistency*, suggested in Shapiro [Sha09]. Information monotonicity means that the risk $\rho(z_1, \ldots, z_T)$ gets smaller if the available information expressed by the σ-fields $\mathcal{F}_t,\ t = 1, \ldots, T$, increases. Since $\rho(z_1, \ldots, z_T)$ is given by a risk-neutral multistage stochastic program, it is time consistent as stated at the beginning of Shapiro [Sha09, section 3].

The need to consider the extended versions from Definition 2.4 is twofold:

(i) *Modeling*: Some (popular) risk measures are extended polyhedral but not polyhedral in the sense of Eichhorn and Römisch [ER05] (see examples at the end of this section).

(ii) *Algorithmic issues*: As announced in the introduction, DP equations can be written for risk-averse versions of (1) defined in terms of extended polyhedral risk measures. Moreover, the convergence of a class of decomposition algorithms applied to the corresponding nested formulation of the risk-averse problem will be proved in section 3 for a subclass of extended polyhedral risk measures that contain some nonpolyhedral risk measures. For this subclass, we have $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$.

In view of (ii) above, extended polyhedral risk measures with $h_t(z_t) = z_t b_t + \tilde{b}_t$ play a particular role when algorithmic issues come into play. In the rest of this section, we study properties of such risk functionals. In this context, the matrices $A_t, B_{t,\tau}$ and the vectors $a_t, b_t, \tilde{b}_t$, and c_t are fixed and deterministic. They have to be chosen such that ρ exhibits desirable risk measure properties. In particular, conditions on these parameters for the corresponding extended polyhedral risk measure to be coherent are given in the Corollary 2.6 of Theorem 2.5, which follows. This theorem gives dual representations for stochastic program (6) when $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$. In what follows, the dimensions of a_t and b_t are, respectively, denoted by $n_{t,1}$ and $n_{t,2}$.

THEOREM 2.5. *Let ρ be a functional of the form (6) on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ with $p \in [1, \infty)$ and $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$. Assume*

(i) *complete recourse: $\{y_1 : A_1 y_1 \le a_1\} \neq \emptyset$ and, for every $t = 2, \ldots, T$, it holds that $\{B_{t,0} y_t : A_t y_t \le a_t\} = \mathbb{R}^{n_{t,2}}$;*

(ii) *dual feasibility: $\{(u, v) : u \in \times_{t=1}^T \mathbb{R}^{n_{t,1}}, v \in \times_{t=2}^T \mathbb{R}^{n_{t,2}}, c_t + A_t^\top u_t + \sum_{\tau=\max(2,t)}^T B_{\tau,\tau-t}^\top v_{\tau-1} = 0,\ t = 1, \ldots, T\} \neq \emptyset$.*

Then ρ is finite, convex, and continuous on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ and with $\frac{1}{p} + \frac{1}{q} = 1$ the following dual representation holds:

(7)
$$\rho(z) = \left\{ \begin{array}{l} \sup\ -\mathbb{E}[\sum_{t=1}^T\ \lambda_{1,t}^\top a_t + \sum_{t=2}^T \lambda_{2,t-1}^\top (z_t b_t + \tilde{b}_t)] \\ \lambda_1 \in \times_{t=1}^T\ L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}),\ \lambda_2 \in \times_{t=2}^T\ L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \\ \lambda_{1,t} \ge 0\ a.s.,\ t = 1, \ldots, T, \\ c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^T\ B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1} | \mathcal{F}_t] = 0\ a.s.,\ t = 1, \ldots, T. \end{array} \right.$$

We also have

$$\rho(z) = \sup \left\{ \mathbb{E}\left[\sum_{t=1}^{T} z_t^* z_t\right] - \rho^*(z^*) \ : \ z^* \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}) \right\}, \tag{8}$$

where ρ^* *is the conjugate of* ρ. *Next, for every* $z^* \in \operatorname{dom}(\rho^*)$, $\rho^*(z^*)$ *is given by*

(9)

$$\rho^*(z^*) = \begin{cases} \inf \ \mathbb{E}[\sum_{t=1}^{T} \ \lambda_{1,t}^\top a_t + \sum_{t=2}^{T} \lambda_{2,t-1}^\top \tilde{b}_t] \\ \lambda_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \ \lambda_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \\ z_t^* = -\lambda_{2,t-1}^\top b_t \ a.s., \ t = 2, \ldots, T, \ \lambda_{1,t} \geq 0 \ a.s., \ t = 1, \ldots, T, \\ c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} \ B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1} | \mathcal{F}_t] = 0 \ a.s., \ t = 1, \ldots, T, \end{cases}$$

where

$$\operatorname{dom}(\rho^*) = \left\{ \begin{array}{l} z^* \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}) \ \textit{such that} \\ \exists \ \lambda_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \\ \lambda_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}) \ \textit{satisfying} \\ \lambda_{1,t} \geq 0 \ a.s., \ t = 1, \ldots, T, \\ c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} \ B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1} | \mathcal{F}_t] = 0 \ a.s., \\ t = 1, \ldots, T, \ \textit{and} \\ z_1^* = 0, \ z_t^* = -\lambda_{2,t-1}^\top b_t \ a.s., \ t = 2, \ldots, T \end{array} \right\}. \tag{10}$$

Proof. We use results on Lagrangian and conjugate duality. Consider the following Banach spaces and their duals:

$$\begin{array}{ll} E := \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), & E^* = \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), \\ Z := \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P}), & Z^* = \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}), \end{array}$$

with bilinear forms

$$\langle e, e^* \rangle_{E/E^*} = \sum_{t=1}^{T} \mathbb{E}[e_t^\top e_t^*] \quad \text{and} \quad \langle z, z^* \rangle_{Z/Z^*} = \sum_{t=1}^{T} \mathbb{E}[z_t z_t^*].$$

Let us introduce the Lagrange multipliers $\lambda_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}})$ (with $\lambda_1 \geq 0$ a.s.) and $\lambda_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}})$ associated to the constraints of (6) and the Lagrangian

$$\begin{aligned} L(y, \lambda_1, \lambda_2) := \mathbb{E} \Bigg[& \sum_{t=1}^{T} c_t^\top y_t + \lambda_{1,t}^\top (A_t y_t - a_t) \\ & + \sum_{t=2}^{T} \lambda_{2,t-1}^\top \left(\sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} - z_t b_t - \tilde{b}_t \right) \Bigg] \\ = \mathbb{E} & \left[\sum_{t=1}^{T} (c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^\top \lambda_{2,\tau-1})^\top y_t \right] \\ & + \mathbb{E} \left[- \sum_{t=1}^{T} \lambda_{1,t}^\top a_t - \sum_{t=2}^{T} \lambda_{2,t-1}^\top (z_t b_t + \tilde{b}_t) \right]. \end{aligned}$$

The dual functional is defined by

$$\theta(\lambda_1, \lambda_2) := \inf_{y \in E} L(y, \lambda_1, \lambda_2), \tag{11}$$

and the Lagrangian dual of (6) is the problem

$$\begin{aligned} \sup_{\lambda_1, \lambda_2} \big\{ \theta(\lambda_1, \lambda_2) \ :\ & \lambda_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \\ & \lambda_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \lambda_1 \geq 0 \,\text{a.s.} \big\}. \end{aligned} \tag{12}$$

Due to Ruszczyński and Shapiro [RS03, Proposition 5, Chapter 1], the conditional expectation operator $\mathbb{E}[\cdot|\mathcal{F}_t]$ and the operation of minimization can be interchanged in (11), which gives for $\theta(\lambda_1, \lambda_2)$ the expression

$$\begin{aligned} & -\mathbb{E}\left[\sum_{t=1}^{T} \lambda_{1,t}^\top a_t + \sum_{t=2}^{T} \lambda_{2,t-1}^\top (z_t b_t + \tilde{b}_t)\right] \\ & \quad + \mathbb{E}\left[\sum_{t=1}^{T} \inf_{y_t \in \mathbb{R}^{k_t}} (c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t])^\top y_t\right]. \end{aligned}$$

Next, $\inf_{y_t \in \mathbb{R}^{k_t}} (c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t])^\top y_t$ is 0 if

$$c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t] = 0$$

and $-\infty$ otherwise. The Lagrangian dual (12) can thus be expressed as

$$\begin{aligned} & \sup\ -\mathbb{E}\left[\sum_{t=1}^{T} \lambda_{1,t}^\top a_t + \sum_{t=2}^{T} \lambda_{2,t-1}^\top (z_t b_t + \tilde{b}_t)\right] \\ & \lambda_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}),\ \lambda_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \lambda_1 \geq 0 \,\text{a.s.}, \\ & c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t] = 0 \text{ a.s.},\ t = 1, \ldots, T. \end{aligned} \tag{13}$$

From weak duality and dual feasibility, we obtain $\rho(z) > -\infty$, and due to the complete recourse assumption $\rho(z) < +\infty$. It follows that $\rho(z)$ is finite. More precisely, dual feasibility and complete recourse imply that there is no duality gap: the optimal value of (6), i.e., $\rho(z)$, is the optimal value of (13). This shows (7).

Next, we use conjugate duality. Let us introduce the vectors $c = (c_1, \ldots, c_T)^\top$, $a = (a_1, \ldots, a_T)^\top$, and $\tilde{b} = (\tilde{b}_2, \ldots, \tilde{b}_T)^\top$ and the matrices

$$A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_T \end{pmatrix}, \qquad \mathcal{B} = \begin{pmatrix} 0 & b_2 & & \\ \vdots & & \ddots & \\ 0 & & & b_T \end{pmatrix},$$

and

$$B = \begin{pmatrix} B_{2,1} & B_{2,0} & 0 & \ldots & 0 \\ B_{3,2} & B_{3,1} & B_{3,0} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ B_{T,T-1} & B_{T,T-2} & B_{T,T-3} & \ldots & B_{T,0} \end{pmatrix}.$$

Let also $Y = \{y \in E \ : \ Ay(\omega) \leq a \text{ for a.e. } \omega \in \Omega\}$ and

$$\begin{array}{rccl} \varphi: & E \times Z & \to & \bar{\mathbb{R}} \\ & (y,z) & \to & \varphi(y,z) = \langle y, c \rangle_{E/E^*} + \delta_Y(y) + \delta_{\{0\}}(By - \mathcal{B}z - \tilde{b}), \end{array}$$

where δ denotes the indicator function taking values 0 and $+\infty$ only. Since Y is closed and convex, φ is lower semicontinuous and convex. With this notation, we can express $\rho(z)$ as $\rho(z) = \inf_{y \in E} \ \varphi(y,z)$ and, due to Bonnans and Shapiro [BS00, Proposition 2.143], ρ is convex. Since ρ is finite valued, [BS00, Proposition 2.152] guarantees the continuity of ρ. Since ρ is proper, convex, and lower semicontinuous, by the Fenchel–Moreau theorem we have that $\rho^{**} = \rho$, where ρ^{**} is the biconjugate of ρ, i.e.,

$$\rho(z) = \rho^{**}(z) = \sup \left\{ \langle z, z^* \rangle_{Z/Z^*} - \rho^*(z^*) \ : \ z^* \in Z^* \right\}, \tag{14}$$

which is (8). Next, $\rho^*(z^*) = \varphi^*(0, z^*)$, where the conjugate φ^* of φ is given by

$$\begin{aligned} \varphi^*(y^*, z^*) &= \sup \left\{ \langle y, y^* \rangle_{E/E^*} + \langle z, z^* \rangle_{Z/Z^*} - \varphi(y,z) \ : \ y \in E, \ z \in Z \right\} \\ &= \sup \left\{ \langle y, y^* - c \rangle_{E/E^*} + \langle z, z^* \rangle_{Z/Z^*} \ : \ Ay \leq a \text{ a.s.}, \ By = \mathcal{B}z + \tilde{b} \text{ a.s.} \right\}. \end{aligned}$$

It follows that

$$\rho^*(z^*) = \left\{ \begin{array}{l} \sup \ \mathbb{E}[\sum_{t=1}^{T} \ (z_t z_t^* - c_t^\top y_t)] \\ y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), \ z_t \in L_q(\Omega, \mathcal{F}_t, \mathbb{P}), \ t = 1, \ldots, T, \\ A_t y_t \leq a_t \text{ a.s.}, \ t = 1, \ldots, T, \\ \sum_{\tau=0}^{t-1} \ B_{t,\tau} y_{t-\tau} = z_t b_t + \tilde{b}_t \text{ a.s.}, \ t = 2, \ldots, T. \end{array} \right. \tag{15}$$

Due to (i) and (ii), complete recourse and dual feasibility also hold for (15) for every $z^* \in \operatorname{dom}(\rho^*)$, where $\operatorname{dom}(\rho^*)$ is given by (10). Using once again Lagrangian duality for problem (15), we obtain for $\rho^*(z^*)$ dual representation (9). ∎

Theorems 2.3 and 2.5 allow us to provide a criterion for an extended polyhedral risk measure to be multiperiod coherent.

COROLLARY 2.6. *Let ρ be a functional on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ of the form (6) with all a_t null and $h_t(z_t) = z_t b_t$ for some vector b_t. Let the conditions of Theorem 2.5 be satisfied (complete recourse and dual feasibility) and let*

$$\mathcal{M}_\rho = \left\{ \begin{array}{l} \lambda \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}) \ \ \textit{such that there exist} \\ \mu_1 \in \times_{t=1}^{T} \ L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \ \mu_2 \in \times_{t=2}^{T} \ L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}) \ \ \textit{satisfying} \\ \mu_{1,t} \geq 0 \ \textit{a.s.}, \ t = 1, \ldots, T, \\ c_t + A_t^\top \mu_{1,t} + \sum_{\tau=\max(2,t)}^{T} \ B_{\tau,\tau-t}^\top \mathbb{E}[\mu_{2,\tau-1} | \mathcal{F}_t] = 0 \ \textit{a.s.}, \\ t = 1, \ldots, T, \ \textit{and} \\ \lambda_1 = 0, \ \lambda_t = \mu_{2,t-1}^\top b_t \ \textit{a.s.}, \ t = 2, \ldots, T, \end{array} \right\}$$

be the (convex) set of dual multipliers. If $\mathcal{M}_\rho \subseteq \mathcal{D}_T$, then ρ is a multiperiod coherent risk measure.

Proof. Using representation (7) of Theorem 2.5, we can write $\rho(z) = \sup_{\lambda \in \mathcal{M}_\rho} - \sum_{t=1}^{T} \ \mathbb{E}[\lambda_t z_t]$. We conclude using Theorem 2.3 with $\mathcal{P}_\rho = \mathcal{M}_\rho$. ∎

Using representation (8) of Theorem 2.5, the properties of ρ can also be characterized by properties of $\operatorname{dom}(\rho^*)$, where $\operatorname{dom}(\rho^*)$ is given by (10).

COROLLARY 2.7. *Let ρ be a functional on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ of the form* (6) *with $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$, and let the conditions of Theorem* 2.5 *be satisfied (complete recourse and dual feasibility). The following hold:*

(i) *ρ is* monotone $\Longleftrightarrow$ *for all $z^* \in \operatorname{dom}(\rho^*)$ we have $z_t^* \le 0$ a.s. for $t = 1, \ldots, T$.*

(ii) *ρ is* translation invariant $\Longleftrightarrow$ *for all $z^* \in \operatorname{dom}(\rho^*)$ we have $\sum_{t=1}^{T} \mathbb{E}[z_t^*] = -1$.*

(iii) *ρ is* positively homogeneous $\Longleftrightarrow$ *for all $z^* \in \operatorname{dom}(\rho^*)$ we have $\rho^*(z^*) = 0$.*

When $T = 2$, since z_1 is deterministic, Definition 2.4 corresponds to one-period extended polyhedral risk measures that assess the riskiness of one random variable z only. For later reference we recall the definition of such risk measures which extend the class of one-period polyhedral risk measures.

DEFINITION 2.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $h(z) = (h_1(z), \ldots, h_{n_{2,2}}(z))^\top$ for given functions $h_1, \ldots, h_{n_{2,2}} : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to L_{p'}(\Omega, \mathcal{F}, \mathbb{P})$ with $1 \le p' \le p$. A risk measure ρ on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty)$ is called* extended polyhedral *if there exist matrices $A_1, A_2, B_{2,0}, B_{2,1}$, and vectors a_1, a_2, c_1, c_2 such that for every random variable $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$*

$$\rho(z) = \left\{ \begin{array}{l} \inf\ c_1^\top y_1 + \mathbb{E}[c_2^\top y_2] \\ y_1 \in \mathbb{R}^{k_1},\ y_2 \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{k_2}), \\ A_1 y_1 \le a_1,\ A_2 y_2 \le a_2\ a.s., \\ B_{2,1} y_1 + B_{2,0} y_2 = h(z)\ a.s. \end{array} \right. \tag{16}$$

For one-period risk measures, dual representations from Theorem 2.5 specialize as follows.

COROLLARY 2.9. *Let ρ be a functional of the form* (16) *on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with some $p \in [1, \infty)$ and $h(z) = z b_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Assume*

(i) *complete recourse: $\{y_1 : A_1 y_1 \le a_1\} \ne \emptyset$ and $\{B_{2,0} y_2 : A_2 y_2 \le a_2\} = \mathbb{R}^{n_{2,2}}$;*

(ii) *dual feasibility: $\{(u, v) : u \in \mathbb{R}^{n_{1,1}} \times \mathbb{R}^{n_{2,1}}, v \in \mathbb{R}^{n_{2,2}}, c_t + A_t^\top u_t + B_{2,2-t}^\top v = 0,\ t = 1, 2\} \ne \emptyset$.*

Then ρ is finite, convex, continuous, and with $\frac{1}{p} + \frac{1}{q} = 1$, ρ admits the dual representation

$$\rho(z) = \left\{ \begin{array}{l} \sup\ -\lambda_1^\top a_1 - \mathbb{E}[\lambda_2^\top a_2 + \lambda_3^\top (z b_2 + \tilde{b}_2)] \\ \lambda_1 \in \mathbb{R}^{n_{1,1}},\ \lambda_2 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}),\ \lambda_3 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}), \\ c_1 + A_1^\top \lambda_1 + B_{2,1}^\top \mathbb{E}[\lambda_3] = 0, \\ c_2 + A_2^\top \lambda_2 + B_{2,0}^\top \lambda_3 = 0\ a.s., \\ \lambda_1 \ge 0,\ \lambda_2 \ge 0,\ a.s. \end{array} \right.$$

We also have

$$\rho(z) = \sup \{\mathbb{E}[z^* z] - \rho^*(z^*)\ :\ z^* \in\ L_q(\Omega, \mathcal{F}, \mathbb{P})\}, \tag{17}$$

where ρ^* is the conjugate of ρ. Next, for every $z^* \in \operatorname{dom}(\rho^*)$, $\rho^*(z^*)$ is given by

$$\rho^*(z^*) = \left\{ \begin{array}{l} \inf\ \mathbb{E}[\lambda_1^\top a_1 + \lambda_2^\top a_2 + \lambda_3^\top \tilde{b}_2] \\ \lambda_1 \in \mathbb{R}^{n_{1,1}},\ \lambda_2 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}),\ \lambda_3 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}), \\ z^* = -\lambda_3^\top b_2\ a.s.,\ \lambda_1 \ge 0,\ \lambda_2 \ge 0\ a.s., \\ c_1 + A_1^\top \lambda_1 + B_{2,1}^\top \mathbb{E}[\lambda_3] = 0, \\ c_2 + A_2^\top \lambda_2 + B_{2,0}^\top \lambda_3 = 0\ a.s., \end{array} \right. \tag{18}$$

where

$$\operatorname{dom}(\rho^*) = \left\{ \begin{array}{l} z^* \in L_q(\Omega, \mathcal{F}, \mathbb{P}) \textit{ such that there exist} \\ \lambda_1 \in \mathbb{R}^{n_{1,1}},\ \lambda_2 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}), \\ \lambda_3 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}) \textit{ satisfying} \\ c_1 + A_1^\top \lambda_1 + B_{2,1}^\top \mathbb{E}[\lambda_3] = 0,\ \lambda_1 \geq 0,\ \lambda_2 \geq 0 \textit{ a.s.}, \\ c_2 + A_2^\top \lambda_2 + B_{2,0}^\top \lambda_3 = 0 \textit{ a.s.}, \textit{ and } \ z^* = -\lambda_3^\top b_2 \textit{ a.s.} \end{array} \right\}. \tag{19}$$

Proof. It suffices to use Theorem 2.5 with $T = 2$. ∎

Definition 2.2 specializes as follows to the one-period case.

DEFINITION 2.10. *A functional $\rho : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \bar{\mathbb{R}}$ is called a* convex risk measure *if it satisfies the following three conditions for all $z, \tilde{z} \in L_p(\Omega, \mathcal{F}, \mathbb{P})$:*

(i) Monotonicity*: if $z \leq \tilde{z}$ a.s., then $\rho(z) \geq \rho(\tilde{z})$.*

(ii) Translation invariance*: for each $r \in \mathbb{R}$ we have $\rho(z + r) = \rho(z) - r$.*

(iii) Convexity*: for all $\mu \in [0, 1]$ we have $\rho(\mu z + (1 - \mu)\tilde{z}) \leq \mu\rho(z) + (1 - \mu)\rho(\tilde{z})$.*

Such a functional ρ is said to be coherent *if it is positively homogeneous, i.e., $\rho(\mu z) = \mu\rho(z)$ for all $\mu \geq 0$ and $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$.*

Using Theorems 2.3 and Corollary 2.9, a sufficient criterion can be provided for a one-period extended polyhedral risk measure to be coherent.

COROLLARY 2.11. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form* (16) *with a_1, a_2 null, $p \in [1, \infty)$, and $h(z) = zb_2$ for some vector b_2. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and let $\mathcal{M}_\rho$ be the following (convex) set of dual multipliers:*

(20)

$$\mathcal{M}_\rho = \left\{ \begin{array}{l} \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}) \textit{ such that there exist} \\ (\mu_1, \mu_2, \mu_3) \in \mathbb{R}^{n_{1,1}} \times L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}) \times L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}) \ \textit{ satisfying} \\ c_1 + A_1^\top \mu_1 + B_{2,1}^\top \mathbb{E}[\mu_3] = 0, \\ c_2 + A_2^\top \mu_2 + B_{2,0}^\top \mu_3 = 0 \textit{ a.s.},\ \mu_1 \geq 0,\ \mu_2 \geq 0 \textit{ a.s. with } \ \lambda = \mu_3^\top b_2 \end{array} \right\}.$$

If $\mathcal{M}_\rho \subseteq \mathcal{D}_1$, then ρ is a (one-period) coherent risk measure.

Proof. From Corollary 2.9, we obtain $\rho(z) = \sup_{\lambda \in \mathcal{M}_\rho} -\mathbb{E}[\lambda z]$, and the result follows taking $\mathcal{P}_\rho = \mathcal{M}_\rho$ in Theorem 2.3. ∎

A dual representation of the second-stage problem for (16) will prove useful for obtaining further properties of one-period risk measures of the form (16).

PROPOSITION 2.12. *Let ρ be a functional of the form* (16) *on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with some $p \in [1, \infty)$ and $h(z) = zb_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility). Assume the feasible set $\mathcal{D}$ of the dual of the second-stage problem is nonempty where*

$$\mathcal{D} = \{\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^{n_{2,2}} \times \mathbb{R}^{n_{2,1}} \ : \ \lambda_2 \leq 0,\ \ B_{2,0}^\top \lambda_1 + A_2^\top \lambda_2 = c_2\}. \tag{21}$$

Then ρ is finite, convex, continuous and is given by

$$\rho(z) = \inf_{A_1 y_1 \leq a_1} \left\{ c_1^\top y_1 + \mathbb{E}\left[\sup_{\lambda \in \mathcal{D}} \ \lambda_1^\top (zb_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2 a_2 \right] \right\}.$$

Proof. Finiteness, convexity, and continuity follow from Corollary 2.9. Next, we write $\rho(z)$ as

$$\rho(z) = \inf_{y_1} \{c_1^\top y_1 + \mathbb{E}[\mathcal{Q}_2(y_1, z)] \ : \ A_1 y_1 \leq a_1\}, \tag{22}$$

where for each y_1 such that $A_1 y_1 \le a_1$ and for each $z \in \mathbb{R}$ we have defined

$$\mathcal{Q}_2(y_1, z) = \inf_{y_2} \{c_2^\top y_2 \ : \ B_{2,0} y_2 = z b_2 + \tilde{b}_2 - B_{2,1} y_1, \ A_2 y_2 \le a_2\}.$$

Finally, since $\mathcal{D} \neq \emptyset$, by duality, we can express $\mathcal{Q}_2(y_1, z)$ as

$$\begin{aligned} \mathcal{Q}_2(y_1, z) = \sup_{(\lambda_1,\lambda_2)} \{&\lambda_1^\top (z b_2 + \tilde{b}_2 - B_{2,1} y_1) \\ &+ \lambda_2^\top a_2 \ : \lambda_2 \le 0, \ B_{2,0}^\top \lambda_1 + A_2^\top \lambda_2 = c_2\}. \end{aligned} \tag{23}$$ ∎

The following proposition provides a sufficient criterion for some extended polyhedral risk measures to be convex risk measures when

$$Y_1 = \{y_1 : A_1 y_1 \le a_1\} \tag{24}$$

is not necessarily a cone (a_1 need not be 0).

PROPOSITION 2.13. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form* (16) *with $p \in [1, \infty)$ and $h(z) = z b_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and let $\mathcal{D}$ be defined as in Proposition* 2.12. *Assume*

(i) *$\mathcal{D} \neq \emptyset$ with $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$;*

(ii) *$c_1 \neq 0$ and b_2 is of the form $b_2 = -B_{2,1}^i / c_1(i)$ for at least one $i \in I = \{j \ : c_1(j) \neq 0\}$ with $y_1(i)$ unconstrained and where $B_{2,1}^i$ denotes the ith column of $B_{2,1}$.*

Then ρ is a finite-valued convex risk measure.

