POLYHEDRAL RISK MEASURES IN STOCHASTIC PROGRAMMING*

A. EICHHORN† AND W. RÖMISCH†

ABSTRACT. Stochastic programs that do not only minimize expected cost but also take into account risk are of great interest in many application fields. We consider stochastic programs with risk measures in the objective and study stability properties as well as decomposition structures. Thereby we place emphasis on dynamic models, i.e., multistage stochastic programs with multiperiod risk measures. In this context, we define the class of polyhedral risk measures such that stochastic programs with risk measures taken from this class have favorable properties. Polyhedral risk measures are defined as optimal values of certain linear stochastic programs where the arguments of the risk measure appear on the right-hand side of the dynamic constraints. Dual representations for polyhedral risk measures are derived and used to deduce criteria for convexity and coherence. As examples of polyhedral risk measures we propose multiperiod extensions of the Conditional-Value-at-Risk.

1. INTRODUCTION

Stochastic programs are essentially known to minimize, maximize, or to bound expected values. From a theoretical point of view they easily offer the possibility to minimize or to bound risk functionals since they rest upon stochastic models. However, in practice it may happen that incorporating risk measures in stochastic programs makes them much harder to solve. In addition, other favorable properties like stability with respect to approximations or duality results may get lost. In this paper considerations are made about the question how risk measures should be designed so that stochastic programs incorporating them show similar properties as stochastic programs based on expected values only. As a result, the class of polyhedral risk measure is introduced.

Of course, when analyzing risk measures with respect to their practicability for stochastic programs, one has to determine first of all what is understood by the expression *risk measure* and what properties are required from the viewpoint of economic considerations. Here, a (one-period) risk measure ρ will be understood as a functional from some set of real random variables $\mathcal{Z}$ to the real numbers, the random variables $z \in \mathcal{Z}$ represent some uncertain (usually monetary) value for which large outcomes are preferred to lower ones. The value $\rho(z)$ gives information about the riskiness of z, i.e., a high value $\rho(z)$ indicates a high danger of reaching low values, whereas a low value represents low danger of low values.

Risk measures are broadly discussed in financial mathematics. For one-period risk measures, i.e., for risk measures that depend on one random variable only, there is a relatively high degree of agreement among the community about the desirable properties. Possibly the