Proof. Let Y_1 be defined by (24). Finiteness and convexity of ρ follow from Corollary 2.9. The monotonicity of ρ follows from (i). Indeed, if $z, \tilde{z} \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ satisfy $z \le \tilde{z}$ a.s., then for every $y_1 \in Y_1$ and every $(\lambda_1, \lambda_2) \in \mathcal{D}$ we have

$$\lambda_1^\top (z b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2 \ge \lambda_1^\top (\tilde{z} b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2.$$

With the notation of Proposition 2.12 and with $\varphi(y_1, z) = c_1^\top y_1 + \mathbb{E}[\mathcal{Q}_2(y_1, z)]$, it follows that for every $y_1 \in Y_1$, we have $\mathbb{E}[\mathcal{Q}_2(y_1, z)] \ge \mathbb{E}[\mathcal{Q}_2(y_1, \tilde{z})]$, $\varphi(y_1, z) \ge \varphi(y_1, \tilde{z})$, and $\rho(z) = \inf_{y_1 \in Y_1} \varphi(y_1, z) \ge \inf_{y_1 \in Y_1} \varphi(y_1, \tilde{z}) = \rho(\tilde{z})$. The translation invariance condition follows from (ii). Indeed, eventually after reordering the components of y_1, c_1, and the columns of $B_{2,1}$, we can always assume that the index i satisfying (ii) is the last k_1th index, i.e., that $c_1, B_{2,1}$, and Y_1 are of the form $c_1 = (\hat{c}_1, \bar{c}_1)^\top$ with $\bar{c}_1 \in \mathbb{R}^*$, $B_{2,1} = [\hat{B}_{2,1}, -\bar{c}_1 b_2]$, and $Y_1 = \{y_1 = (\hat{y}_1, \bar{y}_1) \ : \ \hat{A}_1 \hat{y}_1 \le a_1, \bar{y}_1 \in \mathbb{R}\}$. We then have for each $r \in \mathbb{R}$, for each $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, and setting $\tilde{y}_1 = \bar{y}_1 + \frac{r}{\bar{c}_1} \in \mathbb{R}$

$$\begin{aligned} \rho(z + r) &= \inf_{\hat{A}_1 \hat{y}_1 \le a_1, \ \bar{y}_1 \in \mathbb{R}} \Bigg\{ \hat{c}_1^\top \hat{y}_1 + \bar{c}_1 \bar{y}_1 \\ &\qquad + \mathbb{E}\Bigg[\sup_{(\lambda_1,\lambda_2) \in \mathcal{D}} \lambda_1^\top ((z + r) b_2 + \tilde{b}_2 - \hat{B}_{2,1} \hat{y}_1 + \bar{y}_1 \bar{c}_1 b_2) + \lambda_2^\top a_2 \Bigg] \Bigg\} \\ &= \inf_{\hat{A}_1 \hat{y}_1 \le a_1, \ \tilde{y}_1 \in \mathbb{R}} \Bigg\{ \hat{c}_1^\top \hat{y}_1 + \bar{c}_1 \tilde{y}_1 \\ &\qquad + \mathbb{E}\Bigg[\sup_{(\lambda_1,\lambda_2) \in \mathcal{D}} \lambda_1^\top (z b_2 + \tilde{b}_2 - \hat{B}_{2,1} \hat{y}_1 + \tilde{y}_1 \bar{c}_1 b_2) + \lambda_2^\top a_2 \Bigg] \Bigg\} - r \\ &= \rho(z) - r. \qquad \square \end{aligned}$$

Proposition 2.13 extends the corresponding result in Eichhorn and Römisch [ER05]. Proposition 2.14 below shows that condition (i) in Proposition 2.13 ensures in fact a stronger type of monotonicity than (i) in Definition 2.10. Such monotonicity is based on *stochastic dominance rules* (see Müller and Stoyan [MS02]). For real-valued random variables $z, \tilde{z} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$, stochastic dominance rules are defined by classes of measurable real-valued functions on $\mathbb{R}$. The stochastic dominance rule with respect to class $\mathcal{F}$ is defined by

$$z \preceq_{\mathcal{F}} \tilde{z} \quad :\Longleftrightarrow \quad \forall\, f \in \mathcal{F} \;:\; [\text{ if } \mathbb{E}[f(z)] \text{ and } \mathbb{E}[f(\tilde{z})] \text{ exist, then } \mathbb{E}[f(z)] \leq \mathbb{E}[f(\tilde{z})]]$$

for each $z, \tilde{z} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$. Important special cases are the classes $\mathcal{F}_{nd}$ of nondecreasing functions and $\mathcal{F}_{ndc}$ of nondecreasing concave functions which, respectively, characterize first and second order stochastic dominance rules:

$$\begin{aligned} z \preceq_{FSD} \tilde{z} \quad &:\Longleftrightarrow z \preceq_{\mathcal{F}_{nd}} \tilde{z} \quad \Longleftrightarrow \quad \mathbb{P}(z \leq t) \geq \mathbb{P}(\tilde{z} \leq t) \quad \forall\, t \in \mathbb{R}, \\ z \preceq_{SSD} \tilde{z} \quad &:\Longleftrightarrow z \preceq_{\mathcal{F}_{ndc}} \tilde{z} \quad \Longleftrightarrow \quad \mathbb{E}[\min(z,t)] \leq \mathbb{E}[\min(\tilde{z},t)] \quad \forall\, t \in \mathbb{R}. \end{aligned}$$

In particular, it is said that a risk measure ρ is *consistent with second order stochastic dominance* (see Ogryczak and Ruszczyński [OR02]) if $z \preceq_{SSD} \tilde{z}$ implies $\rho(z) \geq \rho(\tilde{z})$.

PROPOSITION 2.14. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathcal{P})$ of the form* (16) *with $p \in [1, \infty)$ and $h(z) = zb_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and let $\mathcal{D}$ be defined as in Proposition* 2.12. *Assume $\mathcal{D} \neq \emptyset$ with $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$. Then ρ is consistent with second order stochastic dominance.*

Proof. With Y_1 defined as in (24), let g be the function defined for every $y_1 \in Y_1$ and $z \in \mathbb{R}$ by

$$g(y_1, z) = c_1^\top y_1 + \sup_{(\lambda_1, \lambda_2) \in \mathcal{D}} \{\lambda_1^\top (zb_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2\}. \tag{25}$$

For every $y_1 \in Y_1$, $g(y_1, \cdot)$ is convex and, since $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$, it is also nonincreasing. Let $z \preceq_{SSD} \tilde{z}$. For every $y_1 \in Y_1$, since $-g(y_1, \cdot)$ is concave and nondecreasing, $\mathbb{E}[-g(y_1, z)] \leq \mathbb{E}[-g(y_1, \tilde{z})]$ and $\rho(z) = \inf_{y_1 \in Y_1} \mathbb{E}[g(y_1, z)] \geq \inf_{y_1 \in Y_1} \mathbb{E}[g(y_1, \tilde{z})] = \rho(\tilde{z})$. ∎

For a one-period risk measure of the form (16) with $h(z) = zb_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$, the first-stage solution set $S(\rho(z)) \subseteq Y_1$ is given by

$$S(\rho(z)) = \left\{ y_1 \in Y_1 \;:\; \rho(z) = c_1^\top y_1 + \sup_{(\lambda_1, \lambda_2) \in \mathcal{D}} \{\lambda_1^\top (zb_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2\} \right\}. \tag{26}$$

For algorithmic issues considered in sections 3 and 4, it can be useful to have at hand conditions that guarantee the boundedness of $S(\rho(z))$. This question is addressed in the following proposition.

PROPOSITION 2.15. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form* (16) *with $p \in [1, \infty)$, a_2 null, and $h(z) = zb_2$ for some vector b_2. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and assume that $S(\rho(0))$ is nonempty and bounded. Then $S(\rho(z))$ is nonempty, convex, and compact for any $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. The proof follows the proof of Proposition 2.9 in Eichhorn and Römisch [ER05], with, in our case, g given by (25). ∎

We provide examples of extended polyhedral risk measures. The above criteria for coherence and consistency with second order stochastic dominance are applied.

Example 2.16 (spectral risk measures and CVaR). Let $F_z(x) = \mathbb{P}(z \leq x)$ be the distribution function of random variable z, and let $F_z^{\leftarrow}(p) = \inf\{x \;:\; F_z(x) \geq p\}$

be the usual generalized inverse of F_z. Given a risk spectrum $\phi \in L_1([0,1])$ the *spectral risk measure* ρ_ϕ generated by ϕ is given by Acerbi [Ace02]:

$$\rho_\phi(z) = -\int_0^1 F_z^{\leftarrow}(p)\phi(p)dp.$$

Spectral risk measures have been used in a number of applications (portfolio selection in Acerbi and Simonetti [AS], and insurance in Cotter and Kevin [CD06]). The conditional value-at-risk (CVaR) of level $0 < \varepsilon < 1$, also called average value-at-risk (AVaR) in Föllmer and Schied [FS04], is a particular spectral risk measure with a piecewise constant risk function ϕ having a jump at ε: $\phi(u) = \frac{1}{\varepsilon}1_{0\le u\le\varepsilon}$ (Acerbi [Ace02]). Let us consider more generally a piecewise constant risk function $\phi(\cdot)$ with J jumps at $0 < p_1 < p_2 < \cdots < p_J < 1$. Setting $\Delta\phi_k = \phi(p_k^+) - \phi(p_k^-) = \phi(p_k) - \phi(p_{k-1})$ for $k = 1, \dots, J$, with $p_0 = 0$, we assume

$$\text{(i) } \phi(\cdot) \text{ is positive,} \quad \text{(ii) } \Delta\phi_k < 0,\ k = 1, \dots, J, \quad \text{(iii) } \int_0^1 \phi(u)du = 1.$$

With this choice of ϕ, we can express $\rho_\phi(z)$ as the optimal value of a linear programming problem (see Acerbi and Simonetti [AS]):

$$\rho_\phi(z) = \inf_{x\in\mathbb{R}^J} \sum_{k=1}^J \Delta\phi_k[p_k x_k - \mathbb{E}\,[x_k - z]^+] - \phi(1)\mathbb{E}[z]. \tag{27}$$

When $J = 1$, $\Delta\phi_1 = -1/\varepsilon$, $p_1 = \varepsilon$, and $\phi(1) = 0$, the above formula reduces to the formula for the CVaR given by Rockafellar and Uryasev [RU02]:

$$CVaR^\varepsilon[z] = \inf_{x\in\mathbb{R}} \left[x + \frac{1}{\varepsilon}\,\mathbb{E}[z+x]^-\right]. \tag{28}$$

A spectral risk measure with a piecewise constant risk function satisfying (i), (ii), and (iii) above is a coherent extended polyhedral risk measure. Indeed, with respect to (16), we have $c_1 = \Delta\phi \circ p$ with $\Delta\phi = (\Delta\phi_1, \dots, \Delta\phi_J)^\top$, $c_2 = (-\Delta\phi; 0_{J,1}; -\phi(1))$, $B_{2,1} = (I_J; 0_{1,J}), B_{2,0} = (-I_J, I_J, 0_{J,1}; 0_{1,2J}, 1)$, $A_2 = (-I_{2J}, 0_{2J,1})$, and $h(z) = ze$. The matrix A_1 and the vectors a_1 and a_2 are null, b_2 is a $(J+1)$-vector of ones, and $\tilde{b}_2 = 0$. Notice that when $J > 1$ it is not polyhedral in the sense of Eichhorn and Römisch [ER05]. The complete recourse and dual feasibility assumptions from Corollary 2.9 are easily checked. This theorem provides for ρ_ϕ the dual representation

$$\rho_\phi(z) = \begin{cases} \sup\ -\mathbb{E}[\lambda z] \\ \lambda = \mu^\top e + \phi(1),\ \mu \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^J), \\ \mathbb{E}[\mu] = -\Delta\phi \circ p,\ 0 \le \mu \le -\Delta\phi \text{ a.s.} \end{cases} \tag{29}$$

Let $\mathcal{M}_{\rho_\phi}$ be the set of dual multipliers from Corollary 2.11 for ρ_ϕ. For every $\lambda \in \mathcal{M}_{\rho_\phi}$, we have $\lambda \ge 0$ a.s. and

$$\begin{aligned}\mathbb{E}[\lambda] = \mathbb{E}[\phi(1) + \mu^\top e] &= \phi(1) - \sum_{i=1}^J \Delta\phi_i p_i = \phi(1) - \sum_{i=1}^J (\phi(p_i) - \phi(p_{i-1}))p_i \\ &= \phi(0)p_1 + \sum_{i=1}^{J-1} \phi(p_i)(p_{i+1} - p_i) + (1 - p_J)\phi(1) = \int_0^1 \phi(u)du = 1.\end{aligned}$$

It follows that $\mathcal{M}_{\rho_\phi} \subseteq \mathcal{D}_1$ and using Corollary 2.11, ρ_ϕ is a coherent one-period risk measure. Next, the set $\mathcal{D}$ in Proposition 2.14 is given by $\mathcal{D} = \{(\lambda_1, \lambda_2) \in \mathbb{R}^{J+1} \times \mathbb{R}^{2J} \ : \ \lambda_2 \leq 0, \ \lambda_{1,J+1} = -\phi(1), \ \lambda_{1,1:J} = \lambda_{2,J+1:2J}, \ \lambda_{1,1:J} = -\lambda_{2,1:J} + \Delta\phi\}$. For every $(\lambda_1, \lambda_2) \in \mathcal{D}$, we have $\lambda_1^\top b_2 = \lambda_1^\top e \leq 0$. It follows that $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$ and due to Corollary 2.14, ρ_ϕ is consistent with second order stochastic dominance. When $J = 1$, $\Delta\phi_1 = -1/\varepsilon$, $p_1 = \varepsilon$, and $\phi(1) = 0$, $\rho_\phi = CVaR^\varepsilon$ and we recover results given in Eichhorn and Römisch [ER05]: the CVaR is consistent with second order stochastic dominance and is an extended polyhedral risk measure of the form (16) with $c_1 = 1$, $c_2 = (\frac{1}{\varepsilon}; 0)$, $B_{2,1} = -1$, $B_{2,0} = (-1, 1)$, $A_2 = -I_2$, $h(z) = z$, and A_1, a_1, a_2 null. The dual representation (29) becomes

$$CVaR^\varepsilon(z) = \sup\left\{-\mathbb{E}[\lambda z] : \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}), 0 \leq \lambda \leq \frac{1}{\varepsilon} \text{ a.s., } \mathbb{E}[\lambda] = 1\right\}.$$

Example 2.17 (optimized certainty equivalent (OCE) and expected utility). Given a concave nondecreasing utility function u, the optimized certainty equivalent $S_u(z)$ of the random variable z is defined in Ben-Tal and Teboulle [BTT07] by $S_u(z) = \sup_{y_1 \in \mathbb{R}} \ y_1 + \mathbb{E}[u(z - y_1)]$. Considering for u a piecewise affine concave function, we can express the convex function $-u$ as follows (see Rockafellar and Wets [RW98, Example 3.54]:

$$-u(x) = \inf\{c^\top y \ : \ y \in \mathbb{R}^k, y \geq 0, e^\top y = 1, \ b^\top y = x\} \tag{30}$$

for some vectors $b, c \in \mathbb{R}^k$. It follows that if u is a piecewise affine concave function, $\rho(z) = -S_u(z)$ is given by

$$\rho(z) = \left\{ \begin{array}{l} \inf \ -y_1 + \mathbb{E}[c^\top y_2] \\ y_1 \in \mathbb{R}, \ y_2 \in \mathbb{R}^k, \ y_2 \geq 0, \ e^\top y_2 = 1, \ b^\top y_2 = z - y_1. \end{array} \right. \tag{31}$$

In this case, the opposite of the OCE is an extended one-period polyhedral risk measure with h affine: $c_1 = -1$, $c_2 = c$, $A_2 = [-I_k; e^\top; -e^\top]$, $a_2 = [0_{k,1}; 1; -1]$, $B_{2,1} = 1$, $B_{2,0} = b^\top$, $b_2 = 1$, and A_1, a_1, and $\tilde{b}_2$ null. Notice that it is not polyhedral in the sense of Eichhorn and Römisch [ER05] and that complete recourse does not hold. However, properties of the OCE, given in Ben-Tal and Teboulle [BTT07], are easily checked: *monotonicity* follows from the definition of $-S_u$ and the fact that u is nondecreasing; *translation invariance* follows from the change of variable $\bar{y}_1 = y_1 - r$ in (31) (for $\rho(z + r)$) or in the definition of $-S_u(z + r)$; *convexity* can be checked directly from the definition of S_u (or using representation (31) and [BS00, Proposition 2.143], as in the proof of Theorem 2.5).

Let us consider as a special case a piecewise linear utility function of the form

$$u(x) = \gamma_1(x)^+ - \gamma_2(-x)^+, \text{ where } 0 \leq \gamma_1 < 1 < \gamma_2 \tag{32}$$

(note that $u(x) < x$ for $x \neq 0$). The corresponding risk measure $\rho(z) = -S_u(z)$ is an extended polyhedral risk measure with $c_1 = -1$, $c_2 = (-\gamma_1; \gamma_2)$, $B_{2,1} = 1, B_{2,0} = [1 - 1]$, $A_2 = -I_2$, $h(z) = z$, and A_1, a_2, a_2 null. Since complete (and even simple) recourse and dual feasibility hold, Corollary 2.9 provides the following dual representation:

$$\rho(z) = -S_u(z) = \sup\{-\mathbb{E}[\lambda z] : \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}), \ \mathbb{E}[\lambda] = 1, \ \gamma_1 \leq \lambda \leq \gamma_2 \text{ a.s.}\}.$$

Using Corollary 2.11, we deduce that when u is of the form (32), $\rho(z) = -S_u(z)$ is a coherent risk measure. More generally, it is shown in Ben-Tal and Teboulle [BTT07] that if u is a strongly risk-averse function (see Ben-Tal and Teboulle [BTT07]), $\rho(z) = -S_u(z)$ is coherent if and only if u is of the form (32). For $0 < \varepsilon < 1$, $CVaR^{\varepsilon}$ constitutes a particular case with $\gamma_1 = 0$ and $\gamma_2 = \frac{1}{\varepsilon}$. The set $\mathcal{D}$ in Proposition 2.14 is given by $\mathcal{D} = \{(\lambda_1, \lambda_2) : -\gamma_2 \le \lambda_1 \le -\gamma_1,\ \lambda_2 \le 0\}$. Since for every $(\lambda_1, \lambda_2) \in \mathcal{D}$ we have $\lambda_1^\top b_2 = \lambda_1^\top e \le 0$, using Proposition 2.14 we conclude that $-S_u(z)$ is consistent with second order stochastic dominance.

For any concave utility function u, the risk measure $\rho(z) = -\mathbb{E}(u(z))$ is an extended polyhedral risk measure with $h = u$, $B_{2,0} = c_2 = 1$, while the other parameters are null. In the particular case when u is a piecewise affine concave function, representation (30) shows that $-\mathbb{E}(u(z))$ can be written as an extended polyhedral risk measure with $h(z) = z$ and that complete recourse does not hold. However, a dual representation of ρ can be derived from the dual representation

$$-u(x) = \sup\{-\lambda_1 x - \lambda_2 : \lambda \in \mathbb{R}^2,\ \lambda_1 b + \lambda_2 e \le -c\} \tag{33}$$

of $-u$. Applying the expectation operator to both sides of the above equation and using Rockafellar and Wets [RW98, Theorem 14.60] (for switching the inf and expectation operators), we obtain for ρ the dual representation

$$\rho(z) = \sup\{-\mathbb{E}[\lambda_1 z + \lambda_2] : \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^2),\ \lambda_1 b + \lambda_2 e \le -c \text{ a.s.}\}.$$

Since $-u$ is nonincreasing, for every (λ_1, λ_2) in the feasible set of (33) we have $\lambda_1 \ge 0$ (otherwise, there would be positive subgradients of $-u$ at large enough points). It follows that in the above representation of ρ, $\lambda_1 \ge 0$ a.s., which implies that ρ is monotone, convex, and consistent with second order stochastic dominance. The *expected regret* or *expected loss* $\rho(z) = \mathbb{E}(z - \beta)^-$ for some target β is a special case (already considered in Eichhorn and Römisch [ER05]) with utility function $u(z) = -(z - \beta)^-$. Finally, notice that $\rho(z) = \mathbb{E}[(z - \mathbb{E}[z])^k]$ for some $1 \le k \le p - 1$ is an extended polyhedral risk measure with $h(z) = (z - \mathbb{E}[z])^k$.

Example 2.18 (multiperiod extended polyhedral risk measures). We consider functionals ρ on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ ($p \in [1, \infty)$) of the form $\rho(z) = \rho_\phi(\Phi(z))$, where ρ_ϕ is a spectral risk measure of form (27) with $\phi(\cdot)$ satisfying (i), (ii), (iii) in Example 2.16, and the function Φ is defined on $\mathbb{R}^T$ and maps to the extended real numbers.

Then ρ is a finite-valued coherent multiperiod risk measure if the function Φ (i) is concave, (ii) is monotone with respect to the (canonical) partial ordering in $\mathbb{R}^T$, (iii) is positively homogeneous, (iv) satisfies the property $\Phi(\zeta_1 + r, \ldots, \zeta_T + r) = \Phi(\zeta_1, \ldots, \zeta_T) + r$ for all $r \in \mathbb{R}$ and $\zeta \in \mathbb{R}^T$, and (v) has linear growth; i.e., for some constant $L > 0$ it holds $|\Phi(\zeta)| \le L \sum_{t=1}^{T} |\zeta_t|$ for every $\zeta \in \mathbb{R}^T$.

There are three important special cases of the function Φ:

(a) $\Phi(\zeta) = \sum_{t=1}^{T} \gamma_t \zeta_t$ with $\gamma_t \ge 0$, $t = 1, \ldots, T$, such that $\sum_{t=1}^{T} \gamma_t = 1$. Using (27), we have

$$\begin{aligned}
\rho(z) &= \rho_\phi\left(\sum_{t=1}^{T} \gamma_t z_t\right) \\
&= \inf_{x \in \mathbb{R}^J} (\Delta\phi \circ p)^\top x + \mathbb{E}\ \left(-\sum_{k=1}^{J} \Delta\phi_k \left[x_k - \sum_{t=1}^{T} \gamma_t z_t\right]^+ - \phi(1) \sum_{t=1}^{T} \gamma_t z_t\right) \\
&= \left\{\begin{array}{l}
\inf\ (\Delta\phi \circ p)^\top x + \mathbb{E}\ \left(-\sum_{k=1}^{J} \Delta\phi_k w_k - \phi(1) v_T\right) \\
x \in \mathbb{R}^J,\ v_t = v_{t-1} + \gamma_t z_t,\ v_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}),\ t = 1, \ldots, T,\ v_0 = 0, \\
w_k \ge 0,\ \ w_k \ge x_k - v_T,\ w_k \in L_p(\Omega, \mathcal{F}_T, \mathbb{P}),\ k = 1, \ldots, J.
\end{array}\right.
\end{aligned}$$

The stochastic program above can be rewritten in the form (6), and ρ is a multiperiod extended polyhedral coherent risk measure. In the case when $\rho_\phi = CVaR^\varepsilon$, according to the dual representation of $CVaR^\varepsilon$, we obtain

$$\rho(z) = \sup\Bigg\{-\sum_{t=1}^{T}\mathbb{E}(\lambda_t z_t) : \lambda_t \in L_q(\Omega,\mathcal{F}_t,\mathbb{P}),\ \mathbb{E}(\lambda_t)=\gamma_t,\ 0\le\lambda_t\le\frac{\gamma_t}{\varepsilon},\ t=1,\dots,T,$$
$$\gamma_t\mathbb{E}(\lambda_{t+1}|\mathcal{F}_t)=\gamma_{t+1}\lambda_t \text{ a.s., } t=1,\dots,T-1\Bigg\},$$

where $\lambda_t=\gamma_t\mathbb{E}(\lambda|\mathcal{F}_t)$, $t=1,\dots,T$, and $\frac{1}{p}+\frac{1}{q}=1$. Hence, ρ is a multiperiod extended polyhedral coherent risk measure according to Theorems 2.3 and 2.5.

(b) $\Phi(\zeta) = \min_{\gamma\in S}\langle\gamma,\zeta\rangle = \min_{\gamma\in S}\sum_{t=1}^T\gamma_t\zeta_t$, where S denotes the standard simplex $S=\{\gamma\in\mathbb{R}^T : \gamma_t\ge 0, t=1,\dots,T, \sum_{t=1}^T\gamma_t=1\}$, may be used instead of the function Φ in (a). This function satisfies conditions (i)–(v), but avoids specifying the weights γ_t, $t=1,\dots,T$.

(c) $\Phi(\zeta)=\min_{t=1,\dots,T}\zeta_t$ for $\zeta\in\mathbb{R}^T$. Using representation (27), we obtain

$$\begin{aligned}
\rho(z) &= \rho_\phi\Big(\min_{t=1,\dots,T} z_t\Big)\\
&= \inf_{x\in\mathbb{R}^J}(\Delta\phi\circ p)^\top x + \mathbb{E}\left(-\sum_{k=1}^J\Delta\phi_k\Big[x_k-\min_{t=1,\dots,T}z_t\Big]^+ - \phi(1)\min_{t=1,\dots,T}z_t\right)\\
&= \inf_{x\in\mathbb{R}^J}(\Delta\phi\circ p)^\top x + \mathbb{E}\left(-\sum_{k=1}^J\Delta\phi_k\max_{t=1\dots,T}(0,x_k-z_t)+\phi(1)\max_{t=1,\dots,T}-z_t\right)\\
&= \begin{cases}\inf\ (\Delta\phi\circ p)^\top x+\mathbb{E}\Big(-\sum_{k=1}^J\Delta\phi_k v_{kT}+\phi(1)v_T\Big)\\ x\in\mathbb{R}^J,\ v_1\ge -z_1,\ v_t\ge v_{t-1},\ v_t\ge -z_t,\ t=2,\dots,T,\\ v_{kt}\ge v_{kt-1},\ v_{kt}\ge x_k-z_t,\ v_t,v_{k,t}\in L_p(\Omega,\mathcal{F}_t,\mathbb{P}),\\ k=1,\dots,J,\ t=1,\dots,T,\ v_{k0}=0.\end{cases}
\end{aligned}$$

The latter linear stochastic program may be rewritten in the form (6), and ρ is a multiperiod extended polyhedral coherent risk measure. In the case when $\rho_\phi = CVaR^\varepsilon$, we obtain

$$\rho(z)=\inf\Bigg\{x+\frac{1}{\varepsilon}\mathbb{E}(v_T) : v_t\in L_p(\Omega,\mathcal{F}_t,\mathbb{P}), -x-z_t\le v_t, v_{t-1}\le v_t,$$
$$t=1,\dots,T, v_0=0, x\in\mathbb{R}\Bigg\}. \tag{34}$$

Example (34) was first studied by Eichhorn in [Eic07].

3. Risk-averse dynamic programming.

3.1. General setting. When using a multiperiod extended polyhedral risk measure to deal with uncertainty in the multistage stochastic programming framework (4), we consider accumulated revenues $z_t = -\sum_{\tau=1}^t f_\tau(x_\tau,\xi_\tau)$ and the sigma-algebras $\mathcal{F}_t=\sigma(\xi_j, j\le t)$ for $t=1,\dots,T$. Recall that x_0 and $\chi_1(x_0,\xi_1)$ are deterministic and that for any time step $t=1,\dots,T$, we denote by $\xi_{[t]}$ the available realizations of the process up to this time step, i.e., $\xi_{[t]}=(\xi_j, j\le t)$.

We also denote by $\mathcal{Z}_t$ the space of $\mathcal{F}_t$-measurable functions (these sets are embedded: $\mathcal{Z}_1 \subset \cdots \subset \mathcal{Z}_T$). Next, for $t = 1, \ldots, T$, we assume the following:

(H1) the functions $f_t : \mathbb{R}^{N_{t,x}} \times \mathbb{R}^{M_t} \to \mathbb{R}$ are continuous and $\chi_t : \mathbb{R}^{N_{t-1,x}} \times \mathbb{R}^{M_t} \rightrightarrows \mathbb{R}^{N_{t,x}}$ are measurable, bounded, and closed-valued multifunctions.

We are now in a position to define a risk-averse problem for (1) via a multiperiod risk measure. Let $\rho : \mathcal{Z}_1 \times \ldots \mathcal{Z}_T \to \mathbb{R}$ be a multiperiod risk measure and let us introduce the risk-averse problem

$$\begin{aligned} &\inf\ \rho\left(-f_1(x_1,\xi_1), -\sum_{\tau=1}^{2} f_\tau(x_\tau(\xi_{[\tau]}),\xi_\tau), \ldots, -\sum_{\tau=1}^{T} f_\tau(x_\tau(\xi_{[\tau]}),\xi_\tau)\right) \\ &x_t(\xi_{[t]}) \in \chi_t(x_{t-1}(\xi_{[t-1]}),\xi_t),\ t = 1,\ldots,T. \end{aligned} \tag{35}$$

In the above problem, the optimization is performed over $\mathcal{F}_t$-measurable functions $x_t, t = 1, \ldots, T$, satisfying the constraints and such that $f_t(x_t(\cdot), \cdot) \in \mathcal{Z}_t$. The sequence of measurable mappings $x_t(\cdot)$, $t = 1, \ldots, T$, is called a policy. The $\mathcal{F}_t$-measurability of $x_t(\cdot)$ implies the nonanticipativity of the policy, i.e., implies that x_t is a function of $\xi_{[t]}$. The policy obtained from (35) will be said to be risk-averse. A policy is said to be *feasible* if the constraints $x_t(\xi_{[t]}) \in \chi_t(x_{t-1}(\xi_{[t-1]}), \xi_t)$, $t = 1, \ldots, T$, are satisfied with probability one.

In this section, our objective is to provide a class of form (1) problems and a class of multiperiod risk measures ρ having the following two properties:

(P1) DP equations can be written for (35).

(P2) The SDDP algorithm applied to problem (35) decomposed by stages converges to an optimal solution of (35).

We intend to enforce (P2) obtaining DP equations that satisfy conditions given in Philpott and Guan [PG08]. These conditions imply the following:

(P3) The recourse functions are given as the optimal value of a non-risk-averse stochastic program (the objective function is an expectation) where the randomness appears on the right-hand side of the constraints only.

Property (P3) leads us naturally to use the class of extended polyhedral risk measures introduced in the previous section.

3.2. Extended polyhedral risk measures. Taking for ρ a multiperiod extended polyhedral risk measure of the form (6), problem (35) can be written as

$$\begin{aligned} &\inf\ \mathbb{E}\left[\sum_{t=1}^{T} c_t^\top y_t\right] \\ &A_t y_t \le a_t \text{ a.s., } t = 1,\ldots,T, \\ &\sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} = h_t\left(-\sum_{\tau=1}^{t} f_\tau(x_\tau,\xi_\tau)\right) \text{ a.s., } t = 2,\ldots,T, \\ &x_t \in \chi_t(x_{t-1},\xi_t) \text{ a.s., } t = 1,\ldots,T. \end{aligned} \tag{36}$$

Remark 3.1. In (36), the dependence of x_t and y_t with respect to $\xi_{[t]}$ was suppressed to alleviate notation. This will in general be done in what follows.

We first check that (P1) and (P3) hold for problem (36) above. Since we want to write DP equations, we start with the following simple remark.

Remark 3.2. Let us consider the following T-stage optimization problem:

$$P\left\{\begin{array}{l} \inf\ f(x_1,\ldots,x_T) \\ x_t \in X(x_0,\ldots,x_{t-1}),\ t = 1,\ldots,T. \end{array}\right.$$

We decompose f as $f(x) = \sum_{k=1}^{T} f_k(x_{1:k})$, where f_k is the sum of all the functions in the sum of functions defining f which depend on x_k but not on $x_{k+1:T}$ (for a given k, f_k is 0 if no such functions exist). DP equations for P can be written as follows:

$$\mathcal{Q}_t(x_{0:t-1}) = \begin{cases} \inf\limits_{x_t} f_t(x_{1:t}) + \mathcal{Q}_{t+1}(x_{0:t}) \\ x_t \in X(x_{0:t-1}) \end{cases}$$

for $t = 1, \ldots, T$, with $\mathcal{Q}_{T+1} \equiv 0$.

The application of Remark 3.2 to (36) yields the following DP equations: for $t = 1, \ldots, T$, $\mathcal{Q}_t(x_{0:t-1}, \xi_{[t-1]}, y_{1:t-1})$ is given by

$$\begin{aligned} &\mathcal{Q}_t(x_{0:t-1}, \xi_{[t-1]}, y_{1:t-1}) \\ &\quad = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf_{x_t, y_t} \; c_t^\top y_t + \mathcal{Q}_{t+1}(x_{0:t}, \xi_{[t]}, y_{1:t}) \\ A_t y_t \le a_t, \\ (1-\delta_{t1}) \left(\sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} - h_t \left(-\sum_{\tau=1}^{t} f_\tau(x_\tau, \xi_\tau) \right) \right) = 0, \\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right), \end{aligned} \tag{37}$$

where here, and in what follows, $\mathcal{Q}_{T+1} \equiv 0$. Since these DP equations correspond to the stagewise decomposition of risk-averse problem (36), the recourse functions $\mathcal{Q}_t$ in (37) are said to be *risk-averse*. Compared to the DP equations of the original stochastic program, a new state variable y_t and new constraints for it appear in (37) at time t. They serve for computing the multiperiod extended polyhedral risk measure.

Let us now take as a special case for ρ the multiperiod risk measure defined by

$$\rho(z_1, \ldots, z_T) = -\theta_1 \mathbb{E}[z_T] + \sum_{t=2}^{T} \theta_t \rho^t(z_t) \tag{38}$$

for some nonnegative weights θ_t, $t = 1, \ldots, T$, summing to one ($\sum_{t=1}^{T} \theta_t = 1$) and for some one-period coherent extended polyhedral risk measures $\rho^t : \mathcal{Z}_t \to \mathbb{R}$, $t = 2, \ldots, T$.

Remark 3.3. We easily check that ρ in (38) is a multiperiod (coherent) extended polyhedral risk measure.

Observe that since ρ^t is coherent and z_1 deterministic, we have $\rho^t(z_t - z_1) = \rho^t(z_t) + z_1$, and $\rho(z_1, \ldots, z_T)$ in (38) can be expressed as $\rho(z_1, \ldots, z_T) = -z_1 - \theta_1 \mathbb{E}[z_T - z_1] + \sum_{t=2}^{T} \theta_t \rho^t(z_t - z_1)$. This expression reveals that the corresponding objective function in (35) is the sum of the first-stage (deterministic) cost and of a convex combination of the mean future cost and of risk measures of future partial costs. With this choice of ρ, problem (35) becomes

$$\begin{aligned} &\inf \; f_1(x_1, \xi_1) + \theta_1 \mathbb{E}\left[\sum_{t=2}^{T} f_t(x_t, \xi_t) \right] + \sum_{t=2}^{T} \theta_t \rho^t \left(-\sum_{k=2}^{t} f_k(x_k, \xi_k) \right) \\ &x_t \in \chi_t(x_{t-1}, \xi_t), \; t = 1, \ldots, T. \end{aligned} \tag{39}$$

Plugging the expression (16) of the risk measure ρ^t (taking the same for all time steps) into (39), the latter can be written as

$$
\begin{aligned}
&\inf_{x_t,w_t,y_t} f_1(x_1,\xi_1)+\sum_{t=2}^{T}\theta_t c_1^\top w_t+\mathbb{E}\left[\theta_1\sum_{t=2}^{T} f_t(x_t,\xi_t)+\sum_{t=2}^{T}\theta_t c_2^\top y_t\right]\\
&B_{2,1}w_t+B_{2,0}y_t=h\left(-\sum_{k=2}^{t} f_k(x_k,\xi_k)\right),\ t=2,\dots,T,\\
&A_1 w_t\le a_1,\ \ A_2 y_t\le a_2,\ t=2,\dots,T,\\
&x_t\in\chi_t(x_{t-1},\xi_t),\ t=1,\dots,T.
\end{aligned}
$$

In turn, the above optimization problem can be expressed as

$$
\begin{aligned}
&\inf_{x_1,w_{2:T}} f_1(x_1,\xi_1)+\sum_{t=2}^{T}\theta_t c_1^\top w_t+\mathcal{Q}_2(x_1,\xi_{[1]},w_2,\dots,w_T)\\
&A_1 w_t\le a_1,\ t=2,\dots,T,\ x_1\in\chi_1(x_0,\xi_1),
\end{aligned}
\tag{40}
$$

where

$$
\mathcal{Q}_2(x_1,\xi_{[1]},w_{2:T})=\begin{cases}\inf\limits_{x_t,y_t}\ \mathbb{E}\left[\theta_1\sum_{t=2}^{T} f_t(x_t,\xi_t)+\sum_{t=2}^{T}\theta_t c_2^\top y_t\right]\\ B_{2,1}w_t+B_{2,0}y_t=h\left(-\sum_{k=2}^{t} f_k(x_k,\xi_k)\right),\ t=2,\dots,T,\\ A_2 y_t\le a_2,\ t=2,\dots,T,\\ x_t\in\chi_t(x_{t-1},\xi_t),\ t=2,\dots,T.\end{cases}
\tag{41}
$$

The application of Remark 3.2 to optimization problem (41) yields the following DP equations: for $t=2,\dots,T$, $\mathcal{Q}_t(x_{1:t-1},\xi_{[t-1]},w_{t:T})$ is given by

$$
\mathbb{E}_{\xi_t|\xi_{[t-1]}}\begin{pmatrix}\inf\limits_{x_t,y_t}\ \theta_1 f_t(x_t,\xi_t)+\theta_t c_2^\top y_t+\mathcal{Q}_{t+1}(x_{1:t},\xi_{[t]},w_{t+1:T})\\ B_{2,1}w_t+B_{2,0}y_t=h(-\sum_{k=2}^{t} f_k(x_k,\xi_k)),\ A_2 y_t\le a_2,\ x_t\in\chi_t(x_{t-1},\xi_t)\end{pmatrix}.
\tag{42}
$$

In DP equations (37) and (42) obtained for, respectively, risk-averse problems (36) and (39), the state variables memorize the relevant history of the process and of the decisions. For (37) (resp., (42)), we can reduce the size of the state vector replacing the history of the decisions $x_{1:t-1}$ by x_{t-1} and z_{t-1} (resp., x_{t-1} and $\tilde{z}_{t-1}$ with $\tilde{z}_{t-1}=z_{t-1}-z_1$). Variable $\tilde{z}_{t-1}$ represents the total revenue (opposite of the cost) from time step 2 until time step $t-1$ (i.e., the total income until time step $t-1$ for the time steps where the data are random). Variables $\tilde{z}_t$ satisfy $\tilde{z}_t=\tilde{z}_{t-1}-f_t(x_t,\xi_t)$ for $t=2,\dots,T$, with $\tilde{z}_1$ set equal to 0. With this notation, DP equations (37) for problem (36) become

$$
\begin{aligned}
&\mathcal{Q}_t(x_{t-1},\xi_{[t-1]},z_{t-1},y_{1:t-1})\\
&\quad=\mathbb{E}_{\xi_t|\xi_{[t-1]}}\begin{pmatrix}\inf\limits_{x_t,y_t,z_t}\ c_t^\top y_t+\mathcal{Q}_{t+1}(x_t,\xi_{[t]},z_t,y_{1:t})\\ (1-\delta_{t1})\left(\sum_{\tau=0}^{t-1} B_{t,\tau}y_{t-\tau}-h_t(z_t)\right)=0,\ A_t y_t\le a_t,\\ z_t=z_{t-1}-f_t(x_t,\xi_t),\ x_t\in\chi_t(x_{t-1},\xi_t)\end{pmatrix}
\end{aligned}
\tag{43}
$$

for $t = 1, \dots, T$, with $z_0 = 0$. As for the DP equations (40) and (42), they simplify as follows: in (40), $\mathcal{Q}_2(x_1, \xi_{[1]}, w_2, \dots, w_T)$ needs to be replaced by $\mathcal{Q}_2(x_1, \xi_{[1]}, \tilde{z}_1, w_2, \dots, w_T)$ and for $t = 2, \dots, T$ we have

$$\begin{aligned} \text{(44)} \qquad & \mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_{t:T}) \\ & = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf\limits_{x_t, \tilde{z}_t, y_t} \; - \delta_{tT}\theta_1 \tilde{z}_t + \theta_t c_2^\top y_t + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}, \tilde{z}_t, w_{t+1:T}) \\ B_{2,1} w_t + B_{2,0} y_t = h(\tilde{z}_t), \; A_2 y_t \le a_2, \\ \tilde{z}_t = \tilde{z}_{t-1} - f_t(x_t, \xi_t), \; x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right). \end{aligned}$$

Remark 3.4. Comparing the non-risk-averse DP equations (3) with the risk-averse ones (43) or (40) and (44), we see that additional decision and state variables are introduced in the latter cases. More precisely, at the first time step, in the non-risk-averse case the decision x_1 is taken, while in risk-averse case (43) (resp., (40) and (44)), additional decision variables y_1 and z_1 (resp., $(w_2, \dots, w_T)$) are needed. This first-stage problem is deterministic for all models.

For time step $t = 2, \dots, T$, in risk-averse case (43) (resp., (40) and (44)), the state vector is augmented with partial cost z_{t-1} and with the variables $(y_1, \dots, y_{t-1})$ (resp., partial cost $\tilde{z}_{t-1}$ and the variables $(w_t, \dots, w_T)$). For both risk-averse models, additional decisions z_t (or $\tilde{z}_t$) and y_t are needed for stages $t = 2, \dots, T$. This is summarized in Table 1.

TABLE 1

Decision and state variables for the non-risk-averse (NRA) DP equations (3) *as well as for the risk-averse ones* (43) *(RA$_1$), and* (40) *and* (44) *(RA$_2$).*

		First stage	*Stages $t = 2, \dots, T$*
Decision variables	NRA	x_1	x_t
	RA$_1$	(x_1, z_1, y_1)	(x_t, z_t, y_t)
	RA$_2$	$(x_1, w_2, \dots, w_T)$	$(x_t, \tilde{z}_t, y_t)$
State variables	NRA	$(x_0, \xi_{[0]})$	$(x_{t-1}, \xi_{[t-1]})$
	RA$_1$	$(x_0, \xi_{[0]})$	$(x_{t-1}, \xi_{[t-1]}, z_{t-1}, y_1, \dots, y_{t-1})$
	RA$_2$	$(x_0, \xi_{[0]})$	$(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_t, \dots, w_T)$

Remark 3.5. Other special cases for the multiperiod risk measure ρ in (35) for which DP equations can be written are the risk measures from Example 2.18.

Properties (P1) and (P3) thus hold for (36) and hold for (39) when using extended one-period polyhedral risk measures for ρ^t. We now concentrate on (P2). So far, all the developments of this section were valid for a problem of the form (1). To ensure that (P2) holds, we consider the special case when (1) is a stochastic linear program (SLP). Indeed, the convergence of the SDDP algorithm and of related sampling-based algorithms is proved in Philpott and Guan [PG08] for SLP. We observe that if (1) is an SLP, then risk-averse problem (36) (resp., (39)) is an SLP if and only if

$$\begin{aligned} \text{(45)} \qquad & h_t(z) = z b_t + \tilde{b}_t \text{ for some } b_t, \tilde{b}_t \in \mathbb{R}^{n_{t,2}} \\ & (\text{resp., } h(z) = z b_2 + \tilde{b}_2 \text{ for some } b_2, \tilde{b}_2 \in \mathbb{R}^{n_{2,2}}). \end{aligned}$$

Of interest for applications, we now specialize the above DP equations (44) taking extended polyhedral risk measures with $h(\cdot)$ of the kind (45) above. As seen in the previous section, spectral risk measures with piecewise constant spectra are of this kind. We provide the DP equations obtained in this case using directly (27).

3.3. Spectral risk measures. Let ϕ be a piecewise constant risk spectrum satisfying (i), (ii), and (iii) given in Example 2.16 and let $\Delta\phi_k = \phi(p_k) - \phi(p_{k-1}), k = 1, \ldots, J$. If we take for ρ^t a spectral risk measure ρ_ϕ (the same for all time steps), using (27) we can decompose (39) by stages and express it under the form

$$(46)\qquad \begin{array}{l} \inf\ f_1(x_1,\xi_1) + \sum_{t=2}^{T} \theta_t c_1^\top w_t + \mathcal{Q}_2(x_1, \xi_{[1]}, \tilde{z}_1, w_2, \ldots, w_T) \\ x_1 \in \chi_1(x_0, \xi_1),\ w_t \in \mathbb{R}^J,\ t = 2, \ldots, T, \end{array}$$

with $\tilde{z}_1 = 0, c_1 = \Delta\phi \circ p$, and where for $t = 2, \ldots, T$,

$$(47)\qquad \begin{array}{l} \mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_{t:T}) \\ \quad = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf\limits_{x_t, \tilde{z}_t}\ \tilde{f}_t(\tilde{z}_t, w_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}, \tilde{z}_t, w_{t+1:T}) \\ \tilde{z}_t = \tilde{z}_{t-1} - f_t(x_t, \xi_t),\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right) \end{array}$$

with

$$\tilde{f}_t(\tilde{z}_t, w_t) = -(\delta_{tT}\theta_1 + \phi(1)\theta_t)\tilde{z}_t - \theta_t\ \Delta\phi^\top(w_t - \tilde{z}_t e)^+.$$

When the risk spectrum ϕ has one jump, we obtain the CVaR.

3.4. Conditional value-at-risk. When taking $\rho^t = CVaR^{\varepsilon_t}$ and using (28), we can express (39) under the form

$$(48)\qquad \begin{array}{l} \inf\limits_{x_1, w_{2:T}} f_1(x_1,\xi_1) - \sum_{t=2}^{T} \theta_t w_t + \mathcal{Q}_2(x_1, \xi_{[1]}, \tilde{z}_1, w_2, \ldots, w_T) \\ x_1 \in \chi_1(x_0, \xi_1),\ w_t \in \mathbb{R},\ t = 2, \ldots, T, \end{array}$$

with $\tilde{z}_1 = 0$, and where for $t = 2, \ldots, T$,

$$(49)\quad \begin{array}{l} \mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_{t:T}) \\ \quad = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf\limits_{x_t, \tilde{z}_t}\ -\delta_{tT}\theta_1\tilde{z}_t + \dfrac{\theta_t}{\varepsilon_t}(w_t - \tilde{z}_t)^+ + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}, \tilde{z}_t, w_{t+1:T}) \\ \tilde{z}_t = \tilde{z}_{t-1} - f_t(x_t, \xi_t),\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right). \end{array}$$

3.5. Convergence of SDDP in a risk-averse setting. The convergence of the SDDP algorithm and of related sampling-based algorithms is proved in Philpott and Guan [PG08] for SLP with the following properties:

- (A1) Random data only appear on the right-hand side of the constraints.
- (A2) The supports of the distributions of the underlying random vectors are discrete and finite.
- (A3) Random vectors are interstage independent or satisfy a certain type of interstage dependence (see Philpott and Guan [PG08]).
- (A4) The feasible set of the linear program is nonempty and bounded in each stage.

In what follows, we consider multistage SLPs of the form (1) where

$$(50)\qquad f_t(x_t, \xi_t) = d_t^\top x_t \quad \text{and} \quad \chi_t(x_{t-1}, \xi_t) = \{x_t : x_t \geq 0,\ C_t x_t = \xi_t - D_t x_{t-1}\}.$$

For these programs, assumption (A1) holds, and it can be noted that if (A1) holds for (1), then (A1) holds for risk-averse problems (36) and (39). In the remainder of the paper, we assume (A2) and (A3). We also assume that (A4) holds for (1), which, in our context, can be expressed as follows:

(A4) For $t = 1, \dots, T$, for any feasible state x_{t-1}, and for any realization ξ_t^i of ξ_t, the set

$$\chi_t(x_{t-1}, \xi_t^i) = \{x_t \mid x_t \geq 0,\ C_t x_t = \xi_t^i - D_t x_{t-1}\}$$

is bounded and nonempty.

To apply the convergence results from Philpott and Guan [PG08] in our risk-averse setting, (A4) should also hold for risk-averse problems (36) or (39). For (36), (A4) takes the following form:

(A5) $\{y_1 : A_1 y_1 \leq a_1\}$ is bounded and for all $t = 2, \dots, T$, for any feasible states $x_1, y_1, \dots, x_{t-1}, y_{t-1}$, and for any sequence of realizations $\xi_1^i, \dots, \xi_t^i$ of $\xi_1, \dots, \xi_t$, the set $\{y_t : A_t y_t \leq a_t, B_{t,0} y_t = h_t(-\sum_{\tau=1}^{t} f_\tau(x_\tau, \xi_\tau^i)) - \sum_{\tau=1}^{t-1} B_{t,\tau} y_{t-\tau}$ for some $x_t \in \chi_t(x_{t-1}, \xi_t^i)\}$ is bounded and nonempty.

For (39), remembering Proposition 2.15, a condition implying (A4) is the following:

(A6) For $t = 2, \dots, T$, the sets $S(\rho^t(0))$ are nonempty and bounded, where $S(\rho^t(0))$ is defined in (26). $\{y_1 : A_1 y_1 \leq a_1\}$ is bounded and for all $t = 2, \dots, T$, for any feasible $x_1, y_1, \dots, x_{t-1}, y_{t-1}, w_{2:T}$, and for any sequence of realizations $\xi_1^i, \dots, \xi_t^i$ of $\xi_1, \dots, \xi_t$, the set $\{y_t : A_t y_t \leq a_t, \exists\ x_t \in \chi_t(x_{t-1}, \xi_t^i), B_{2,0} y_t = h(-\sum_{\tau=2}^{t} f_\tau(x_\tau, \xi_\tau^i)) - B_{2,1} w_t\}$ is bounded and nonempty.

Indeed, with respect to the non-risk-averse setting, recall that the additional decision variables for (39) are $\tilde{z}_t$ (bounded, due to (A4)), y_t, and w_t. Variables $w_t, t = 2, \dots, T$, are first-stage decision variables and, due to Proposition 2.15, if $S(\rho^t(0))$ is nonempty and bounded, then optimal w_t are bounded. Next, condition (A6) guarantees the boundedness of optimal y_t.

However, even if the feasible set at each stage for (36) or (39) is not bounded, we may be able to show, in some cases, that these feasible sets can be replaced by bounded feasible sets without changing the problems, i.e., that the solutions are bounded. Such is the case for problems (46) and (48). Indeed, for these problems, the only additional variables with respect to the non-risk-averse case are $\tilde{z}_t$ (bounded, due to (A4)) and first-stage variables $w_2, \dots, w_T$. For the spectral risk measure $\rho^t = \rho_\phi, t = 2, \dots, T$, considered in (46), the sets $S(\rho^t(0)) = S(\rho_\phi(0)) = \{0\}, t = 2, \dots, T$, are nonempty and bounded. Using Proposition 2.15, optimal values of w_t in (46) are bounded. This result can also be easily proved directly.

LEMMA 3.6. *Let assumption* (A4) *hold, and let ϕ be a piecewise risk spectrum satisfying* (i), (ii), *and* (iii) *given in Example* 2.16. *Let $w_2^*, \dots, w_T^*$ be optimal values of $w_2, \dots, w_T$ for* (46). *Then $w_t^*(k)$ is finite for every $t = 2, \dots, T$, and $k = 1, \dots, J$.*

Proof. Since $\chi_t, t = 1, \dots, T$, are bounded and $\Delta\phi < 0$, we can bound from below the objective function of (46) by $L_1(w) = K_1 + \sum_{t=2}^{T} \theta_t (\Delta\phi \circ p)^\top w_t$ and $L_2(w) = K_2 + \sum_{t=2}^{T} \theta_t (\Delta\phi \circ (p - e))^\top w_t$ for some constants K_1 and K_2. Since $\Delta\phi \circ p < 0$, if one component $w_t(k) = -\infty$, then $L_1(w) = +\infty$, the objective function is therefore $+\infty$, and such $w_t(k)$ cannot be an optimal value of $w_t(k)$. Similarly, since $\Delta\phi \circ (p - e) > 0$, if one $w_t(k) = +\infty$, then $L_2(w) = +\infty$, the objective function is $+\infty$, and such $w_t(k)$ cannot be an optimal value of $w_t(k)$. $\square$

The following corollary is an immediate consequence of this lemma.

COROLLARY 3.7. *Let assumption* (A4) *hold. Let $w_2^*, \dots, w_T^*$ be optimal values of $w_2, \dots, w_T$ for* (48). *Then w_t^* is finite for every $t = 2, \dots, T$.*

It follows that we can add (sufficiently large) box constraints on w_t in (46) and (48) without changing the optimal solutions of (46) and (48). Gathering our observations, we come to the following proposition.

PROPOSITION 3.8 (convergence of SDDP in a risk-averse setting). *Consider multistage SLPs of the form* (1) *with f_t and χ_t given by* (50). *Assume that for such multistage programs, assumptions* (A1), (A2), (A3), *and* (A4) *hold. Consider the risk-averse formulations* (46), (47) *and* (48), (49). *Then an SDDP algorithm applied on these DP equations will converge if the sampling procedures satisfy the FPSP and BPSP assumptions (see Philpott and Guan* [PG08]*).*

The same convergence result holds for the following two risk-averse formulations:

(1) *assuming* (A5), *for risk-averse program* (36) *decomposed by stages as* (43) *with $h_t(\cdot)$ given by* (45)*;*

(2) *assuming* (A6), *for risk-averse program* (39) *decomposed by stages as* (40), (44) *with $h(\cdot)$ given by* (45).

In the next section, we detail the SDDP algorithm for interstage independent risk-averse problems of form (35). The developments can be easily adapted to the case when the process affinely depends on previous values. Our notation follows closely that of Birge and Donohue [BD06].

4. Decomposition algorithms for a class of risk-averse stochastic programs. We consider the risk-averse recourse functions (43) from section 3 in the case when f_t and χ_t are given by (50) and $h_t(\cdot)$ is given by (45). Recall that risk-averse DP equations (43) satisfy (P3) (like the non-risk-averse DP equations (3) but with additional state and control variables). We assume interstage independence and relatively complete recourse for (1). We also assume that the hypotheses of Proposition 3.8 hold. In this context, relatively complete recourse also holds for risk-averse problems (43). As a result, the SDDP algorithm [PP91], [Sha11] can be applied to obtain approximations of the corresponding risk-averse recourse functions. At each iteration, this algorithm consists of a forward pass followed by a backward pass. The backward pass builds cuts for the recourse functions (hyperplanes lying below these functions) at some points computed in the forward pass. If H cuts are built for each recourse function at each iteration, iteration i ends with a lower bounding approximation of form

(51)
$$\mathcal{Q}_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}) = \max_{j=0,1,\ldots,iH} \left[-E_{t-1}^j x_{t-1} - Z_{t-1}^j z_{t-1} - \sum_{\tau=1}^{t-1} Y_{t-1}^{j,\tau} y_\tau + e_{t-1}^j \right]$$

for $\mathcal{Q}_t$, knowing that the algorithm starts taking for $\mathcal{Q}_t^0$ a known lower bounding affine approximation of $\mathcal{Q}_t$ while $\mathcal{Q}_{T+1}^i \equiv 0$. In the above expression, $Z_{t-1}^j \in \mathbb{R}$, while E_{t-1}^j and $Y_{t-1}^{j,\tau}$ are row vectors of appropriate dimensions.

The forward pass of iteration i samples H scenarios $(\xi_2^k, \ldots, \xi_T^k)$, $k = (i-1)H + 1, \ldots, iH$, from the distribution of $(\xi_2, \ldots, \xi_T)$. On scenario $(\xi_2^k, \ldots, \xi_T^k)$, the decisions $(x_1^k, \ldots, x_T^k, y_1^k, \ldots, y_T^k)$ as well as the partial costs $(z_1^k, \ldots, z_T^k)$ are computed replacing recourse functions $\mathcal{Q}_t$ by $\mathcal{Q}_t^{i-1}$ for $t = 2, \ldots, T+1$. The stopping criterion is discussed in [Sha11].

The cuts are computed from time step $T+1$ down to time step 2. For time step $T+1$, since $\mathcal{Q}_{T+1}^i = \mathcal{Q}_{T+1} = 0$, cuts for $\mathcal{Q}_{T+1}$ are obtained taking null values for $E_T^k, Z_T^k, Y_T^{k,\tau}$, and e_T^k for $k = (i-1)H + 1, \ldots, iH$. At $t = 2, \ldots, T$, cuts for $\mathcal{Q}_t$ are computed at $(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k)$, $k = (i-1)H + 1, \ldots, iH$. More precisely, having at hand the lower bounding approximation $\mathcal{Q}_{t+1}^i$ of $\mathcal{Q}_{t+1}$, we can bound from below $\mathcal{Q}_t(x_{t-1}, z_{t-1}, y_{1:t-1})$ by $\mathbb{E}_{\xi_t}[Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t)]$ with $Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t)$

given as the optimal value of the following linear program:

$$
(52) \quad \begin{array}{ll}
\inf\limits_{x_t, y_t, z_t, \tilde{\theta}_t} & c_t^\top y_t + \tilde{\theta}_t \\
& A_t y_t \le a_t, \; x_t \ge 0, \\
& \sum\limits_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} - z_t b_t = \tilde{b}_t, \qquad \text{(a)} \\
& z_t + d_t^\top x_t = z_{t-1}, \qquad \text{(b)} \\
& C_t x_t = \xi_t - D_t x_{t-1}, \qquad \text{(c)} \\
& \overrightarrow{E}_t^i x_t + \overrightarrow{Z}_t^i z_t + e\tilde{\theta}_t \ge -\sum\limits_{\tau=1}^{t} \overrightarrow{Y}_t^{i,\tau} y_\tau + \overrightarrow{e}_t^i, \qquad \text{(d)}
\end{array}
$$

where $\overrightarrow{Z}_t^i = (Z_t^0, Z_t^1, \ldots, Z_t^{iH})^\top$ and $\overrightarrow{Y}_t^{i,\tau}$ is the matrix whose $(j+1)$th line is $Y_t^{j,\tau}$ for $j = 0, \ldots, iH$. We denote by ξ_t^j, $j = 1, \ldots, q_t < +\infty$, the possible realizations of ξ_t with $p(t,j) = \mathbb{P}(\xi_t = \xi_t^j)$. We also denote by $\sigma_t^{k,j}, \mu_t^{k,j}, \pi_t^{k,j}$, and $\rho_t^{k,j}$ the (row vectors) optimal Lagrange multipliers associated to constraints (52)-(a), (52)-(b), (52)-(c), and (52)-(d) for the problem defining $Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j)$. With this notation, the following theorem provides the cuts computed for $\mathcal{Q}_t$ at iteration i.

THEOREM 4.1. *Let $\mathcal{Q}_t, t = 2, \ldots, T+1$, be the risk-averse recourse functions given by* (43) *with $h_t(\cdot)$ given by* (45). *In the backward pass of iteration i of the SDDP algorithm, the following cuts are computed for these recourse functions. For $t = T+1$, we set $E_{t-1}^k, Z_{t-1}^k, Y_{t-1}^{k,\tau}$ and e_{t-1}^k to 0 for $k = (i-1)H+1, \ldots, iH$ and $\tau = 1, \ldots, T$. For $t = 2, \ldots, T$ and $k = (i-1)H+1, \ldots, iH$, $E_{t-1}^k = \sum_{j=1}^{q_t} p(t,j)\pi_t^{k,j} D_t$ and*

$$
Z_{t-1}^k = -\sum_{j=1}^{q_t} p(t,j)\mu_t^{k,j}, Y_{t-1}^{k,\tau} = \sum_{j=1}^{q_t} p(t,j)(\sigma_t^{k,j} B_{t,t-\tau} + \rho_t^{k,j} \overrightarrow{Y}_t^{i,\tau}), \; \tau = 1, \ldots, t-1.
$$

Next, e_{t-1}^k is given by

$$
\sum_{j=1}^{q_t} p(t,j) \Bigg[Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j) - \mu_t^{k,j} z_{t-1}^k \\
+ \sum_{\tau=1}^{t-1} (\sigma_t^{k,j} B_{t,t-\tau} + \rho_t^{k,j} \overrightarrow{Y}_t^{i,\tau}) y_\tau^k + \pi_t^{k,j} D_t x_{t-1}^k \Bigg].
$$

Proof. Since relatively complete recourse and assumptions (A4) and (A5) hold, the linear program defining $Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j)$ has a nonempty feasible set and its optimal value is finite. As a result, both this primal problem and its dual have the same optimal value. Since a dual solution is a subgradient of the value function for problem (52), we obtain for $Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t^j)$ the lower bound

$$
Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j) - \sum_{\tau=1}^{t-1} \sigma_t^{k,j} B_{t,\tau}(y_{t-\tau} - y_{t-\tau}^k) - \sum_{\tau=1}^{t-1} \rho_t^{k,j} \overrightarrow{Y}_t^{i,\tau}(y_\tau - y_\tau^k) \\
+ \mu_t^{k,j}(z_{t-1} - z_{t-1}^k) - \pi_t^{k,j} D_t(x_{t-1} - x_{t-1}^k).
$$

Plugging this bound into the relation $\mathcal{Q}_t(x_{t-1}, z_{t-1}, y_{1:t-1}) \ge \sum_{j=1}^{q_t} p(t,j) Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t^j)$, rearranging the terms, and identifying with (51) gives the announced cuts. ∎

The above cuts can be easily specialized to DP equations (46)–(47) (based on spectral risk measures) or to (44) with $h(\cdot)$ as in (45).

5. Conclusion. The class of extended polyhedral risk measures was introduced in this paper. Dual representations of these risk measures were obtained and used to provide conditions for coherence, convexity, and consistency with second order stochastic dominance.

This class allowed us to write risk-averse dynamic programming equations for some risk-averse problems with risk measures taken from this class. We then detailed a stochastic dual dynamic programming algorithm for approximating the corresponding risk-averse recourse functions for some stochastic linear programs. In particular, conditions were given to guarantee convergence. The methodology can be easily adapted if the recourse functions are approximated using other sampling-based decomposition algorithms such as AND (Birge and Donohue [BD06]) and DOASA (Philpott and Guan [PG08]).

A forthcoming work will assess the proposed approach on a midterm multistage production management problem Guigues [Gui].

REFERENCES

[Ace02] C. Acerbi, *Spectral measures of risk: A coherent representation of subjective risk aversion*, J. Banking and Finance, 7 (2002), pp. 1505–1518.

[ADE+] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku, *Coherent Multiperiod Risk Measurement*, available online from http://www.math.ethz.ch/~delbaen, 2003.

[ADE+07] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku, *Coherent multiperiod risk adjusted values and Bellman's principle*, Ann. Oper. Res., 152 (2007), pp. 5–22.

[AS] C. Acerbi and P. Simonetti, *Portfolio Optimization with Spectral Measures of Risk*, Abaxbank internal report, 2002; available online from http://www.gloriamundi.org/.

[BD06] J. R. Birge and C. J. Donohue, *The abridged nested decomposition method for multistage stochastic linear programs with relatively complete recourse*, Algorithmic Oper. Res., 1 (2006), pp. 20–30.

[BL88] J. R. Birge and F. V. Louveaux, *A multicut algorithm for two-stage stochastic linear programs*, European J. Oper. Res., 34 (1988), pp. 384–392.

[BS00] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[BTT07] A. Ben-Tal and M. Teboulle, *An old-new concept of convex risk measures: The optimized certainty equivalent*, Math. Finance, 17 (2007), pp. 449–476.

[CD06] J. Cotter and K. Dowd, *Extreme spectral risk measures: An application to futures clearinghouse variation margin requirements*, J. Banking and Finance, 30 (2006), pp. 3469–3485.

[CPR] R. A. Collado, D. Papp, and A. Ruszczyński, *Scenario Decomposition of Risk-Averse Multistage Stochastic Programming Problems*, 2010; available online from http://www.optimization-online.org/DB_HTML/2010/08/2717.html.

[DG90] G. B. Dantzig and P. W. Glynn, *Parallel processors for planning under uncertainty*, Ann. Oper. Res., 22 (1990), pp. 1–21.

[Eic07] A. Eichhorn, *Stochastic Programming Recourse Models: Approximation, Risk Aversion, Applications in Energy*, Logos Verlag, Berlin, 2007.

[ER05] A. Eichhorn and W. Römisch, *Polyhedral risk measures in stochastic programming*, SIAM J. Optim., 16 (2005), pp. 69–95.

[FS04] H. Föllmer and A. Schied, *Stochastic Finance. An Introduction in Discrete Time*, 2nd ed., Walter de Gruyter, Berlin, 2004.

[FS05] M. Fritelli and G. Scandalo, *Risk measures and capital requirements for processes*, Math. Finance, 16 (2005), pp. 589–612.

[Gui] V. Guigues, *SDDP for Some Interstage Dependent Risk Averse Problems and Application to Hydro-thermal Planning*, preprint, 2011; available online from http://www.optimization-online.org/DB_HTML/2011/03/2970.html.

[HRS10] J. L. Higle, B. Rayco, and S. Sen, *Stochastic scenario decomposition for multistage stochastic programs*, IMA J. Manag. Math., 21 (2010), pp. 39–66.

[HS96] J. L. Higle and S. Sen, *Stochastic Decomposition*, Kluwer Academic, Dordrecht, The Netherlands, 1996.

[Inf92] G. Infanger, *Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs*, Ann. Oper. Res., 39 (1992), pp. 69–95.

[KP09] R. Kovacevic and G. Ch. Pflug, *Time consistency and information monotonicity of multiperiod acceptability functionals*, in Advanced Financial Modeling, H. Albrecher, W. J. Rungaldier, and W. Schachermayer, eds., Radon Ser. Comput. Appl. Math. 8, Walter de Gruyter, Berlin, 2009, pp. 347–369.

[MS02] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester, UK, 2002.

[OR02] W. Ogryczak and A. Ruszczyński, *Dual stochastic dominance and related mean-risk models*, SIAM J. Optim., 13 (2002), pp. 60–78.

[PG08] A. B. Philpott and Z. Guan, *On the convergence of stochastic dual dynamic programming and related methods*, Oper. Res. Lett., 36 (2008), pp. 450–455.

[PP91] M. V. F. Pereira and L. M. V. G. Pinto, *Multi-stage stochastic optimization applied to energy planning*, Math. Programming, 52 (1991), pp. 359–375.

[PR07] G. Ch. Pflug and W. Römisch, *Modeling, Measuring, and Managing Risk*, World Scientific, Singapore, 2007.

[RS03] A. Ruszczyński and A. Shapiro, *Stochastic Programming*, Handbooks Oper. Res. Management Sci. 10, Elsevier, Amsterdam, 2003.

[RS06a] A. Ruszczyński and A. Shapiro, *Conditional risk mappings*, Math. Oper. Res., 31 (2006), pp. 544–561.

[RS06b] A. Ruszczyński and A. Shapiro, *Optimization of convex risk functions*, Math. Oper. Res., 31 (2006), pp. 433–452.

[RU02] R. T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Banking and Finance, 26 (2002), pp. 1443–1471.

[Rus86] A. Ruszczyński, *A multicut regularized decomposition method for minimizing a sum of polyhedral functions*, Math. Programming, 35 (1986), pp. 309–333.

[RW98] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[SDR09] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[Sha09] A. Shapiro, *On a time consistency concept in risk averse multistage programming*, Oper. Res. Lett., 37 (2009), pp. 143–147.

[Sha11] A. Shapiro, *Analysis of stochastic dual dynamic programming method*, European J. Oper. Res., 209 (2011), pp. 63–72.

SIAM J. OPTIM.
Vol. 22, No. 2, pp. 286–312

SAMPLING-BASED DECOMPOSITION METHODS FOR MULTISTAGE STOCHASTIC PROGRAMS BASED ON EXTENDED POLYHEDRAL RISK MEASURES*

VINCENT GUIGUES† AND WERNER RÖMISCH‡

Abstract. We define a risk-averse nonanticipative feasible policy for multistage stochastic programs and propose a methodology to implement it. The approach is based on dynamic programming equations written for a risk-averse formulation of the problem. This formulation relies on a new class of multiperiod risk functionals called extended polyhedral risk measures. Dual representations of such risk functionals are given and used to derive conditions of coherence. In the one-period case, conditions for convexity and consistency with second order stochastic dominance are also provided. The risk-averse dynamic programming equations are specialized considering convex combinations of one-period extended polyhedral risk measures such as spectral risk measures. To implement the proposed policy, the approximation of the risk-averse recourse functions for stochastic linear programs is discussed. In this context, we detail a stochastic dual dynamic programming algorithm which converges to the optimal value of the risk-averse problem.

Key words. convex risk measure, coherent risk measure, stochastic programming, risk-averse optimization, decomposition algorithms, Monte-Carlo sampling, spectral risk measure, CVaR

AMS subject classifications. 90C15, 91B30

DOI. 10.1137/100811696

1. Introduction.

Let us consider a T-stage optimization problem of the form

$$\begin{aligned}&\inf \mathbb{E}\left[\sum_{t=1}^{T} f_t(x_t,\xi_t)\right]\\ &x_t \in \chi_t(x_{t-1},\xi_t) \text{ a.s., } x_t\ \mathcal{F}_t\text{-measurable, } t=1,\ldots,T,\end{aligned} \tag{1}$$

where $(\xi_t)_{t=1}^T$ is a stochastic process, $\mathcal{F}_t$ is the sigma-algebra $\mathcal{F}_t := \sigma(\xi_j, j \le t)$, and $\chi_t : \mathbb{R}^{N_{t-1,x}} \times \mathbb{R}^{M_t} \rightrightarrows \mathbb{R}^{N_{t,x}}$ are given multifunctions. In this setting, multistage stochastic optimization problems set two challenging questions. The first question refers to modeling: how does one deal with uncertainty in this context? Once a model is chosen, the second question is, how does one design suitable solution methods?

For the first of these questions, we are interested in defining *nonanticipative* policies. This means that the decision we make at any time step should be a function of the available history $\xi_{[t]}$ of the process at this time step. This is a necessary condition for a *policy* to be implementable since a decision has to be made on the basis of the available information. We will focus on models with recourse. More precisely, introducing a recourse function $\mathcal{Q}_{t+1}$ for time step t and given x_{t-1}, the decision x_t is found by solving the problem

$$\begin{aligned}&\inf_{x_t} f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}(x_t,\xi_{[t]})\\ &x_t \in \chi_t(x_{t-1},\xi_t)\end{aligned} \tag{2}$$

*Received by the editors October 14, 2010; accepted for publication (in revised form) December 23, 2011; published electronically April 4, 2012.
http://www.siam.org/journals/siopt/22-2/81169.html

†UFRJ, Escola Politécnica, Departamento de Engenharia Industrial Ilha do Fundão, CT, Bloco F, Rio de Janeiro, Brazil (vguigues@impa.br).
‡Institute of Mathematics, Humboldt-University Berlin, 10099 Berlin, Germany (romisch@math.hu-berlin.de).

at time step t. In this problem, we have assumed that ξ_t is available at time step t and thus $\xi_{[t]}$ gathers all the realizations of ξ_j up to time step t. The policy depends crucially on the choice of the recourse function $\mathcal{Q}_{t+1}$ used in (2). Given x_0 and the information $\xi_{[1]}$, a non-risk-averse model uses the recourse functions defined by

$$\mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}) = \mathbb{E}_{\xi_t | \xi_{[t-1]}} \left(\begin{array}{l} \inf_{x_t} f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) \\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right) \tag{3}$$

for $t = 1, \ldots, T$, with $\mathcal{Q}_{T+1} \equiv 0$. These dynamic programming (DP) equations are associated to the non-risk-averse model

$$\begin{array}{l} \inf \mathbb{E} \left[\sum_{t=1}^{T} f_t(x_t(\xi_{[t]}), \xi_t) \right] \\ x_t(\xi_{[t]}) \in \chi_t(x_{t-1}(\xi_{[t-1]}), \xi_t), \ t = 1, \ldots, T. \end{array} \tag{4}$$

For the second of these questions, most of the efforts so far have been placed on solution methods that approximate the recourse functions (3) in the case of multistage stochastic linear programs. In this paper, we contribute to these two questions as follows.

From the modeling point of view, we define risk-averse recourse functions. For this purpose, a common approach (Ruszczyński and Shapiro [RS06a], [RS06b]) is based on a risk-averse nested formulation of the problem using conditional (coherent) risk measures. In this situation, it is in general difficult, even for simple risk measures such as the *conditional value-at-risk* (CVaR) (Rockafellar and Uryasev [RU02]), to determine a risk-averse problem (using a risk measure that has a physical interpretation) whose stagewise decomposition is given by these DP equations. However, such an interpretation is important. This is why we define instead a risk-averse problem for (1) that is then decomposed by stages to obtain DP equations. A similar idea appears in the recent book by Shapiro, Dentcheva, and Ruszczyński [SDR09, Chapter 6, p. 326], where a convex combination of the expectation and of the CVaR of the final wealth is used for a portfolio selection problem. Instead, we control partial costs (the sum of the costs up to the current time step) and use a new class of risk measures that is suitable for decomposing the risk-averse problem by stages. This class of multiperiod risk measures called *extended polyhedral risk measures* has three appealing properties. First, the class is large: it contains the polyhedral risk measures (Eichhorn and Römisch [ER05]); in the one-period case some special cases include the optimized certainty equivalent (Ben-Tal and Teboulle [BTT07]), some spectral risk measures (Acerbi [Ace02]), and the CVaR. More generally, conditions for such functionals to be coherent or convex are provided. Second, as stated above, it allows us to define DP equations for our risk-averse problem. Finally, these equations are suitable for proposing convergent solution methods for a class of stochastic linear programs.

Regarding algorithmic issues, exact decomposition algorithms such as the nested decomposition (ND) algorithm have shown their superiority to direct solution methods for obtaining approximations of the recourse functions. Each iteration of these algorithms computes upper and lower bounds on the optimal mean cost. If an optimal solution to the problem exists, the algorithm finds an optimal solution after a finite number of iterations. These exact algorithms build at each iteration and each node of the scenario tree a cut for the recourse functions. These cuts form an outer linearization of these recourse functions.

There are two important variants of the ND algorithm: a variant that adds quadratic proximal terms in the objective functions of the master problems and a variant that uses multicuts (Ruszczyński [Rus86]).

The purpose of the first variant is to discourage the solution from moving too far from the best solution found so far, and this can significantly accelerate the convergence of the method even if the master problems are quadratic programs with this approach. The proximal term penalties are positive and can be dynamically modified in the course of the algorithm.

In the ND algorithm, for a given node in the scenario tree and a given input state x_{t-1} at t, the subproblems associated to all the realizations in stage $t+1$ are solved to obtain their optimal simplex multipliers. These multipliers are then aggregated to obtain a single cut for each node in each iteration. In the multicut variant, there are as many cuts as descendant realizations that are built at each iteration. More information is thus passed from the children nodes to their immediate ancestor by sending disaggregate cuts. The size of the master programs increases, but we expect fewer iterations (see Birge and Louveaux [BL88]).

However, in some applications, the number of scenarios may become so large that even these improved variants are difficult to apply since they entail prohibitive computational efforts.

Monte Carlo sampling-based algorithms constitute an interesting alternative in such situations. Higle and Sen [HS96] introduced a stochastic cutting plane method for two-stage stochastic programs and showed its convergence with probability one. Recently, Higle, Rayco, and Sen [HRS10] extended this idea to multistage models by applying a stochastic cutting plane method to the dual problem resulting when dualizing nonanticipativity constraints. Their method is, hence, based on scenario decomposition. A different approach for two-stage problems based on Monte Carlo (importance) sampling within the L-shaped method was introduced by Dantzig and Glynn [DG90] and Infanger [Inf92]. For multistage stochastic linear programs whose number of immediate descendant nodes is small but with many stages, Pereira and Pinto [PP91] proposed sampling in the forward pass of the ND. This sampling-based variant of the ND is the so-called stochastic dual dynamic programming algorithm on which we focus our attention. More precisely, we detail a *stochastic dual dynamic programming* (SDDP) algorithm (Pereira and Pinto [PP91]) to approximate our risk-averse recourse functions, to be used in (2) in place of $\mathcal{Q}_{t+1}$. The computation of the cuts in the backward pass of SDDP are detailed in this risk-averse setting.

Our developments can be easily extended to other sampling-based decomposition methods such as AND and DOASA.

The abridged nested decomposition (AND) algorithm proposed by Birge and Donohue [BD06] is a variant of SDDP that also involves sampling in the forward pass. This algorithm determines in a different manner the sequence of states and scenarios in the forward pass. The numerical simulations in Birge and Donohue [BD06] report lower computational time on average for the AND algorithm in comparison with SDDP.

When the number of immediate descendant nodes is large (possibly infinite) and when the problem has many stages, we also can (or even must) sample in the backward pass. In this case, for a given node on a forward path k, not all the optimal simplex multipliers associated to the descendant subproblems are computed. Only the descendant subproblems associated with some realizations are solved. As explained in the cut calculation algorithm (CCA) in Philpott and Guan [PG08], it is, however, possible in this situation to replace the “missing” multipliers by some coefficients so that the cuts built still lie below the corresponding recourse functions. This gives rise to dynamic outer approximation sampling algorithms (DOASA) described in Philpott and Guan [PG08].

The paper is organized as follows. In the second section, we introduce the class of multiperiod extended polyhedral risk measures and study their properties: dual representations are derived and used to provide criteria for convexity and coherence and, in the one-period case, for convexity and consistency with second order stochastic dominance. In section 3, we derive DP equations for a risk-averse problem defined in terms of extended polyhedral risk measures. We also provide conditions that guarantee the convergence of SDDP in this risk-averse setting. Finally, in section 4, we propose to use SDDP to approximate the risk-averse recourse functions from section 3 for some stochastic linear programs. In particular, formulas for the cuts in the backward pass are given.

We mention that after writing our paper we became aware of two recent and closely related papers: Collado, Papp, and Ruszczyński [CPR], based on scenario decomposition, and Shapiro [Sha11], which suggests using SDDP to approximate risk-averse recourse functions defined from a nested risk-averse formulation of a multistage stochastic program.

We start by setting down some notation:

- For $x \in \mathbb{R}^n$, the vectors x^+ and x^- are defined by $x^+(i) = \max(x(i), 0)$ and $x^-(i) = \max(-x(i), 0)$ for $i = 1, \ldots, n$.
- For a nonempty set $X \subseteq \mathbb{R}^n$, the polar cone X^* is defined by $X^* = \{x^* : \langle x, x^* \rangle \leq 0 \; \forall x \in X\}$, where $\langle \cdot, \cdot \rangle$ is the standard scalar product on $\mathbb{R}^n$.
- e is a column vector of all ones.
- If A is an $m_1 \times n$ matrix and B an $m_2 \times n$ matrix, $(A; B)$ denotes the $(m_1 + m_2) \times n$ matrix $\left(\begin{smallmatrix} A \\ B \end{smallmatrix}\right)$.
- For vectors $x_1, \ldots, x_T \in \mathbb{R}^n$ and $1 \leq t_1 \leq t_2 \leq T$, we denote $(x_{t_1}, \ldots, x_{t_2}) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ by $x_{t_1:t_2}$.
- For $x, y \in \mathbb{R}^n$, the vector $x \circ y \in \mathbb{R}^n$ is defined by $(x \circ y)(i) = x(i)y(i),\ i = 1, \ldots, n$.
- I_n is the $n \times n$ identity matrix, and $0_{m,n}$ is an $m \times n$ matrix of zeros.
- δ_{ij} is the Kronecker delta defined for i, j integers by $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.
- $\mathcal{Q}_{t+1}$ denotes a (generic) recourse function used at time step $t = 1, \ldots, T$, i.e., $\mathcal{Q}_{T+1} \equiv 0$, and if $t < T$, then $\mathcal{Q}_{t+1}(x_t, \xi_{[t]})$ represents a cost over the period $t + 1, \ldots, T$. Various recourse functions at t will be defined using the same notation $\mathcal{Q}_{t+1}$. Which $\mathcal{Q}_{t+1}$ is relevant will be clear from the context.

As is usually done in the stochastic programming literature and to alleviate notation, we use the same notation for a random variable and for a particular realization of this random variable, the context allowing us to know which concept is being referred to.

2. Extended polyhedral risk measures. We consider multiperiod risk functionals ρ whose arguments are sequences of random variables. We confine ourselves to discrete-time processes with a finite time horizon as in Ruszczyński and Shapiro [RS06a]. Such risk functionals have to assess the riskiness of a finite sequence $z_1, \ldots, z_T$ of random variables for some integer $T \geq 2$. To reflect the evolution of information as time goes by, we assume that z_t is measurable with respect to some σ-field $\mathcal{F}_t$, where $\mathcal{F}_1, \ldots, \mathcal{F}_T$ form a filtration, i.e., $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_T = \mathcal{F}$, with $\mathcal{F}_1 = \{\emptyset, \Omega\}$. In this setting, z_1 is deterministic, and a multiperiod risk functional ρ will be seen as a mapping $\rho : \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P}) \to \bar{\mathbb{R}}$ for some $p \in [1, +\infty)$.

Remark 2.1. Throughout the paper, the arguments $(z_1,\dots,z_T)$ of the risk functionals will be interpreted as accumulated revenues (for which higher values are preferred). More precisely, if $\tilde z_t$ is the revenue for time step t, we consider the accumulated revenues $z_t=\sum_{\tau=1}^t \tilde z_\tau,\ t=1,\dots,T$.

For future use, we recall the definition of multiperiod convex risk measures (from Artzner et al. [ADE$^+$], [ADE$^+$07], Föllmer and Schied [FS04]) which are multiperiod risk functionals of special interest when the random variables z_t represent revenues (accumulated or not).

DEFINITION 2.2. *A functional ρ on $\times_{t=1}^T L_p(\Omega,\mathcal{F}_t,\mathbb{P})$ is called a* multiperiod convex risk measure *if conditions* (i)–(iii) *below hold:*

(i) Monotonicity*: if $z_t\le\tilde z_t$ a.s, $t=1,\dots,T$, then $\rho(z_1,\dots,z_T)\ge\rho(\tilde z_1,\dots,\tilde z_T)$.*
(ii) Translation invariance*: for each $r\in\mathbb{R}$ we have $\rho(z_1+r,\dots,z_T+r)=\rho(z_1,\dots,z_T)-r$.*
(iii) Convexity*: for each $\lambda\in[0,1]$ and $z,\tilde z\in\times_{t=1}^T L_p(\Omega,\mathcal{F}_t,\mathbb{P})$ we have $\rho(\lambda z+(1-\lambda)\tilde z)\le\lambda\rho(z)+(1-\lambda)\rho(\tilde z)$.*

It is called a multiperiod coherent risk measure *if in addition condition* (iv) *holds:*

(iv) Positive homogeneity*: for each $\lambda\ge 0$ we have $\rho(\lambda z_1,\dots,\lambda z_T)=\lambda\rho(z_1,\dots,z_T)$.*

In the literature, there appear different requirements instead of the translation invariance (ii) above, e.g., Fritelli and Scandalo [FS05] and Pflug and Römisch [PR07].

Convex duality can be used to obtain dual representations of multiperiod convex risk measures. Next, we cite such a representation that uses the set $\mathcal{D}_T$ of generalized density functions where

$$\mathcal{D}_T:=\left\{\lambda\in\times_{t=1}^T L_1(\Omega,\mathcal{F}_t,\mathbb{P})\ :\ \lambda_t\ge 0 \text{ a.s.}, t=1,\dots,T,\ \sum_{t=1}^T\mathbb{E}[\lambda_t]=1\right\}.$$

THEOREM 2.3. *Let $\rho:\times_{t=1}^T L_p(\Omega,\mathcal{F}_t,\mathbb{P})\to\bar{\mathbb{R}}$ and assume that ρ is proper (i.e., ρ is finite on the nonempty set* $\operatorname{dom}\rho=\{z:\rho(z)<\infty\}$*) and lower semicontinuous. Then ρ is a multiperiod convex risk measure if and only if it admits the representation*

$$\rho(z)=\sup\left\{\mathbb{E}\left(-\sum_{t=1}^T\lambda_t z_t\right)-\rho^*(\lambda)\ :\ \lambda\in\mathcal{P}_\rho\right\}\tag{5}$$

for some convex closed subset $\mathcal{P}_\rho\subseteq\mathcal{D}_T$ of the space $\times_{t=1}^T L_q(\Omega,\mathcal{F}_t,\mathbb{P})$ $(\frac{1}{p}+\frac{1}{q}=1)$ on which the conjugate ρ^* of ρ is given too. The functional ρ is coherent if and only if the conjugate ρ^* in (5) *is the indicator function of $\mathcal{P}_\rho$.*

A proof of the above theorem can be found in, e.g., Ruszczyński and Shapiro [RS06b]. We are now in a position to define the class of multiperiod extended polyhedral risk measures.

DEFINITION 2.4. *A risk measure ρ on $\times_{t=1}^T L_p(\Omega,\mathcal{F}_t,\mathbb{P})$ is called* multiperiod extended polyhedral *if there exist matrices $A_t, B_{t,\tau}$, vectors a_t, c_t, and functions $h_t(z)=(h_{t,1}(z),\dots,h_{t,n_{t,2}}(z))^\top$ for given functions $h_{t,1},\dots,h_{t,n_{t,2}}:L_p(\Omega,\mathcal{F}_t,\mathbb{P})\to L_{p'}(\Omega,\mathcal{F}_t,\mathbb{P})$ with $1\le p'\le p$ such that*

$$\rho(z_1,\dots,z_T)=\left\{\begin{array}{l}\inf\ \mathbb{E}[\sum_{t=1}^T c_t^\top y_t]\\ y_t\in L_p(\Omega,\mathcal{F}_t,\mathbb{P};\mathbb{R}^{k_t}),\ t=1,\dots,T,\\ A_t y_t\le a_t \ a.s., t=1,\dots,T,\\ \sum_{\tau=0}^{t-1}B_{t,\tau}y_{t-\tau}=h_t(z_t)\ a.s.,\ t=2,\dots,T.\end{array}\right.\tag{6}$$

Another less general extension of polyhedral risk measures is due to Eichhorn [Eic07]. Like a multiperiod polyhedral risk measure (Eichhorn and Römisch [ER05]), a multiperiod extended polyhedral risk measure is given as the optimal value of a T-stage linear stochastic program where the arguments of the risk measure appear on the right-hand side of the dynamic constraints. Multiperiod polyhedral risk measures constitute a particular case with $a_t = 0,\ t = 2, \dots, T,\ B_{t,\tau}$ row vectors, and $h_t(z_t) = h_{t,1}(z_t) = z_t$ (i.e., $n_{t,2} = 1$).

We mention that multiperiod extended polyhedral risk measures satisfy two additional properties that were recently discussed in the literature: *information monotonicity* (see Kovacevic and Pflug [KP09]) and *time consistency*, suggested in Shapiro [Sha09]. Information monotonicity means that the risk $\rho(z_1, \dots, z_T)$ gets smaller if the available information expressed by the σ-fields $\mathcal{F}_t,\ t = 1, \dots, T$, increases. Since $\rho(z_1, \dots, z_T)$ is given by a risk-neutral multistage stochastic program, it is time consistent as stated at the beginning of Shapiro [Sha09, section 3].

The need to consider the extended versions from Definition 2.4 is twofold:

(i) *Modeling*: Some (popular) risk measures are extended polyhedral but not polyhedral in the sense of Eichhorn and Römisch [ER05] (see examples at the end of this section).

(ii) *Algorithmic issues*: As announced in the introduction, DP equations can be written for risk-averse versions of (1) defined in terms of extended polyhedral risk measures. Moreover, the convergence of a class of decomposition algorithms applied to the corresponding nested formulation of the risk-averse problem will be proved in section 3 for a subclass of extended polyhedral risk measures that contain some nonpolyhedral risk measures. For this subclass, we have $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$.

In view of (ii) above, extended polyhedral risk measures with $h_t(z_t) = z_t b_t + \tilde{b}_t$ play a particular role when algorithmic issues come into play. In the rest of this section, we study properties of such risk functionals. In this context, the matrices $A_t, B_{t,\tau}$ and the vectors $a_t, b_t, \tilde{b}_t$, and c_t are fixed and deterministic. They have to be chosen such that ρ exhibits desirable risk measure properties. In particular, conditions on these parameters for the corresponding extended polyhedral risk measure to be coherent are given in the Corollary 2.6 of Theorem 2.5, which follows. This theorem gives dual representations for stochastic program (6) when $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$. In what follows, the dimensions of a_t and b_t are, respectively, denoted by $n_{t,1}$ and $n_{t,2}$.

THEOREM 2.5. *Let ρ be a functional of the form (6) on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ with $p \in [1, \infty)$ and $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$. Assume*

(i) *complete recourse: $\{y_1 : A_1 y_1 \le a_1\} \neq \emptyset$ and, for every $t = 2, \dots, T$, it holds that $\{B_{t,0} y_t : A_t y_t \le a_t\} = \mathbb{R}^{n_{t,2}}$;*

(ii) *dual feasibility: $\{(u, v) : u \in \times_{t=1}^T \mathbb{R}^{n_{t,1}}, v \in \times_{t=2}^T \mathbb{R}^{n_{t,2}}, c_t + A_t^\top u_t + \sum_{\tau=\max(2,t)}^T B_{\tau,\tau-t}^\top v_{\tau-1} = 0,\ t = 1, \dots, T\} \neq \emptyset$.*

Then ρ is finite, convex, and continuous on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ and with $\frac{1}{p} + \frac{1}{q} = 1$ the following dual representation holds:

(7)
$$\rho(z) = \begin{cases} \sup\ -\mathbb{E}[\sum_{t=1}^T\ \lambda_{1,t}^\top a_t + \sum_{t=2}^T \lambda_{2,t-1}^\top (z_t b_t + \tilde{b}_t)] \\ \lambda_1 \in \times_{t=1}^T\ L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}),\ \lambda_2 \in \times_{t=2}^T\ L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \\ \lambda_{1,t} \ge 0 \ a.s.,\ t = 1, \dots, T, \\ c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^T\ B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1} | \mathcal{F}_t] = 0 \ a.s.,\ t = 1, \dots, T. \end{cases}$$

We also have

$$\rho(z) = \sup \left\{ \mathbb{E}\left[\sum_{t=1}^{T} z_t^* z_t\right] - \rho^*(z^*) \ : \ z^* \in \times_{t=1}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}) \right\}, \tag{8}$$

where ρ^* *is the conjugate of* ρ*. Next, for every* $z^* \in \mathrm{dom}(\rho^*)$*,* $\rho^*(z^*)$ *is given by*

$$\rho^*(z^*) = \left\{ \begin{array}{l} \inf \ \mathbb{E}[\sum_{t=1}^T \ \lambda_{1,t}^\top a_t + \sum_{t=2}^T \lambda_{2,t-1}^\top \tilde{b}_t] \\ \lambda_1 \in \times_{t=1}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \ \lambda_2 \in \times_{t=2}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \\ z_t^* = -\lambda_{2,t-1}^\top b_t \ a.s., \ t = 2, \ldots, T, \ \lambda_{1,t} \geq 0 \ a.s., \ t = 1, \ldots, T, \\ c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^T \ B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1} | \mathcal{F}_t] = 0 \ a.s., \ t = 1, \ldots, T, \end{array} \right. \tag{9}$$

where

$$\mathrm{dom}(\rho^*) = \left\{ \begin{array}{l} z^* \in \times_{t=1}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}) \ \textit{such that} \\ \exists \, \lambda_1 \in \times_{t=1}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \\ \lambda_2 \in \times_{t=2}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}) \ \textit{satisfying} \\ \lambda_{1,t} \geq 0 \ a.s., \ t = 1, \ldots, T, \\ c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^T \ B_{\tau,\tau-t}^\top \mathbb{E}[\lambda_{2,\tau-1} | \mathcal{F}_t] = 0 \ a.s., \\ t = 1, \ldots, T, \ \textit{and} \\ z_1^* = 0, \ z_t^* = -\lambda_{2,t-1}^\top b_t \ a.s., \ t = 2, \ldots, T \end{array} \right\}. \tag{10}$$

Proof. We use results on Lagrangian and conjugate duality. Consider the following Banach spaces and their duals:

$$\begin{array}{ll} E := \times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), & E^* = \times_{t=1}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), \\ Z := \times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P}), & Z^* = \times_{t=1}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}), \end{array}$$

with bilinear forms

$$\langle e, e^* \rangle_{E/E^*} = \sum_{t=1}^T \mathbb{E}[e_t^\top e_t^*] \quad \text{and} \quad \langle z, z^* \rangle_{Z/Z^*} = \sum_{t=1}^T \mathbb{E}[z_t z_t^*].$$

Let us introduce the Lagrange multipliers $\lambda_1 \in \times_{t=1}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}})$ (with $\lambda_1 \geq 0$ a.s.) and $\lambda_2 \in \times_{t=2}^T L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}})$ associated to the constraints of (6) and the Lagrangian

$$\begin{aligned} L(y, \lambda_1, \lambda_2) := \mathbb{E} & \left[\sum_{t=1}^T c_t^\top y_t + \lambda_{1,t}^\top (A_t y_t - a_t) \right. \\ & \left. + \sum_{t=2}^T \lambda_{2,t-1}^\top \left(\sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} - z_t b_t - \tilde{b}_t \right) \right] \\ = \mathbb{E} & \left[\sum_{t=1}^T (c_t + A_t^\top \lambda_{1,t} + \sum_{\tau=\max(2,t)}^T B_{\tau,\tau-t}^\top \lambda_{2,\tau-1})^\top y_t \right] \\ & + \mathbb{E} \left[-\sum_{t=1}^T \lambda_{1,t}^\top a_t - \sum_{t=2}^T \lambda_{2,t-1}^\top (z_t b_t + \tilde{b}_t) \right]. \end{aligned}$$

The dual functional is defined by

$$\theta(\lambda_1, \lambda_2) := \inf_{y \in E} L(y, \lambda_1, \lambda_2), \tag{11}$$

and the Lagrangian dual of (6) is the problem

$$\begin{aligned} \sup_{\lambda_1,\lambda_2} \big\{ \theta(\lambda_1, \lambda_2) \ : \ & \lambda_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \\ & \lambda_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \lambda_1 \geq 0 \text{ a.s.} \big\}. \end{aligned} \tag{12}$$

Due to Ruszczyński and Shapiro [RS03, Proposition 5, Chapter 1], the conditional expectation operator $\mathbb{E}[\cdot|\mathcal{F}_t]$ and the operation of minimization can be interchanged in (11), which gives for $\theta(\lambda_1, \lambda_2)$ the expression

$$\begin{aligned} & -\mathbb{E}\left[\sum_{t=1}^{T} \lambda_{1,t}^{\top} a_t + \sum_{t=2}^{T} \lambda_{2,t-1}^{\top}(z_t b_t + \tilde{b}_t) \right] \\ & \quad + \mathbb{E}\left[\sum_{t=1}^{T} \inf_{y_t \in \mathbb{R}^{k_t}} (c_t + A_t^{\top} \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^{\top} \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t])^{\top} y_t \right]. \end{aligned}$$

Next, $\inf_{y_t \in \mathbb{R}^{k_t}} (c_t + A_t^{\top} \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^{\top} \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t])^{\top} y_t$ is 0 if

$$c_t + A_t^{\top} \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^{\top} \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t] = 0$$

and $-\infty$ otherwise. The Lagrangian dual (12) can thus be expressed as

$$\begin{aligned} & \sup \ -\mathbb{E}\left[\sum_{t=1}^{T} \lambda_{1,t}^{\top} a_t + \sum_{t=2}^{T} \lambda_{2,t-1}^{\top}(z_t b_t + \tilde{b}_t) \right] \\ & \lambda_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \ \lambda_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}), \lambda_1 \geq 0 \text{ a.s.}, \\ & c_t + A_t^{\top} \lambda_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^{\top} \mathbb{E}[\lambda_{2,\tau-1}|\mathcal{F}_t] = 0 \text{ a.s.}, \ t = 1, \ldots, T. \end{aligned} \tag{13}$$

From weak duality and dual feasibility, we obtain $\rho(z) > -\infty$, and due to the complete recourse assumption $\rho(z) < +\infty$. It follows that $\rho(z)$ is finite. More precisely, dual feasibility and complete recourse imply that there is no duality gap: the optimal value of (6), i.e., $\rho(z)$, is the optimal value of (13). This shows (7).

Next, we use conjugate duality. Let us introduce the vectors $c = (c_1, \ldots, c_T)^{\top}$, $a = (a_1, \ldots, a_T)^{\top}$, and $\tilde{b} = (\tilde{b}_2, \ldots, \tilde{b}_T)^{\top}$ and the matrices

$$A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_T \end{pmatrix}, \qquad \mathcal{B} = \begin{pmatrix} 0 & b_2 & & \\ \vdots & & \ddots & \\ 0 & & & b_T \end{pmatrix},$$

and

$$B = \begin{pmatrix} B_{2,1} & B_{2,0} & 0 & \ldots & 0 \\ B_{3,2} & B_{3,1} & B_{3,0} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ B_{T,T-1} & B_{T,T-2} & B_{T,T-3} & \ldots & B_{T,0} \end{pmatrix}.$$

Let also $Y = \{y \in E \ : \ Ay(\omega) \le a$ for a.e. $\omega \in \Omega\}$ and

$$\begin{array}{rrcl} \varphi : & E \times Z & \to & \bar{\mathbb{R}} \\ & (y,z) & \to & \varphi(y,z) = \langle y, c \rangle_{E/E^*} + \delta_Y(y) + \delta_{\{0\}}(By - \mathcal{B}z - \tilde{b}), \end{array}$$

where δ denotes the indicator function taking values 0 and $+\infty$ only. Since Y is closed and convex, φ is lower semicontinuous and convex. With this notation, we can express $\rho(z)$ as $\rho(z) = \inf_{y \in E} \ \varphi(y,z)$ and, due to Bonnans and Shapiro [BS00, Proposition 2.143], ρ is convex. Since ρ is finite valued, [BS00, Proposition 2.152] guarantees the continuity of ρ. Since ρ is proper, convex, and lower semicontinuous, by the Fenchel–Moreau theorem we have that $\rho^{**} = \rho$, where ρ^{**} is the biconjugate of ρ, i.e.,

$$\rho(z) = \rho^{**}(z) = \sup \{\langle z, z^* \rangle_{Z/Z^*} - \rho^*(z^*) \ : \ z^* \in Z^*\}, \tag{14}$$

which is (8). Next, $\rho^*(z^*) = \varphi^*(0, z^*)$, where the conjugate φ^* of φ is given by

$$\begin{aligned} \varphi^*(y^*, z^*) &= \sup \{\langle y, y^* \rangle_{E/E^*} + \langle z, z^* \rangle_{Z/Z^*} - \varphi(y,z) \ : \ y \in E, \ z \in Z\} \\ &= \sup \{\langle y, y^* - c \rangle_{E/E^*} + \langle z, z^* \rangle_{Z/Z^*} \ : \ Ay \le a \text{ a.s.}, \ By = \mathcal{B}z + \tilde{b} \text{ a.s.}\}. \end{aligned}$$

It follows that

$$\rho^*(z^*) = \left\{ \begin{array}{l} \sup \ \mathbb{E}[\sum_{t=1}^{T} (z_t z_t^* - c_t^\top y_t)] \\ y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), \ z_t \in L_q(\Omega, \mathcal{F}_t, \mathbb{P}), \ t = 1, \ldots, T, \\ A_t y_t \le a_t \text{ a.s.}, \ t = 1, \ldots, T, \\ \sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} = z_t b_t + \tilde{b}_t \text{ a.s.}, \ t = 2, \ldots, T. \end{array} \right. \tag{15}$$

Due to (i) and (ii), complete recourse and dual feasibility also hold for (15) for every $z^* \in \operatorname{dom}(\rho^*)$, where $\operatorname{dom}(\rho^*)$ is given by (10). Using once again Lagrangian duality for problem (15), we obtain for $\rho^*(z^*)$ dual representation (9). ∎

Theorems 2.3 and 2.5 allow us to provide a criterion for an extended polyhedral risk measure to be multiperiod coherent.

COROLLARY 2.6. *Let ρ be a functional on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ of the form* (6) *with all a_t null and $h_t(z_t) = z_t b_t$ for some vector b_t. Let the conditions of Theorem* 2.5 *be satisfied (complete recourse and dual feasibility) and let*

$$\mathcal{M}_\rho = \left\{ \begin{array}{l} \lambda \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}) \ \textit{ such that there exist} \\ \mu_1 \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,1}}), \ \mu_2 \in \times_{t=2}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{n_{t,2}}) \ \textit{ satisfying} \\ \mu_{1,t} \ge 0 \textit{ a.s.}, \ t = 1, \ldots, T, \\ c_t + A_t^\top \mu_{1,t} + \sum_{\tau=\max(2,t)}^{T} B_{\tau,\tau-t}^\top \mathbb{E}[\mu_{2,\tau-1} | \mathcal{F}_t] = 0 \textit{ a.s.}, \\ t = 1, \ldots, T, \textit{ and} \\ \lambda_1 = 0, \ \lambda_t = \mu_{2,t-1}^\top b_t \textit{ a.s.}, \ t = 2, \ldots, T, \end{array} \right\}$$

be the (convex) set of dual multipliers. If $\mathcal{M}_\rho \subseteq \mathcal{D}_T$, then ρ is a multiperiod coherent risk measure.

Proof. Using representation (7) of Theorem 2.5, we can write $\rho(z) = \sup_{\lambda \in \mathcal{M}_\rho} - \sum_{t=1}^{T} \mathbb{E}[\lambda_t z_t]$. We conclude using Theorem 2.3 with $\mathcal{P}_\rho = \mathcal{M}_\rho$. ∎

Using representation (8) of Theorem 2.5, the properties of ρ can also be characterized by properties of $\operatorname{dom}(\rho^*)$, where $\operatorname{dom}(\rho^*)$ is given by (10).

COROLLARY 2.7. *Let ρ be a functional on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ of the form* (6) *with $h_t(z_t) = z_t b_t + \tilde{b}_t$ for some vectors $b_t, \tilde{b}_t$, and let the conditions of Theorem* 2.5 *be satisfied (complete recourse and dual feasibility). The following hold:*

(i) *ρ is* monotone $\Longleftrightarrow$ *for all $z^* \in \mathrm{dom}(\rho^*)$ we have $z_t^* \leq 0$ a.s. for $t = 1, \ldots, T$.*

(ii) *ρ is* translation invariant $\Longleftrightarrow$ *for all $z^* \in \mathrm{dom}(\rho^*)$ we have $\sum_{t=1}^{T} \mathbb{E}[z_t^*] = -1$.*

(iii) *ρ is* positively homogeneous $\Longleftrightarrow$ *for all $z^* \in \mathrm{dom}(\rho^*)$ we have $\rho^*(z^*) = 0$.*

When $T = 2$, since z_1 is deterministic, Definition 2.4 corresponds to one-period extended polyhedral risk measures that assess the riskiness of one random variable z only. For later reference we recall the definition of such risk measures which extend the class of one-period polyhedral risk measures.

DEFINITION 2.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $h(z) = (h_1(z), \ldots, h_{n_{2,2}}(z))^\top$ for given functions $h_1, \ldots, h_{n_{2,2}} : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to L_{p'}(\Omega, \mathcal{F}, \mathbb{P})$ with $1 \leq p' \leq p$. A risk measure ρ on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty)$ is called* extended polyhedral *if there exist matrices $A_1, A_2, B_{2,0}, B_{2,1}$, and vectors a_1, a_2, c_1, c_2 such that for every random variable $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$*

$$\rho(z) = \begin{cases} \inf\ c_1^\top y_1 + \mathbb{E}[c_2^\top y_2] \\ y_1 \in \mathbb{R}^{k_1},\ y_2 \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{k_2}), \\ A_1 y_1 \leq a_1,\ A_2 y_2 \leq a_2\ a.s., \\ B_{2,1} y_1 + B_{2,0} y_2 = h(z)\ a.s. \end{cases} \tag{16}$$

For one-period risk measures, dual representations from Theorem 2.5 specialize as follows.

COROLLARY 2.9. *Let ρ be a functional of the form* (16) *on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with some $p \in [1, \infty)$ and $h(z) = z b_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Assume*

(i) *complete recourse: $\{y_1 : A_1 y_1 \leq a_1\} \neq \emptyset$ and $\{B_{2,0} y_2 : A_2 y_2 \leq a_2\} = \mathbb{R}^{n_{2,2}}$;*

(ii) *dual feasibility: $\{(u, v) : u \in \mathbb{R}^{n_{1,1}} \times \mathbb{R}^{n_{2,1}}, v \in \mathbb{R}^{n_{2,2}}, c_t + A_t^\top u_t + B_{2,2-t}^\top v = 0,\ t = 1, 2\} \neq \emptyset$.*

Then ρ is finite, convex, continuous, and with $\frac{1}{p} + \frac{1}{q} = 1$, ρ admits the dual representation

$$\rho(z) = \begin{cases} \sup\ -\lambda_1^\top a_1 - \mathbb{E}[\lambda_2^\top a_2 + \lambda_3^\top (z b_2 + \tilde{b}_2)] \\ \lambda_1 \in \mathbb{R}^{n_{1,1}},\ \lambda_2 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}),\ \lambda_3 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}), \\ c_1 + A_1^\top \lambda_1 + B_{2,1}^\top \mathbb{E}[\lambda_3] = 0, \\ c_2 + A_2^\top \lambda_2 + B_{2,0}^\top \lambda_3 = 0\ a.s., \\ \lambda_1 \geq 0,\ \lambda_2 \geq 0,\ a.s. \end{cases}$$

We also have

$$\rho(z) = \sup \left\{ \mathbb{E}[z^* z] - \rho^*(z^*)\ :\ z^* \in\ L_q(\Omega, \mathcal{F}, \mathbb{P}) \right\}, \tag{17}$$

where ρ^* is the conjugate of ρ. Next, for every $z^* \in \mathrm{dom}(\rho^*)$, $\rho^*(z^*)$ is given by

$$\rho^*(z^*) = \begin{cases} \inf\ \mathbb{E}[\lambda_1^\top a_1 + \lambda_2^\top a_2 + \lambda_3^\top \tilde{b}_2] \\ \lambda_1 \in \mathbb{R}^{n_{1,1}},\ \lambda_2 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}),\ \lambda_3 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}), \\ z^* = -\lambda_3^\top b_2\ a.s.,\ \lambda_1 \geq 0,\ \lambda_2 \geq 0\ a.s., \\ c_1 + A_1^\top \lambda_1 + B_{2,1}^\top \mathbb{E}[\lambda_3] = 0, \\ c_2 + A_2^\top \lambda_2 + B_{2,0}^\top \lambda_3 = 0\ a.s., \end{cases} \tag{18}$$

where

$$\operatorname{dom}(\rho^*) = \left\{ \begin{array}{l} z^* \in L_q(\Omega, \mathcal{F}, \mathbb{P}) \textit{ such that there exist} \\ \lambda_1 \in \mathbb{R}^{n_{1,1}},\ \lambda_2 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}), \\ \lambda_3 \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}) \textit{ satisfying} \\ c_1 + A_1^\top \lambda_1 + B_{2,1}^\top \mathbb{E}[\lambda_3] = 0,\ \lambda_1 \geq 0,\ \lambda_2 \geq 0 \ a.s., \\ c_2 + A_2^\top \lambda_2 + B_{2,0}^\top \lambda_3 = 0 \ a.s., \textit{ and } \ z^* = -\lambda_3^\top b_2 \ a.s. \end{array} \right\}. \tag{19}$$

Proof. It suffices to use Theorem 2.5 with $T = 2$. ∎

Definition 2.2 specializes as follows to the one-period case.

DEFINITION 2.10. *A functional $\rho : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \bar{\mathbb{R}}$ is called a* convex risk measure *if it satisfies the following three conditions for all $z, \tilde{z} \in L_p(\Omega, \mathcal{F}, \mathbb{P})$:*

(i) Monotonicity*: if $z \leq \tilde{z}$ a.s., then $\rho(z) \geq \rho(\tilde{z})$.*

(ii) Translation invariance*: for each $r \in \mathbb{R}$ we have $\rho(z + r) = \rho(z) - r$.*

(iii) Convexity*: for all $\mu \in [0, 1]$ we have $\rho(\mu z + (1 - \mu)\tilde{z}) \leq \mu\rho(z) + (1 - \mu)\rho(\tilde{z})$.*

Such a functional ρ is said to be coherent *if it is positively homogeneous, i.e., $\rho(\mu z) = \mu\rho(z)$ for all $\mu \geq 0$ and $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$.*

Using Theorems 2.3 and Corollary 2.9, a sufficient criterion can be provided for a one-period extended polyhedral risk measure to be coherent.

COROLLARY 2.11. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form* (16) *with a_1, a_2 null, $p \in [1, \infty)$, and $h(z) = zb_2$ for some vector b_2. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and let $\mathcal{M}_\rho$ be the following (convex) set of dual multipliers:*

$$\mathcal{M}_\rho = \left\{ \begin{array}{l} \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}) \textit{ such that there exist} \\ (\mu_1, \mu_2, \mu_3) \in \mathbb{R}^{n_{1,1}} \times L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,1}}) \times L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_{2,2}}) \ \textit{ satisfying} \\ c_1 + A_1^\top \mu_1 + B_{2,1}^\top \mathbb{E}[\mu_3] = 0, \\ c_2 + A_2^\top \mu_2 + B_{2,0}^\top \mu_3 = 0 \ a.s.,\ \mu_1 \geq 0,\ \mu_2 \geq 0 \ a.s. \textit{ with } \ \lambda = \mu_3^\top b_2 \end{array} \right\}. \tag{20}$$

If $\mathcal{M}_\rho \subseteq \mathcal{D}_1$, then ρ is a (one-period) coherent risk measure.

Proof. From Corollary 2.9, we obtain $\rho(z) = \sup_{\lambda \in \mathcal{M}_\rho} -\mathbb{E}[\lambda z]$, and the result follows taking $\mathcal{P}_\rho = \mathcal{M}_\rho$ in Theorem 2.3. ∎

A dual representation of the second-stage problem for (16) will prove useful for obtaining further properties of one-period risk measures of the form (16).

PROPOSITION 2.12. *Let ρ be a functional of the form* (16) *on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with some $p \in [1, \infty)$ and $h(z) = zb_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility). Assume the feasible set $\mathcal{D}$ of the dual of the second-stage problem is nonempty where*

$$\mathcal{D} = \{\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^{n_{2,2}} \times \mathbb{R}^{n_{2,1}} \ : \ \lambda_2 \leq 0,\ \ B_{2,0}^\top \lambda_1 + A_2^\top \lambda_2 = c_2\}. \tag{21}$$

Then ρ is finite, convex, continuous and is given by

$$\rho(z) = \inf_{A_1 y_1 \leq a_1} \left\{ c_1^\top y_1 + \mathbb{E}\left[\sup_{\lambda \in \mathcal{D}} \ \lambda_1^\top (z b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2 a_2 \right] \right\}.$$

Proof. Finiteness, convexity, and continuity follow from Corollary 2.9. Next, we write $\rho(z)$ as

$$\rho(z) = \inf_{y_1} \{c_1^\top y_1 + \mathbb{E}[\mathcal{Q}_2(y_1, z)] \ : \ A_1 y_1 \leq a_1\}, \tag{22}$$

where for each y_1 such that $A_1 y_1 \leq a_1$ and for each $z \in \mathbb{R}$ we have defined

$$\mathcal{Q}_2(y_1, z) = \inf_{y_2} \{c_2^\top y_2 \ : \ B_{2,0} y_2 = z b_2 + \tilde{b}_2 - B_{2,1} y_1, \ A_2 y_2 \leq a_2\}.$$

Finally, since $\mathcal{D} \neq \emptyset$, by duality, we can express $\mathcal{Q}_2(y_1, z)$ as

$$\begin{aligned} \mathcal{Q}_2(y_1, z) = \sup_{(\lambda_1, \lambda_2)} \{&\lambda_1^\top (z b_2 + \tilde{b}_2 - B_{2,1} y_1) \\ &+ \lambda_2^\top a_2 \ : \lambda_2 \leq 0, \ B_{2,0}^\top \lambda_1 + A_2^\top \lambda_2 = c_2\}. \qquad \square \end{aligned} \tag{23}$$

The following proposition provides a sufficient criterion for some extended polyhedral risk measures to be convex risk measures when

$$Y_1 = \{y_1 : A_1 y_1 \leq a_1\} \tag{24}$$

is not necessarily a cone (a_1 need not be 0).

Proposition 2.13. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form* (16) *with $p \in [1, \infty)$ and $h(z) = z b_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and let $\mathcal{D}$ be defined as in Proposition* 2.12. *Assume*

(i) *$\mathcal{D} \neq \emptyset$ with $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$;*

(ii) *$c_1 \neq 0$ and b_2 is of the form $b_2 = -B_{2,1}^i / c_1(i)$ for at least one $i \in I = \{j \ : \ c_1(j) \neq 0\}$ with $y_1(i)$ unconstrained and where $B_{2,1}^i$ denotes the ith column of $B_{2,1}$.*

Then ρ is a finite-valued convex risk measure.

Proof. Let Y_1 be defined by (24). Finiteness and convexity of ρ follow from Corollary 2.9. The monotonicity of ρ follows from (i). Indeed, if $z, \tilde{z} \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ satisfy $z \leq \tilde{z}$ a.s., then for every $y_1 \in Y_1$ and every $(\lambda_1, \lambda_2) \in \mathcal{D}$ we have

$$\lambda_1^\top (z b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2 \geq \lambda_1^\top (\tilde{z} b_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2.$$

With the notation of Proposition 2.12 and with $\varphi(y_1, z) = c_1^\top y_1 + \mathbb{E}[\mathcal{Q}_2(y_1, z)]$, it follows that for every $y_1 \in Y_1$, we have $\mathbb{E}[\mathcal{Q}_2(y_1, z)] \geq \mathbb{E}[\mathcal{Q}_2(y_1, \tilde{z})]$, $\varphi(y_1, z) \geq \varphi(y_1, \tilde{z})$, and $\rho(z) = \inf_{y_1 \in Y_1} \varphi(y_1, z) \geq \inf_{y_1 \in Y_1} \varphi(y_1, \tilde{z}) = \rho(\tilde{z})$. The translation invariance condition follows from (ii). Indeed, eventually after reordering the components of y_1, c_1, and the columns of $B_{2,1}$, we can always assume that the index i satisfying (ii) is the last k_1th index, i.e., that $c_1, B_{2,1}$, and Y_1 are of the form $c_1 = (\hat{c}_1, \bar{c}_1)^\top$ with $\bar{c}_1 \in \mathbb{R}^*$, $B_{2,1} = [\hat{B}_{2,1}, -\bar{c}_1 b_2]$, and $Y_1 = \{y_1 = (\hat{y}_1, \bar{y}_1) \ : \ \hat{A}_1 \hat{y}_1 \leq a_1, \bar{y}_1 \in \mathbb{R}\}$. We then have for each $r \in \mathbb{R}$, for each $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, and setting $\tilde{y}_1 = \bar{y}_1 + \frac{r}{\bar{c}_1} \in \mathbb{R}$

$$\begin{aligned} \rho(z + r) = \inf_{\hat{A}_1 \hat{y}_1 \leq a_1, \ \bar{y}_1 \in \mathbb{R}} &\Bigg\{ \hat{c}_1^\top \hat{y}_1 + \bar{c}_1 \bar{y}_1 \\ &+ \mathbb{E}\Bigg[\sup_{(\lambda_1, \lambda_2) \in \mathcal{D}} \lambda_1^\top ((z + r) b_2 + \tilde{b}_2 - \hat{B}_{2,1} \hat{y}_1 + \bar{y}_1 \bar{c}_1 b_2) + \lambda_2^\top a_2 \Bigg] \Bigg\} \\ = \inf_{\hat{A}_1 \hat{y}_1 \leq a_1, \ \tilde{y}_1 \in \mathbb{R}} &\Bigg\{ \hat{c}_1^\top \hat{y}_1 + \bar{c}_1 \tilde{y}_1 \\ &+ \mathbb{E}\Bigg[\sup_{(\lambda_1, \lambda_2) \in \mathcal{D}} \lambda_1^\top (z b_2 + \tilde{b}_2 - \hat{B}_{2,1} \hat{y}_1 + \tilde{y}_1 \bar{c}_1 b_2) + \lambda_2^\top a_2 \Bigg] \Bigg\} - r \\ = \rho(z) - r. \qquad \square& \end{aligned}$$

Proposition 2.13 extends the corresponding result in Eichhorn and Römisch [ER05]. Proposition 2.14 below shows that condition (i) in Proposition 2.13 ensures in fact a stronger type of monotonicity than (i) in Definition 2.10. Such monotonicity is based on *stochastic dominance rules* (see Müller and Stoyan [MS02]). For real-valued random variables $z, \tilde{z} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$, stochastic dominance rules are defined by classes of measurable real-valued functions on $\mathbb{R}$. The stochastic dominance rule with respect to class $\mathcal{F}$ is defined by

$$z \preceq_{\mathcal{F}} \tilde{z} \quad :\Longleftrightarrow \quad \forall\, f \in \mathcal{F} \;:\; [\text{ if } \mathbb{E}[f(z)] \text{ and } \mathbb{E}[f(\tilde{z})] \text{ exist, then } \mathbb{E}[f(z)] \le \mathbb{E}[f(\tilde{z})]]$$

for each $z, \tilde{z} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$. Important special cases are the classes $\mathcal{F}_{nd}$ of nondecreasing functions and $\mathcal{F}_{ndc}$ of nondecreasing concave functions which, respectively, characterize first and second order stochastic dominance rules:

$$\begin{array}{lll} z \preceq_{FSD} \tilde{z} \;:\Longleftrightarrow z \preceq_{\mathcal{F}_{nd}} \tilde{z} & \Longleftrightarrow & \mathbb{P}(z \le t) \ge \mathbb{P}(\tilde{z} \le t) \quad \forall\, t \in \mathbb{R}, \\ z \preceq_{SSD} \tilde{z} \;:\Longleftrightarrow z \preceq_{\mathcal{F}_{ndc}} \tilde{z} & \Longleftrightarrow & \mathbb{E}[\min(z,t)] \le \mathbb{E}[\min(\tilde{z},t)] \quad \forall\, t \in \mathbb{R}. \end{array}$$

In particular, it is said that a risk measure ρ is *consistent with second order stochastic dominance* (see Ogryczak and Ruszczyński [OR02]) if $z \preceq_{SSD} \tilde{z}$ implies $\rho(z) \ge \rho(\tilde{z})$.

PROPOSITION 2.14. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathcal{P})$ of the form* (16) *with $p \in [1, \infty)$ and $h(z) = zb_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and let $\mathcal{D}$ be defined as in Proposition* 2.12. *Assume $\mathcal{D} \neq \emptyset$ with $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$. Then ρ is consistent with second order stochastic dominance.*

Proof. With Y_1 defined as in (24), let g be the function defined for every $y_1 \in Y_1$ and $z \in \mathbb{R}$ by

$$g(y_1, z) = c_1^\top y_1 + \sup_{(\lambda_1,\lambda_2) \in \mathcal{D}} \{\lambda_1^\top (zb_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2\}. \tag{25}$$

For every $y_1 \in Y_1$, $g(y_1, \cdot)$ is convex and, since $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$, it is also nonincreasing. Let $z \preceq_{SSD} \tilde{z}$. For every $y_1 \in Y_1$, since $-g(y_1, \cdot)$ is concave and nondecreasing, $\mathbb{E}[-g(y_1, z)] \le \mathbb{E}[-g(y_1, \tilde{z})]$ and $\rho(z) = \inf_{y_1 \in Y_1} \mathbb{E}[g(y_1, z)] \ge \inf_{y_1 \in Y_1} \mathbb{E}[g(y_1, \tilde{z})] = \rho(\tilde{z})$. ∎

For a one-period risk measure of the form (16) with $h(z) = zb_2 + \tilde{b}_2$ for some vectors $b_2, \tilde{b}_2$, the first-stage solution set $S(\rho(z)) \subseteq Y_1$ is given by

$$S(\rho(z)) = \left\{ y_1 \in Y_1 \;:\; \rho(z) = c_1^\top y_1 + \sup_{(\lambda_1,\lambda_2) \in \mathcal{D}} \{\lambda_1^\top (zb_2 + \tilde{b}_2 - B_{2,1} y_1) + \lambda_2^\top a_2\} \right\}. \tag{26}$$

For algorithmic issues considered in sections 3 and 4, it can be useful to have at hand conditions that guarantee the boundedness of $S(\rho(z))$. This question is addressed in the following proposition.

PROPOSITION 2.15. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form* (16) *with $p \in [1, \infty)$, a_2 null, and $h(z) = zb_2$ for some vector b_2. Let the conditions of Corollary* 2.9 *be satisfied (complete recourse and dual feasibility), and assume that $S(\rho(0))$ is nonempty and bounded. Then $S(\rho(z))$ is nonempty, convex, and compact for any $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. The proof follows the proof of Proposition 2.9 in Eichhorn and Römisch [ER05], with, in our case, g given by (25). ∎

We provide examples of extended polyhedral risk measures. The above criteria for coherence and consistency with second order stochastic dominance are applied.

Example 2.16 (spectral risk measures and CVaR). Let $F_z(x) = \mathbb{P}(z \le x)$ be the distribution function of random variable z, and let $F_z^{\leftarrow}(p) = \inf\{x \;:\; F_z(x) \ge p\}$

be the usual generalized inverse of F_z. Given a risk spectrum $\phi \in L_1([0,1])$ the *spectral risk measure* ρ_ϕ generated by ϕ is given by Acerbi [Ace02]:

$$\rho_\phi(z) = -\int_0^1 F_z^{\leftarrow}(p)\phi(p)dp.$$

Spectral risk measures have been used in a number of applications (portfolio selection in Acerbi and Simonetti [AS], and insurance in Cotter and Kevin [CD06]). The conditional value-at-risk (CVaR) of level $0 < \varepsilon < 1$, also called average value-at-risk (AVaR) in Föllmer and Schied [FS04], is a particular spectral risk measure with a piecewise constant risk function ϕ having a jump at ε: $\phi(u) = \frac{1}{\varepsilon}1_{0\le u\le\varepsilon}$ (Acerbi [Ace02]). Let us consider more generally a piecewise constant risk function $\phi(\cdot)$ with J jumps at $0 < p_1 < p_2 < \cdots < p_J < 1$. Setting $\Delta\phi_k = \phi(p_k^+) - \phi(p_k^-) = \phi(p_k) - \phi(p_{k-1})$ for $k = 1,\ldots,J$, with $p_0 = 0$, we assume

$$\text{(i) } \phi(\cdot) \text{ is positive,} \quad \text{(ii) } \Delta\phi_k < 0,\ k = 1,\ldots,J, \quad \text{(iii) } \int_0^1 \phi(u)du = 1.$$

With this choice of ϕ, we can express $\rho_\phi(z)$ as the optimal value of a linear programming problem (see Acerbi and Simonetti [AS]):

$$\rho_\phi(z) = \inf_{x\in\mathbb{R}^J} \sum_{k=1}^{J} \Delta\phi_k[p_k x_k - \mathbb{E}\,[x_k - z]^+] - \phi(1)\mathbb{E}[z]. \tag{27}$$

When $J = 1$, $\Delta\phi_1 = -1/\varepsilon$, $p_1 = \varepsilon$, and $\phi(1) = 0$, the above formula reduces to the formula for the CVaR given by Rockafellar and Uryasev [RU02]:

$$CVaR^\varepsilon[z] = \inf_{x\in\mathbb{R}} \left[x + \frac{1}{\varepsilon}\,\mathbb{E}[z + x]^-\right]. \tag{28}$$

A spectral risk measure with a piecewise constant risk function satisfying (i), (ii), and (iii) above is a coherent extended polyhedral risk measure. Indeed, with respect to (16), we have $c_1 = \Delta\phi \circ p$ with $\Delta\phi = (\Delta\phi_1,\ldots,\Delta\phi_J)^\top$, $c_2 = (-\Delta\phi; 0_{J,1}; -\phi(1))$, $B_{2,1} = (I_J; 0_{1,J}), B_{2,0} = (-I_J, I_J, 0_{J,1}; 0_{1,2J}, 1)$, $A_2 = (-I_{2J}, 0_{2J,1})$, and $h(z) = ze$. The matrix A_1 and the vectors a_1 and a_2 are null, b_2 is a $(J+1)$-vector of ones, and $\tilde{b}_2 = 0$. Notice that when $J > 1$ it is not polyhedral in the sense of Eichhorn and Römisch [ER05]. The complete recourse and dual feasibility assumptions from Corollary 2.9 are easily checked. This theorem provides for ρ_ϕ the dual representation

$$\rho_\phi(z) = \begin{cases} \sup\ -\mathbb{E}[\lambda z] \\ \lambda = \mu^\top e + \phi(1),\ \mu \in L_q(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^J), \\ \mathbb{E}[\mu] = -\Delta\phi \circ p,\ 0 \le \mu \le -\Delta\phi \text{ a.s.} \end{cases} \tag{29}$$

Let $\mathcal{M}_{\rho_\phi}$ be the set of dual multipliers from Corollary 2.11 for ρ_ϕ. For every $\lambda \in \mathcal{M}_{\rho_\phi}$, we have $\lambda \ge 0$ a.s. and

$$\begin{aligned}\mathbb{E}[\lambda] = \mathbb{E}[\phi(1) + \mu^\top e] &= \phi(1) - \sum_{i=1}^{J} \Delta\phi_i p_i = \phi(1) - \sum_{i=1}^{J} (\phi(p_i) - \phi(p_{i-1}))p_i \\ &= \phi(0)p_1 + \sum_{i=1}^{J-1} \phi(p_i)(p_{i+1} - p_i) + (1 - p_J)\phi(1) = \int_0^1 \phi(u)du = 1.\end{aligned}$$

It follows that $\mathcal{M}_{\rho_\phi} \subseteq \mathcal{D}_1$ and using Corollary 2.11, ρ_ϕ is a coherent one-period risk measure. Next, the set $\mathcal{D}$ in Proposition 2.14 is given by $\mathcal{D} = \{(\lambda_1, \lambda_2) \in \mathbb{R}^{J+1} \times \mathbb{R}^{2J} \ : \ \lambda_2 \leq 0, \ \lambda_{1,J+1} = -\phi(1), \ \lambda_{1,1:J} = \lambda_{2,J+1:2J}, \ \lambda_{1,1:J} = -\lambda_{2,1:J} + \Delta\phi\}$. For every $(\lambda_1, \lambda_2) \in \mathcal{D}$, we have $\lambda_1^\top b_2 = \lambda_1^\top e \leq 0$. It follows that $\mathcal{D} \subseteq \{b_2\}^* \times \mathbb{R}^{n_{2,1}}$ and due to Corollary 2.14, ρ_ϕ is consistent with second order stochastic dominance. When $J = 1$, $\Delta\phi_1 = -1/\varepsilon$, $p_1 = \varepsilon$, and $\phi(1) = 0$, $\rho_\phi = CVaR^\varepsilon$ and we recover results given in Eichhorn and Römisch [ER05]: the CVaR is consistent with second order stochastic dominance and is an extended polyhedral risk measure of the form (16) with $c_1 = 1$, $c_2 = (\frac{1}{\varepsilon}; 0)$, $B_{2,1} = -1$, $B_{2,0} = (-1, 1)$, $A_2 = -I_2$, $h(z) = z$, and A_1, a_1, a_2 null. The dual representation (29) becomes

$$CVaR^\varepsilon(z) = \sup\left\{-\mathbb{E}[\lambda z] : \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}), 0 \leq \lambda \leq \frac{1}{\varepsilon} \text{ a.s., } \mathbb{E}[\lambda] = 1\right\}.$$

Example 2.17 (optimized certainty equivalent (OCE) and expected utility). Given a concave nondecreasing utility function u, the optimized certainty equivalent $S_u(z)$ of the random variable z is defined in Ben-Tal and Teboulle [BTT07] by $S_u(z) = \sup_{y_1 \in \mathbb{R}} \ y_1 + \mathbb{E}[u(z - y_1)]$. Considering for u a piecewise affine concave function, we can express the convex function $-u$ as follows (see Rockafellar and Wets [RW98, Example 3.54]:

$$-u(x) = \inf\{c^\top y \ : \ y \in \mathbb{R}^k, y \geq 0, e^\top y = 1, \ b^\top y = x\} \tag{30}$$

for some vectors $b, c \in \mathbb{R}^k$. It follows that if u is a piecewise affine concave function, $\rho(z) = -S_u(z)$ is given by

$$\rho(z) = \left\{ \begin{array}{l} \inf \ -y_1 + \mathbb{E}[c^\top y_2] \\ y_1 \in \mathbb{R}, \ y_2 \in \mathbb{R}^k, \ y_2 \geq 0, \ e^\top y_2 = 1, \ b^\top y_2 = z - y_1. \end{array} \right. \tag{31}$$

In this case, the opposite of the OCE is an extended one-period polyhedral risk measure with h affine: $c_1 = -1$, $c_2 = c$, $A_2 = [-I_k; e^\top; -e^\top]$, $a_2 = [0_{k,1}; 1; -1]$, $B_{2,1} = 1$, $B_{2,0} = b^\top$, $b_2 = 1$, and A_1, a_1, and $\tilde{b}_2$ null. Notice that it is not polyhedral in the sense of Eichhorn and Römisch [ER05] and that complete recourse does not hold. However, properties of the OCE, given in Ben-Tal and Teboulle [BTT07], are easily checked: *monotonicity* follows from the definition of $-S_u$ and the fact that u is nondecreasing; *translation invariance* follows from the change of variable $\bar{y}_1 = y_1 - r$ in (31) (for $\rho(z + r)$) or in the definition of $-S_u(z + r)$; *convexity* can be checked directly from the definition of S_u (or using representation (31) and [BS00, Proposition 2.143], as in the proof of Theorem 2.5).

Let us consider as a special case a piecewise linear utility function of the form

$$u(x) = \gamma_1(x)^+ - \gamma_2(-x)^+, \text{ where } 0 \leq \gamma_1 < 1 < \gamma_2 \tag{32}$$

(note that $u(x) < x$ for $x \neq 0$). The corresponding risk measure $\rho(z) = -S_u(z)$ is an extended polyhedral risk measure with $c_1 = -1$, $c_2 = (-\gamma_1; \gamma_2)$, $B_{2,1} = 1, B_{2,0} = [1 - 1]$, $A_2 = -I_2$, $h(z) = z$, and A_1, a_2, a_2 null. Since complete (and even simple) recourse and dual feasibility hold, Corollary 2.9 provides the following dual representation:

$$\rho(z) = -S_u(z) = \sup\{-\mathbb{E}[\lambda z] : \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}), \ \mathbb{E}[\lambda] = 1, \ \gamma_1 \leq \lambda \leq \gamma_2 \text{ a.s.}\}.$$

Using Corollary 2.11, we deduce that when u is of the form (32), $\rho(z) = -S_u(z)$ is a coherent risk measure. More generally, it is shown in Ben-Tal and Teboulle [BTT07] that if u is a strongly risk-averse function (see Ben-Tal and Teboulle [BTT07]), $\rho(z) = -S_u(z)$ is coherent if and only if u is of the form (32). For $0 < \varepsilon < 1$, $CVaR^{\varepsilon}$ constitutes a particular case with $\gamma_1 = 0$ and $\gamma_2 = \frac{1}{\varepsilon}$. The set $\mathcal{D}$ in Proposition 2.14 is given by $\mathcal{D} = \{(\lambda_1, \lambda_2) : -\gamma_2 \leq \lambda_1 \leq -\gamma_1,\ \lambda_2 \leq 0\}$. Since for every $(\lambda_1, \lambda_2) \in \mathcal{D}$ we have $\lambda_1^\top b_2 = \lambda_1^\top e \leq 0$, using Proposition 2.14 we conclude that $-S_u(z)$ is consistent with second order stochastic dominance.

For any concave utility function u, the risk measure $\rho(z) = -\mathbb{E}(u(z))$ is an extended polyhedral risk measure with $h = u$, $B_{2,0} = c_2 = 1$, while the other parameters are null. In the particular case when u is a piecewise affine concave function, representation (30) shows that $-\mathbb{E}(u(z))$ can be written as an extended polyhedral risk measure with $h(z) = z$ and that complete recourse does not hold. However, a dual representation of ρ can be derived from the dual representation

$$-u(x) = \sup\{-\lambda_1 x - \lambda_2 : \lambda \in \mathbb{R}^2,\ \lambda_1 b + \lambda_2 e \leq -c\} \tag{33}$$

of $-u$. Applying the expectation operator to both sides of the above equation and using Rockafellar and Wets [RW98, Theorem 14.60] (for switching the inf and expectation operators), we obtain for ρ the dual representation

$$\rho(z) = \sup\{-\mathbb{E}[\lambda_1 z + \lambda_2] : \lambda \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^2),\ \lambda_1 b + \lambda_2 e \leq -c \text{ a.s.}\}.$$

Since $-u$ is nonincreasing, for every (λ_1, λ_2) in the feasible set of (33) we have $\lambda_1 \geq 0$ (otherwise, there would be positive subgradients of $-u$ at large enough points). It follows that in the above representation of ρ, $\lambda_1 \geq 0$ a.s., which implies that ρ is monotone, convex, and consistent with second order stochastic dominance. The *expected regret* or *expected loss* $\rho(z) = \mathbb{E}(z - \beta)^-$ for some target β is a special case (already considered in Eichhorn and Römisch [ER05]) with utility function $u(z) = -(z - \beta)^-$. Finally, notice that $\rho(z) = \mathbb{E}[(z - \mathbb{E}[z])^k]$ for some $1 \leq k \leq p - 1$ is an extended polyhedral risk measure with $h(z) = (z - \mathbb{E}[z])^k$.

Example 2.18 (multiperiod extended polyhedral risk measures). We consider functionals ρ on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ ($p \in [1, \infty)$) of the form $\rho(z) = \rho_\phi(\Phi(z))$, where ρ_ϕ is a spectral risk measure of form (27) with $\phi(\cdot)$ satisfying (i), (ii), (iii) in Example 2.16, and the function Φ is defined on $\mathbb{R}^T$ and maps to the extended real numbers.

Then ρ is a finite-valued coherent multiperiod risk measure if the function Φ (i) is concave, (ii) is monotone with respect to the (canonical) partial ordering in $\mathbb{R}^T$, (iii) is positively homogeneous, (iv) satisfies the property $\Phi(\zeta_1 + r, \ldots, \zeta_T + r) = \Phi(\zeta_1, \ldots, \zeta_T) + r$ for all $r \in \mathbb{R}$ and $\zeta \in \mathbb{R}^T$, and (v) has linear growth; i.e., for some constant $L > 0$ it holds $|\Phi(\zeta)| \leq L \sum_{t=1}^T |\zeta_t|$ for every $\zeta \in \mathbb{R}^T$.

There are three important special cases of the function Φ:

(a) $\Phi(\zeta) = \sum_{t=1}^T \gamma_t \zeta_t$ with $\gamma_t \geq 0$, $t = 1, \ldots, T$, such that $\sum_{t=1}^T \gamma_t = 1$. Using (27), we have

$$\begin{aligned}
\rho(z) &= \rho_\phi\left(\sum_{t=1}^T \gamma_t z_t\right)\\
&= \inf_{x \in \mathbb{R}^J} (\Delta\phi \circ p)^\top x + \mathbb{E}\ \left(-\sum_{k=1}^J \Delta\phi_k \left[x_k - \sum_{t=1}^T \gamma_t z_t\right]^+ - \phi(1) \sum_{t=1}^T \gamma_t z_t\right)\\
&= \left\{\begin{array}{l}
\inf\ (\Delta\phi \circ p)^\top x + \mathbb{E}\left(-\sum_{k=1}^J \Delta\phi_k w_k - \phi(1) v_T\right)\\
x \in \mathbb{R}^J,\ v_t = v_{t-1} + \gamma_t z_t,\ v_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}),\ t = 1, \ldots, T,\ v_0 = 0,\\
w_k \geq 0,\ \ w_k \geq x_k - v_T,\ w_k \in L_p(\Omega, \mathcal{F}_T, \mathbb{P}),\ k = 1, \ldots, J.
\end{array}\right.
\end{aligned}$$

The stochastic program above can be rewritten in the form (6), and ρ is a multiperiod extended polyhedral coherent risk measure. In the case when $\rho_\phi = CVaR^\varepsilon$, according to the dual representation of $CVaR^\varepsilon$, we obtain

$$\rho(z) = \sup\left\{-\sum_{t=1}^{T}\mathbb{E}(\lambda_t z_t) : \lambda_t \in L_q(\Omega, \mathcal{F}_t, \mathbb{P}),\, \mathbb{E}(\lambda_t) = \gamma_t,\, 0 \le \lambda_t \le \frac{\gamma_t}{\varepsilon},\, t = 1, \ldots, T,\right.$$
$$\left.\gamma_t \mathbb{E}(\lambda_{t+1}|\mathcal{F}_t) = \gamma_{t+1}\lambda_t \text{ a.s., } t = 1, \ldots, T-1\right\},$$

where $\lambda_t = \gamma_t \mathbb{E}(\lambda|\mathcal{F}_t)$, $t = 1, \ldots, T$, and $\frac{1}{p} + \frac{1}{q} = 1$. Hence, ρ is a multiperiod extended polyhedral coherent risk measure according to Theorems 2.3 and 2.5.

(b) $\Phi(\zeta) = \min_{\gamma\in S}\langle\gamma, \zeta\rangle = \min_{\gamma\in S}\sum_{t=1}^{T}\gamma_t\zeta_t$, where S denotes the standard simplex $S = \{\gamma \in \mathbb{R}^T : \gamma_t \ge 0, t = 1, \ldots, T, \sum_{t=1}^{T}\gamma_t = 1\}$, may be used instead of the function Φ in (a). This function satisfies conditions (i)–(v), but avoids specifying the weights γ_t, $t = 1, \ldots, T$.

(c) $\Phi(\zeta) = \min_{t=1,\ldots,T}\zeta_t$ for $\zeta \in \mathbb{R}^T$. Using representation (27), we obtain

$$\begin{aligned}
\rho(z) &= \rho_\phi\Big(\min_{t=1,\ldots,T} z_t\Big)\\
&= \inf_{x\in\mathbb{R}^J} (\Delta\phi \circ p)^\top x + \mathbb{E}\left(-\sum_{k=1}^{J}\Delta\phi_k\Big[x_k - \min_{t=1,\ldots,T} z_t\Big]^+ - \phi(1)\min_{t=1,\ldots,T} z_t\right)\\
&= \inf_{x\in\mathbb{R}^J} (\Delta\phi \circ p)^\top x + \mathbb{E}\left(-\sum_{k=1}^{J}\Delta\phi_k \max_{t=1\ldots,T}(0, x_k - z_t) + \phi(1)\max_{t=1,\ldots,T} -z_t\right)\\
&= \begin{cases}
\inf\, (\Delta\phi \circ p)^\top x + \mathbb{E}\Big(-\sum_{k=1}^{J}\Delta\phi_k v_{kT} + \phi(1)v_T\Big)\\
x \in \mathbb{R}^J,\ v_1 \ge -z_1,\ v_t \ge v_{t-1},\ v_t \ge -z_t,\ t = 2, \ldots, T,\\
v_{kt} \ge v_{kt-1},\ v_{kt} \ge x_k - z_t,\ v_t, v_{k,t} \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}),\\
k = 1, \ldots, J,\ t = 1, \ldots, T,\ v_{k0} = 0.
\end{cases}
\end{aligned}$$

The latter linear stochastic program may be rewritten in the form (6), and ρ is a multiperiod extended polyhedral coherent risk measure. In the case when $\rho_\phi = CVaR^\varepsilon$, we obtain

$$\rho(z) = \inf\left\{x + \frac{1}{\varepsilon}\mathbb{E}(v_T) : v_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}), -x - z_t \le v_t, v_{t-1} \le v_t,\right.$$
$$\left. t = 1, \ldots, T, v_0 = 0, x \in \mathbb{R}\right\}. \tag{34}$$

Example (34) was first studied by Eichhorn in [Eic07].

3. Risk-averse dynamic programming.

3.1. General setting. When using a multiperiod extended polyhedral risk measure to deal with uncertainty in the multistage stochastic programming framework (4), we consider accumulated revenues $z_t = -\sum_{\tau=1}^{t} f_\tau(x_\tau, \xi_\tau)$ and the sigma-algebras $\mathcal{F}_t = \sigma(\xi_j, j \le t)$ for $t = 1, \ldots, T$. Recall that x_0 and $\chi_1(x_0, \xi_1)$ are deterministic and that for any time step $t = 1, \ldots, T$, we denote by $\xi_{[t]}$ the available realizations of the process up to this time step, i.e., $\xi_{[t]} = (\xi_j, j \le t)$.

We also denote by $\mathcal{Z}_t$ the space of $\mathcal{F}_t$-measurable functions (these sets are embedded: $\mathcal{Z}_1 \subset \cdots \subset \mathcal{Z}_T$). Next, for $t = 1, \ldots, T$, we assume the following:

(H1) the functions $f_t : \mathbb{R}^{N_{t,x}} \times \mathbb{R}^{M_t} \to \mathbb{R}$ are continuous and $\chi_t : \mathbb{R}^{N_{t-1,x}} \times \mathbb{R}^{M_t} \rightrightarrows \mathbb{R}^{N_{t,x}}$ are measurable, bounded, and closed-valued multifunctions.

We are now in a position to define a risk-averse problem for (1) via a multiperiod risk measure. Let $\rho : \mathcal{Z}_1 \times \ldots \mathcal{Z}_T \to \mathbb{R}$ be a multiperiod risk measure and let us introduce the risk-averse problem

$$
\begin{aligned}
&\inf \ \rho\left(-f_1(x_1,\xi_1), -\sum_{\tau=1}^{2} f_\tau(x_\tau(\xi_{[\tau]}),\xi_\tau), \ldots, -\sum_{\tau=1}^{T} f_\tau(x_\tau(\xi_{[\tau]}),\xi_\tau)\right)\\
&x_t(\xi_{[t]}) \in \chi_t(x_{t-1}(\xi_{[t-1]}),\xi_t),\ t = 1,\ldots,T.
\end{aligned} \tag{35}
$$

In the above problem, the optimization is performed over $\mathcal{F}_t$-measurable functions $x_t, t = 1, \ldots, T$, satisfying the constraints and such that $f_t(x_t(\cdot), \cdot) \in \mathcal{Z}_t$. The sequence of measurable mappings $x_t(\cdot)$, $t = 1, \ldots, T$, is called a policy. The $\mathcal{F}_t$-measurability of $x_t(\cdot)$ implies the nonanticipativity of the policy, i.e., implies that x_t is a function of $\xi_{[t]}$. The policy obtained from (35) will be said to be risk-averse. A policy is said to be *feasible* if the constraints $x_t(\xi_{[t]}) \in \chi_t(x_{t-1}(\xi_{[t-1]}), \xi_t)$, $t = 1, \ldots, T$, are satisfied with probability one.

In this section, our objective is to provide a class of form (1) problems and a class of multiperiod risk measures ρ having the following two properties:

(P1) DP equations can be written for (35).

(P2) The SDDP algorithm applied to problem (35) decomposed by stages converges to an optimal solution of (35).

We intend to enforce (P2) obtaining DP equations that satisfy conditions given in Philpott and Guan [PG08]. These conditions imply the following:

(P3) The recourse functions are given as the optimal value of a non-risk-averse stochastic program (the objective function is an expectation) where the randomness appears on the right-hand side of the constraints only.

Property (P3) leads us naturally to use the class of extended polyhedral risk measures introduced in the previous section.

3.2. Extended polyhedral risk measures. Taking for ρ a multiperiod extended polyhedral risk measure of the form (6), problem (35) can be written as

$$
\begin{aligned}
&\inf \ \mathbb{E}\left[\sum_{t=1}^{T} c_t^\top y_t\right]\\
&A_t y_t \le a_t \text{ a.s.},\ t = 1,\ldots,T,\\
&\sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} = h_t\left(-\sum_{\tau=1}^{t} f_\tau(x_\tau,\xi_\tau)\right) \text{a.s.},\ t = 2,\ldots,T,\\
&x_t \in \chi_t(x_{t-1},\xi_t) \text{ a.s.},\ t = 1,\ldots,T.
\end{aligned} \tag{36}
$$

Remark 3.1. In (36), the dependence of x_t and y_t with respect to $\xi_{[t]}$ was suppressed to alleviate notation. This will in general be done in what follows.

We first check that (P1) and (P3) hold for problem (36) above. Since we want to write DP equations, we start with the following simple remark.

Remark 3.2. Let us consider the following T-stage optimization problem:

$$
P\left\{\begin{array}{l} \inf\ f(x_1,\ldots,x_T)\\ x_t \in X(x_0,\ldots,x_{t-1}),\ t = 1,\ldots,T. \end{array}\right.
$$

We decompose f as $f(x) = \sum_{k=1}^{T} f_k(x_{1:k})$, where f_k is the sum of all the functions in the sum of functions defining f which depend on x_k but not on $x_{k+1:T}$ (for a given k, f_k is 0 if no such functions exist). DP equations for P can be written as follows:

$$\mathcal{Q}_t(x_{0:t-1}) = \begin{cases} \inf\limits_{x_t} f_t(x_{1:t}) + \mathcal{Q}_{t+1}(x_{0:t}) \\ x_t \in X(x_{0:t-1}) \end{cases}$$

for $t = 1, \ldots, T$, with $\mathcal{Q}_{T+1} \equiv 0$.

The application of Remark 3.2 to (36) yields the following DP equations: for $t = 1, \ldots, T$, $\mathcal{Q}_t(x_{0:t-1}, \xi_{[t-1]}, y_{1:t-1})$ is given by

$$\begin{aligned} &\mathcal{Q}_t(x_{0:t-1}, \xi_{[t-1]}, y_{1:t-1}) \\ &\quad = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf_{x_t, y_t} \; c_t^\top y_t + \mathcal{Q}_{t+1}(x_{0:t}, \xi_{[t]}, y_{1:t}) \\ A_t y_t \le a_t, \\ (1-\delta_{t1}) \left(\sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} - h_t \left(-\sum_{\tau=1}^{t} f_\tau(x_\tau, \xi_\tau) \right) \right) = 0, \\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right), \end{aligned} \tag{37}$$

where here, and in what follows, $\mathcal{Q}_{T+1} \equiv 0$. Since these DP equations correspond to the stagewise decomposition of risk-averse problem (36), the recourse functions $\mathcal{Q}_t$ in (37) are said to be *risk-averse*. Compared to the DP equations of the original stochastic program, a new state variable y_t and new constraints for it appear in (37) at time t. They serve for computing the multiperiod extended polyhedral risk measure.

Let us now take as a special case for ρ the multiperiod risk measure defined by

$$\rho(z_1, \ldots, z_T) = -\theta_1 \mathbb{E}[z_T] + \sum_{t=2}^{T} \theta_t \rho^t(z_t) \tag{38}$$

for some nonnegative weights θ_t, $t = 1, \ldots, T$, summing to one ($\sum_{t=1}^{T} \theta_t = 1$) and for some one-period coherent extended polyhedral risk measures $\rho^t : \mathcal{Z}_t \to \mathbb{R}$, $t = 2, \ldots, T$.

Remark 3.3. We easily check that ρ in (38) is a multiperiod (coherent) extended polyhedral risk measure.

Observe that since ρ^t is coherent and z_1 deterministic, we have $\rho^t(z_t - z_1) = \rho^t(z_t) + z_1$, and $\rho(z_1, \ldots, z_T)$ in (38) can be expressed as $\rho(z_1, \ldots, z_T) = -z_1 - \theta_1 \mathbb{E}[z_T - z_1] + \sum_{t=2}^{T} \theta_t \rho^t(z_t - z_1)$. This expression reveals that the corresponding objective function in (35) is the sum of the first-stage (deterministic) cost and of a convex combination of the mean future cost and of risk measures of future partial costs. With this choice of ρ, problem (35) becomes

$$\begin{aligned} &\inf \; f_1(x_1, \xi_1) + \theta_1 \mathbb{E}\left[\sum_{t=2}^{T} f_t(x_t, \xi_t) \right] + \sum_{t=2}^{T} \theta_t \rho^t \left(-\sum_{k=2}^{t} f_k(x_k, \xi_k) \right) \\ &x_t \in \chi_t(x_{t-1}, \xi_t), \; t = 1, \ldots, T. \end{aligned} \tag{39}$$

Plugging the expression (16) of the risk measure ρ^t (taking the same for all time steps) into (39), the latter can be written as

$$\begin{aligned}
&\inf_{x_t,w_t,y_t} f_1(x_1,\xi_1)+\sum_{t=2}^{T}\theta_t c_1^\top w_t+\mathbb{E}\left[\theta_1\sum_{t=2}^{T} f_t(x_t,\xi_t)+\sum_{t=2}^{T}\theta_t c_2^\top y_t\right]\\
&B_{2,1}w_t+B_{2,0}y_t=h\left(-\sum_{k=2}^{t} f_k(x_k,\xi_k)\right),\ t=2,\dots,T,\\
&A_1 w_t\le a_1,\ \ A_2 y_t\le a_2,\ t=2,\dots,T,\\
&x_t\in\chi_t(x_{t-1},\xi_t),\ t=1,\dots,T.
\end{aligned}$$

In turn, the above optimization problem can be expressed as

$$\begin{aligned}
&\inf_{x_1,w_{2:T}} f_1(x_1,\xi_1)+\sum_{t=2}^{T}\theta_t c_1^\top w_t+\mathcal{Q}_2(x_1,\xi_{[1]},w_2,\dots,w_T)\\
&A_1 w_t\le a_1,\ t=2,\dots,T,\ x_1\in\chi_1(x_0,\xi_1),
\end{aligned}\tag{40}$$

where

$$\mathcal{Q}_2(x_1,\xi_{[1]},w_{2:T})=\begin{cases}\inf\limits_{x_t,y_t}\mathbb{E}\left[\theta_1\sum_{t=2}^{T} f_t(x_t,\xi_t)+\sum_{t=2}^{T}\theta_t c_2^\top y_t\right]\\
B_{2,1}w_t+B_{2,0}y_t=h\left(-\sum_{k=2}^{t} f_k(x_k,\xi_k)\right),\ t=2,\dots,T,\\
A_2 y_t\le a_2,\ t=2,\dots,T,\\
x_t\in\chi_t(x_{t-1},\xi_t),\ t=2,\dots,T.\end{cases}\tag{41}$$

The application of Remark 3.2 to optimization problem (41) yields the following DP equations: for $t=2,\dots,T$, $\mathcal{Q}_t(x_{1:t-1},\xi_{[t-1]},w_{t:T})$ is given by

$$\mathbb{E}_{\xi_t|\xi_{[t-1]}}\begin{pmatrix}\inf\limits_{x_t,y_t}\theta_1 f_t(x_t,\xi_t)+\theta_t c_2^\top y_t+\mathcal{Q}_{t+1}(x_{1:t},\xi_{[t]},w_{t+1:T})\\
B_{2,1}w_t+B_{2,0}y_t=h(-\sum_{k=2}^{t} f_k(x_k,\xi_k)),\ A_2 y_t\le a_2,\ x_t\in\chi_t(x_{t-1},\xi_t)\end{pmatrix}.\tag{42}$$

In DP equations (37) and (42) obtained for, respectively, risk-averse problems (36) and (39), the state variables memorize the relevant history of the process and of the decisions. For (37) (resp., (42)), we can reduce the size of the state vector replacing the history of the decisions $x_{1:t-1}$ by x_{t-1} and z_{t-1} (resp., x_{t-1} and $\tilde{z}_{t-1}$ with $\tilde{z}_{t-1}=z_{t-1}-z_1$). Variable $\tilde{z}_{t-1}$ represents the total revenue (opposite of the cost) from time step 2 until time step $t-1$ (i.e., the total income until time step $t-1$ for the time steps where the data are random). Variables $\tilde{z}_t$ satisfy $\tilde{z}_t=\tilde{z}_{t-1}-f_t(x_t,\xi_t)$ for $t=2,\dots,T$, with $\tilde{z}_1$ set equal to 0. With this notation, DP equations (37) for problem (36) become

$$\begin{aligned}
&\mathcal{Q}_t(x_{t-1},\xi_{[t-1]},z_{t-1},y_{1:t-1})\\
&\quad=\mathbb{E}_{\xi_t|\xi_{[t-1]}}\begin{pmatrix}\inf\limits_{x_t,y_t,z_t} c_t^\top y_t+\mathcal{Q}_{t+1}(x_t,\xi_{[t]},z_t,y_{1:t})\\
(1-\delta_{t1})\left(\sum_{\tau=0}^{t-1} B_{t,\tau}y_{t-\tau}-h_t(z_t)\right)=0,\ A_t y_t\le a_t,\\
z_t=z_{t-1}-f_t(x_t,\xi_t),\ x_t\in\chi_t(x_{t-1},\xi_t)\end{pmatrix}
\end{aligned}\tag{43}$$

for $t = 1, \dots, T$, with $z_0 = 0$. As for the DP equations (40) and (42), they simplify as follows: in (40), $\mathcal{Q}_2(x_1, \xi_{[1]}, w_2, \dots, w_T)$ needs to be replaced by $\mathcal{Q}_2(x_1, \xi_{[1]}, \tilde{z}_1, w_2, \dots, w_T)$ and for $t = 2, \dots, T$ we have

$$
(44) \qquad \begin{aligned} &\mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_{t:T}) \\ &= \mathbb{E}_{\xi_t|\xi_{[t-1]}} \begin{pmatrix} \inf\limits_{x_t, \tilde{z}_t, y_t} \; - \delta_{tT}\theta_1 \tilde{z}_t + \theta_t c_2^\top y_t + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}, \tilde{z}_t, w_{t+1:T}) \\ B_{2,1} w_t + B_{2,0} y_t = h(\tilde{z}_t), \; A_2 y_t \le a_2, \\ \tilde{z}_t = \tilde{z}_{t-1} - f_t(x_t, \xi_t), \; x_t \in \chi_t(x_{t-1}, \xi_t) \end{pmatrix}. \end{aligned}
$$

Remark 3.4. Comparing the non-risk-averse DP equations (3) with the risk-averse ones (43) or (40) and (44), we see that additional decision and state variables are introduced in the latter cases. More precisely, at the first time step, in the non-risk-averse case the decision x_1 is taken, while in risk-averse case (43) (resp., (40) and (44)), additional decision variables y_1 and z_1 (resp., $(w_2, \dots, w_T)$) are needed. This first-stage problem is deterministic for all models.

For time step $t = 2, \dots, T$, in risk-averse case (43) (resp., (40) and (44)), the state vector is augmented with partial cost z_{t-1} and with the variables $(y_1, \dots, y_{t-1})$ (resp., partial cost $\tilde{z}_{t-1}$ and the variables $(w_t, \dots, w_T)$). For both risk-averse models, additional decisions z_t (or $\tilde{z}_t$) and y_t are needed for stages $t = 2, \dots, T$. This is summarized in Table 1.

Table 1

Decision and state variables for the non-risk-averse (NRA) DP equations (3) *as well as for the risk-averse ones* (43) *(RA$_1$), and* (40) *and* (44) *(RA$_2$).*

		First stage	*Stages $t = 2, \dots, T$*
Decision variables	NRA	x_1	x_t
	RA$_1$	(x_1, z_1, y_1)	(x_t, z_t, y_t)
	RA$_2$	$(x_1, w_2, \dots, w_T)$	$(x_t, \tilde{z}_t, y_t)$
State variables	NRA	$(x_0, \xi_{[0]})$	$(x_{t-1}, \xi_{[t-1]})$
	RA$_1$	$(x_0, \xi_{[0]})$	$(x_{t-1}, \xi_{[t-1]}, z_{t-1}, y_1, \dots, y_{t-1})$
	RA$_2$	$(x_0, \xi_{[0]})$	$(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_t, \dots, w_T)$

Remark 3.5. Other special cases for the multiperiod risk measure ρ in (35) for which DP equations can be written are the risk measures from Example 2.18.

Properties (P1) and (P3) thus hold for (36) and hold for (39) when using extended one-period polyhedral risk measures for ρ^t. We now concentrate on (P2). So far, all the developments of this section were valid for a problem of the form (1). To ensure that (P2) holds, we consider the special case when (1) is a stochastic linear program (SLP). Indeed, the convergence of the SDDP algorithm and of related sampling-based algorithms is proved in Philpott and Guan [PG08] for SLP. We observe that if (1) is an SLP, then risk-averse problem (36) (resp., (39)) is an SLP if and only if

$$
(45) \qquad \begin{aligned} &h_t(z) = z b_t + \tilde{b}_t \text{ for some } b_t, \tilde{b}_t \in \mathbb{R}^{n_{t,2}} \\ &(\text{resp., } h(z) = z b_2 + \tilde{b}_2 \text{ for some } b_2, \tilde{b}_2 \in \mathbb{R}^{n_{2,2}}). \end{aligned}
$$

Of interest for applications, we now specialize the above DP equations (44) taking extended polyhedral risk measures with $h(\cdot)$ of the kind (45) above. As seen in the previous section, spectral risk measures with piecewise constant spectra are of this kind. We provide the DP equations obtained in this case using directly (27).

3.3. Spectral risk measures. Let ϕ be a piecewise constant risk spectrum satisfying (i), (ii), and (iii) given in Example 2.16 and let $\Delta\phi_k = \phi(p_k) - \phi(p_{k-1}), k = 1, \ldots, J$. If we take for ρ^t a spectral risk measure ρ_ϕ (the same for all time steps), using (27) we can decompose (39) by stages and express it under the form

$$\begin{array}{ll} \inf & f_1(x_1, \xi_1) + \sum_{t=2}^{T} \theta_t c_1^\top w_t + \mathcal{Q}_2(x_1, \xi_{[1]}, \tilde{z}_1, w_2, \ldots, w_T) \\ & x_1 \in \chi_1(x_0, \xi_1),\ w_t \in \mathbb{R}^J,\ t = 2, \ldots, T, \end{array} \tag{46}$$

with $\tilde{z}_1 = 0, c_1 = \Delta\phi \circ p$, and where for $t = 2, \ldots, T$,

$$\begin{aligned} &\mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_{t:T}) \\ &\quad = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf_{x_t, \tilde{z}_t} \tilde{f}_t(\tilde{z}_t, w_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}, \tilde{z}_t, w_{t+1:T}) \\ \tilde{z}_t = \tilde{z}_{t-1} - f_t(x_t, \xi_t),\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right) \end{aligned} \tag{47}$$

with

$$\tilde{f}_t(\tilde{z}_t, w_t) = -(\delta_{tT}\theta_1 + \phi(1)\theta_t)\tilde{z}_t - \theta_t\, \Delta\phi^\top (w_t - \tilde{z}_t e)^+.$$

When the risk spectrum ϕ has one jump, we obtain the CVaR.

3.4. Conditional value-at-risk. When taking $\rho^t = CVaR^{\varepsilon_t}$ and using (28), we can express (39) under the form

$$\begin{array}{ll} \inf_{x_1, w_{2:T}} & f_1(x_1, \xi_1) - \sum_{t=2}^{T} \theta_t w_t + \mathcal{Q}_2(x_1, \xi_{[1]}, \tilde{z}_1, w_2, \ldots, w_T) \\ & x_1 \in \chi_1(x_0, \xi_1),\ w_t \in \mathbb{R},\ t = 2, \ldots, T, \end{array} \tag{48}$$

with $\tilde{z}_1 = 0$, and where for $t = 2, \ldots, T$,

$$\begin{aligned} &\mathcal{Q}_t(x_{t-1}, \xi_{[t-1]}, \tilde{z}_{t-1}, w_{t:T}) \\ &\quad = \mathbb{E}_{\xi_t|\xi_{[t-1]}} \left(\begin{array}{l} \inf_{x_t, \tilde{z}_t} - \delta_{tT}\theta_1\tilde{z}_t + \frac{\theta_t}{\varepsilon_t}(w_t - \tilde{z}_t)^+ + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}, \tilde{z}_t, w_{t+1:T}) \\ \tilde{z}_t = \tilde{z}_{t-1} - f_t(x_t, \xi_t),\ x_t \in \chi_t(x_{t-1}, \xi_t) \end{array} \right). \end{aligned} \tag{49}$$

3.5. Convergence of SDDP in a risk-averse setting. The convergence of the SDDP algorithm and of related sampling-based algorithms is proved in Philpott and Guan [PG08] for SLP with the following properties:

- (A1) Random data only appear on the right-hand side of the constraints.
- (A2) The supports of the distributions of the underlying random vectors are discrete and finite.
- (A3) Random vectors are interstage independent or satisfy a certain type of interstage dependence (see Philpott and Guan [PG08]).
- (A4) The feasible set of the linear program is nonempty and bounded in each stage.

In what follows, we consider multistage SLPs of the form (1) where

$$f_t(x_t, \xi_t) = d_t^\top x_t \quad \text{and} \quad \chi_t(x_{t-1}, \xi_t) = \{x_t : x_t \geq 0,\ C_t x_t = \xi_t - D_t x_{t-1}\}. \tag{50}$$

For these programs, assumption (A1) holds, and it can be noted that if (A1) holds for (1), then (A1) holds for risk-averse problems (36) and (39). In the remainder of the paper, we assume (A2) and (A3). We also assume that (A4) holds for (1), which, in our context, can be expressed as follows:

(A4) For $t = 1, \ldots, T$, for any feasible state x_{t-1}, and for any realization ξ_t^i of ξ_t, the set

$$\chi_t(x_{t-1}, \xi_t^i) = \{x_t \mid x_t \geq 0,\ C_t x_t = \xi_t^i - D_t x_{t-1}\}$$

is bounded and nonempty.

To apply the convergence results from Philpott and Guan [PG08] in our risk-averse setting, (A4) should also hold for risk-averse problems (36) or (39). For (36), (A4) takes the following form:

(A5) $\{y_1 : A_1 y_1 \leq a_1\}$ is bounded and for all $t = 2, \ldots, T$, for any feasible states $x_1, y_1, \ldots, x_{t-1}, y_{t-1}$, and for any sequence of realizations $\xi_1^i, \ldots, \xi_t^i$ of $\xi_1, \ldots, \xi_t$, the set $\{y_t : A_t y_t \leq a_t, B_{t,0} y_t = h_t(-\sum_{\tau=1}^{t} f_\tau(x_\tau, \xi_\tau^i)) - \sum_{\tau=1}^{t-1} B_{t,\tau} y_{t-\tau}$ for some $x_t \in \chi_t(x_{t-1}, \xi_t^i)\}$ is bounded and nonempty.

For (39), remembering Proposition 2.15, a condition implying (A4) is the following:

(A6) For $t = 2, \ldots, T$, the sets $S(\rho^t(0))$ are nonempty and bounded, where $S(\rho^t(0))$ is defined in (26). $\{y_1 : A_1 y_1 \leq a_1\}$ is bounded and for all $t = 2, \ldots, T$, for any feasible $x_1, y_1, \ldots, x_{t-1}, y_{t-1}, w_{2:T}$, and for any sequence of realizations $\xi_1^i, \ldots, \xi_t^i$ of $\xi_1, \ldots, \xi_t$, the set $\{y_t : A_t y_t \leq a_t, \exists\ x_t \in \chi_t(x_{t-1}, \xi_t^i), B_{2,0} y_t = h(-\sum_{\tau=2}^{t} f_\tau(x_\tau, \xi_\tau^i)) - B_{2,1} w_t\}$ is bounded and nonempty.

Indeed, with respect to the non-risk-averse setting, recall that the additional decision variables for (39) are $\tilde{z}_t$ (bounded, due to (A4)), y_t, and w_t. Variables $w_t, t = 2, \ldots, T$, are first-stage decision variables and, due to Proposition 2.15, if $S(\rho^t(0))$ is nonempty and bounded, then optimal w_t are bounded. Next, condition (A6) guarantees the boundedness of optimal y_t.

However, even if the feasible set at each stage for (36) or (39) is not bounded, we may be able to show, in some cases, that these feasible sets can be replaced by bounded feasible sets without changing the problems, i.e., that the solutions are bounded. Such is the case for problems (46) and (48). Indeed, for these problems, the only additional variables with respect to the non-risk-averse case are $\tilde{z}_t$ (bounded, due to (A4)) and first-stage variables $w_2, \ldots, w_T$. For the spectral risk measure $\rho^t = \rho_\phi, t = 2, \ldots, T$, considered in (46), the sets $S(\rho^t(0)) = S(\rho_\phi(0)) = \{0\}, t = 2, \ldots, T$, are nonempty and bounded. Using Proposition 2.15, optimal values of w_t in (46) are bounded. This result can also be easily proved directly.

LEMMA 3.6. *Let assumption* (A4) *hold, and let* ϕ *be a piecewise risk spectrum satisfying* (i), (ii), *and* (iii) *given in Example* 2.16. *Let* $w_2^*, \ldots, w_T^*$ *be optimal values of* $w_2, \ldots, w_T$ *for* (46). *Then* $w_t^*(k)$ *is finite for every* $t = 2, \ldots, T$, *and* $k = 1, \ldots, J$.

Proof. Since $\chi_t, t = 1, \ldots, T$, are bounded and $\Delta\phi < 0$, we can bound from below the objective function of (46) by $L_1(w) = K_1 + \sum_{t=2}^{T} \theta_t (\Delta\phi \circ p)^\top w_t$ and $L_2(w) = K_2 + \sum_{t=2}^{T} \theta_t (\Delta\phi \circ (p - e))^\top w_t$ for some constants K_1 and K_2. Since $\Delta\phi \circ p < 0$, if one component $w_t(k) = -\infty$, then $L_1(w) = +\infty$, the objective function is therefore $+\infty$, and such $w_t(k)$ cannot be an optimal value of $w_t(k)$. Similarly, since $\Delta\phi \circ (p - e) > 0$, if one $w_t(k) = +\infty$, then $L_2(w) = +\infty$, the objective function is $+\infty$, and such $w_t(k)$ cannot be an optimal value of $w_t(k)$. ∎

The following corollary is an immediate consequence of this lemma.

COROLLARY 3.7. *Let assumption* (A4) *hold. Let* $w_2^*, \ldots, w_T^*$ *be optimal values of* $w_2, \ldots, w_T$ *for* (48). *Then* w_t^* *is finite for every* $t = 2, \ldots, T$.

It follows that we can add (sufficiently large) box constraints on w_t in (46) and (48) without changing the optimal solutions of (46) and (48). Gathering our observations, we come to the following proposition.

Proposition 3.8 (convergence of SDDP in a risk-averse setting). *Consider multistage SLPs of the form* (1) *with f_t and χ_t given by* (50). *Assume that for such multistage programs, assumptions* (A1), (A2), (A3), *and* (A4) *hold. Consider the risk-averse formulations* (46), (47) *and* (48), (49). *Then an SDDP algorithm applied on these DP equations will converge if the sampling procedures satisfy the FPSP and BPSP assumptions (see Philpott and Guan* [PG08]*).*

The same convergence result holds for the following two risk-averse formulations:

(1) *assuming* (A5), *for risk-averse program* (36) *decomposed by stages as* (43) *with $h_t(\cdot)$ given by* (45)*;*

(2) *assuming* (A6), *for risk-averse program* (39) *decomposed by stages as* (40), (44) *with $h(\cdot)$ given by* (45).

In the next section, we detail the SDDP algorithm for interstage independent risk-averse problems of form (35). The developments can be easily adapted to the case when the process affinely depends on previous values. Our notation follows closely that of Birge and Donohue [BD06].

4. Decomposition algorithms for a class of risk-averse stochastic programs. We consider the risk-averse recourse functions (43) from section 3 in the case when f_t and χ_t are given by (50) and $h_t(\cdot)$ is given by (45). Recall that risk-averse DP equations (43) satisfy (P3) (like the non-risk-averse DP equations (3) but with additional state and control variables). We assume interstage independence and relatively complete recourse for (1). We also assume that the hypotheses of Proposition 3.8 hold. In this context, relatively complete recourse also holds for risk-averse problems (43). As a result, the SDDP algorithm [PP91], [Sha11] can be applied to obtain approximations of the corresponding risk-averse recourse functions. At each iteration, this algorithm consists of a forward pass followed by a backward pass. The backward pass builds cuts for the recourse functions (hyperplanes lying below these functions) at some points computed in the forward pass. If H cuts are built for each recourse function at each iteration, iteration i ends with a lower bounding approximation of form

(51)
$$\mathcal{Q}_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}) = \max_{j=0,1,\ldots,iH} \left[-E_{t-1}^j x_{t-1} - Z_{t-1}^j z_{t-1} - \sum_{\tau=1}^{t-1} Y_{t-1}^{j,\tau} y_\tau + e_{t-1}^j \right]$$

for $\mathcal{Q}_t$, knowing that the algorithm starts taking for $\mathcal{Q}_t^0$ a known lower bounding affine approximation of $\mathcal{Q}_t$ while $\mathcal{Q}_{T+1}^i \equiv 0$. In the above expression, $Z_{t-1}^j \in \mathbb{R}$, while E_{t-1}^j and $Y_{t-1}^{j,\tau}$ are row vectors of appropriate dimensions.

The forward pass of iteration i samples H scenarios $(\xi_2^k, \ldots, \xi_T^k)$, $k = (i-1)H + 1, \ldots, iH$, from the distribution of $(\xi_2, \ldots, \xi_T)$. On scenario $(\xi_2^k, \ldots, \xi_T^k)$, the decisions $(x_1^k, \ldots, x_T^k, y_1^k, \ldots, y_T^k)$ as well as the partial costs $(z_1^k, \ldots, z_T^k)$ are computed replacing recourse functions $\mathcal{Q}_t$ by $\mathcal{Q}_t^{i-1}$ for $t = 2, \ldots, T+1$. The stopping criterion is discussed in [Sha11].

The cuts are computed from time step $T+1$ down to time step 2. For time step $T+1$, since $\mathcal{Q}_{T+1}^i = \mathcal{Q}_{T+1} = 0$, cuts for $\mathcal{Q}_{T+1}$ are obtained taking null values for $E_T^k, Z_T^k, Y_T^{k,\tau}$, and e_T^k for $k = (i-1)H + 1, \ldots, iH$. At $t = 2, \ldots, T$, cuts for $\mathcal{Q}_t$ are computed at $(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k)$, $k = (i-1)H + 1, \ldots, iH$. More precisely, having at hand the lower bounding approximation $\mathcal{Q}_{t+1}^i$ of $\mathcal{Q}_{t+1}$, we can bound from below $\mathcal{Q}_t(x_{t-1}, z_{t-1}, y_{1:t-1})$ by $\mathbb{E}_{\xi_t}[Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t)]$ with $Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t)$

given as the optimal value of the following linear program:

$$
(52) \quad \begin{array}{ll}
\inf_{x_t, y_t, z_t, \tilde{\theta}_t} c_t^\top y_t + \tilde{\theta}_t & \\
A_t y_t \leq a_t, \; x_t \geq 0, & \\
\sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} - z_t b_t = \tilde{b}_t, & \text{(a)} \\
z_t + d_t^\top x_t = z_{t-1}, & \text{(b)} \\
C_t x_t = \xi_t - D_t x_{t-1}, & \text{(c)} \\
\overrightarrow{E}_t^i x_t + \overrightarrow{Z}_t^i z_t + e\tilde{\theta}_t \geq -\sum_{\tau=1}^{t} \overrightarrow{Y}_t^{i,\tau} y_\tau + \overrightarrow{e}_t^i, & \text{(d)}
\end{array}
$$

where $\overrightarrow{Z}_t^i = (Z_t^0, Z_t^1, \ldots, Z_t^{iH})^\top$ and $\overrightarrow{Y}_t^{i,\tau}$ is the matrix whose $(j+1)$th line is $Y_t^{j,\tau}$ for $j = 0, \ldots, iH$. We denote by ξ_t^j, $j = 1, \ldots, q_t < +\infty$, the possible realizations of ξ_t with $p(t,j) = \mathbb{P}(\xi_t = \xi_t^j)$. We also denote by $\sigma_t^{k,j}, \mu_t^{k,j}, \pi_t^{k,j}$, and $\rho_t^{k,j}$ the (row vectors) optimal Lagrange multipliers associated to constraints (52)-(a), (52)-(b), (52)-(c), and (52)-(d) for the problem defining $Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j)$. With this notation, the following theorem provides the cuts computed for $\mathcal{Q}_t$ at iteration i.

THEOREM 4.1. *Let $\mathcal{Q}_t, t = 2, \ldots, T+1$, be the risk-averse recourse functions given by* (43) *with $h_t(\cdot)$ given by* (45)*. In the backward pass of iteration i of the SDDP algorithm, the following cuts are computed for these recourse functions. For $t = T+1$, we set $E_{t-1}^k, Z_{t-1}^k, Y_{t-1}^{k,\tau}$ and e_{t-1}^k to 0 for $k = (i-1)H+1, \ldots, iH$ and $\tau = 1, \ldots, T$. For $t = 2, \ldots, T$ and $k = (i-1)H+1, \ldots, iH$, $E_{t-1}^k = \sum_{j=1}^{q_t} p(t,j)\pi_t^{k,j} D_t$ and*

$$
Z_{t-1}^k = -\sum_{j=1}^{q_t} p(t,j)\mu_t^{k,j}, Y_{t-1}^{k,\tau} = \sum_{j=1}^{q_t} p(t,j)(\sigma_t^{k,j} B_{t,t-\tau} + \rho_t^{k,j} \overrightarrow{Y}_t^{i,\tau}), \; \tau = 1, \ldots, t-1.
$$

Next, e_{t-1}^k is given by

$$
\sum_{j=1}^{q_t} p(t,j) \Bigg[Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j) - \mu_t^{k,j} z_{t-1}^k + \sum_{\tau=1}^{t-1} (\sigma_t^{k,j} B_{t,t-\tau} + \rho_t^{k,j} \overrightarrow{Y}_t^{i,\tau}) y_\tau^k + \pi_t^{k,j} D_t x_{t-1}^k \Bigg].
$$

Proof. Since relatively complete recourse and assumptions (A4) and (A5) hold, the linear program defining $Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j)$ has a nonempty feasible set and its optimal value is finite. As a result, both this primal problem and its dual have the same optimal value. Since a dual solution is a subgradient of the value function for problem (52), we obtain for $Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t^j)$ the lower bound

$$
\begin{aligned}
Q_t^i(x_{t-1}^k, z_{t-1}^k, y_{1:t-1}^k, \xi_t^j) &- \sum_{\tau=1}^{t-1} \sigma_t^{k,j} B_{t,\tau}(y_{t-\tau} - y_{t-\tau}^k) - \sum_{\tau=1}^{t-1} \rho_t^{k,j} \overrightarrow{Y}_t^{i,\tau}(y_\tau - y_\tau^k) \\
&+ \mu_t^{k,j}(z_{t-1} - z_{t-1}^k) - \pi_t^{k,j} D_t (x_{t-1} - x_{t-1}^k).
\end{aligned}
$$

Plugging this bound into the relation $\mathcal{Q}_t(x_{t-1}, z_{t-1}, y_{1:t-1}) \geq \sum_{j=1}^{q_t} p(t,j) Q_t^i(x_{t-1}, z_{t-1}, y_{1:t-1}, \xi_t^j)$, rearranging the terms, and identifying with (51) gives the announced cuts. ∎

The above cuts can be easily specialized to DP equations (46)–(47) (based on spectral risk measures) or to (44) with $h(\cdot)$ as in (45).

5. Conclusion. The class of extended polyhedral risk measures was introduced in this paper. Dual representations of these risk measures were obtained and used to provide conditions for coherence, convexity, and consistency with second order stochastic dominance.

This class allowed us to write risk-averse dynamic programming equations for some risk-averse problems with risk measures taken from this class. We then detailed a stochastic dual dynamic programming algorithm for approximating the corresponding risk-averse recourse functions for some stochastic linear programs. In particular, conditions were given to guarantee convergence. The methodology can be easily adapted if the recourse functions are approximated using other sampling-based decomposition algorithms such as AND (Birge and Donohue [BD06]) and DOASA (Philpott and Guan [PG08]).

A forthcoming work will assess the proposed approach on a midterm multistage production management problem Guigues [Gui].

REFERENCES

[Ace02] C. Acerbi, *Spectral measures of risk: A coherent representation of subjective risk aversion*, J. Banking and Finance, 7 (2002), pp. 1505–1518.

[ADE+] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku, *Coherent Multiperiod Risk Measurement*, available online from http://www.math.ethz.ch/∼delbaen, 2003.

[ADE+07] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku, *Coherent multiperiod risk adjusted values and Bellman's principle*, Ann. Oper. Res., 152 (2007), pp. 5–22.

[AS] C. Acerbi and P. Simonetti, *Portfolio Optimization with Spectral Measures of Risk*, Abaxbank internal report, 2002; available online from http://www.gloriamundi.org/.

[BD06] J. R. Birge and C. J. Donohue, *The abridged nested decomposition method for multistage stochastic linear programs with relatively complete recourse*, Algorithmic Oper. Res., 1 (2006), pp. 20–30.

[BL88] J. R. Birge and F. V. Louveaux, *A multicut algorithm for two-stage stochastic linear programs*, European J. Oper. Res., 34 (1988), pp. 384–392.

[BS00] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[BTT07] A. Ben-Tal and M. Teboulle, *An old-new concept of convex risk measures: The optimized certainty equivalent*, Math. Finance, 17 (2007), pp. 449–476.

[CD06] J. Cotter and K. Dowd, *Extreme spectral risk measures: An application to futures clearinghouse variation margin requirements*, J. Banking and Finance, 30 (2006), pp. 3469–3485.

[CPR] R. A. Collado, D. Papp, and A. Ruszczyński, *Scenario Decomposition of Risk-Averse Multistage Stochastic Programming Problems*, 2010; available online from http://www.optimization-online.org/DB_HTML/2010/08/2717.html.

[DG90] G. B. Dantzig and P. W. Glynn, *Parallel processors for planning under uncertainty*, Ann. Oper. Res., 22 (1990), pp. 1–21.

[Eic07] A. Eichhorn, *Stochastic Programming Recourse Models: Approximation, Risk Aversion, Applications in Energy*, Logos Verlag, Berlin, 2007.

[ER05] A. Eichhorn and W. Römisch, *Polyhedral risk measures in stochastic programming*, SIAM J. Optim., 16 (2005), pp. 69–95.

[FS04] H. Föllmer and A. Schied, *Stochastic Finance. An Introduction in Discrete Time*, 2nd ed., Walter de Gruyter, Berlin, 2004.

[FS05] M. Fritelli and G. Scandalo, *Risk measures and capital requirements for processes*, Math. Finance, 16 (2005), pp. 589–612.

[Gui] V. Guigues, *SDDP for Some Interstage Dependent Risk Averse Problems and Application to Hydro-thermal Planning*, preprint, 2011; available online from http://www.optimization-online.org/DB_HTML/2011/03/2970.html.

[HRS10] J. L. Higle, B. Rayco, and S. Sen, *Stochastic scenario decomposition for multistage stochastic programs*, IMA J. Manag. Math., 21 (2010), pp. 39–66.

[HS96] J. L. Higle and S. Sen, *Stochastic Decomposition*, Kluwer Academic, Dordrecht, The Netherlands, 1996.

[Inf92] G. Infanger, *Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs*, Ann. Oper. Res., 39 (1992), pp. 69–95.

[KP09] R. Kovacevic and G. Ch. Pflug, *Time consistency and information monotonicity of multiperiod acceptability functionals*, in Advanced Financial Modeling, H. Albrecher, W. J. Rungaldier, and W. Schachermayer, eds., Radon Ser. Comput. Appl. Math. 8, Walter de Gruyter, Berlin, 2009, pp. 347–369.

[MS02] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester, UK, 2002.

[OR02] W. Ogryczak and A. Ruszczyński, *Dual stochastic dominance and related mean-risk models*, SIAM J. Optim., 13 (2002), pp. 60–78.

[PG08] A. B. Philpott and Z. Guan, *On the convergence of stochastic dual dynamic programming and related methods*, Oper. Res. Lett., 36 (2008), pp. 450–455.

[PP91] M. V. F. Pereira and L. M. V. G. Pinto, *Multi-stage stochastic optimization applied to energy planning*, Math. Programming, 52 (1991), pp. 359–375.

[PR07] G. Ch. Pflug and W. Römisch, *Modeling, Measuring, and Managing Risk*, World Scientific, Singapore, 2007.

[RS03] A. Ruszczyński and A. Shapiro, *Stochastic Programming*, Handbooks Oper. Res. Management Sci. 10, Elsevier, Amsterdam, 2003.

[RS06a] A. Ruszczyński and A. Shapiro, *Conditional risk mappings*, Math. Oper. Res., 31 (2006), pp. 544–561.

[RS06b] A. Ruszczyński and A. Shapiro, *Optimization of convex risk functions*, Math. Oper. Res., 31 (2006), pp. 433–452.

[RU02] R. T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Banking and Finance, 26 (2002), pp. 1443–1471.

[Rus86] A. Ruszczyński, *A multicut regularized decomposition method for minimizing a sum of polyhedral functions*, Math. Programming, 35 (1986), pp. 309–333.

[RW98] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[SDR09] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[Sha09] A. Shapiro, *On a time consistency concept in risk averse multistage programming*, Oper. Res. Lett., 37 (2009), pp. 143–147.

[Sha11] A. Shapiro, *Analysis of stochastic dual dynamic programming method*, European J. Oper. Res., 209 (2011), pp. 63–72.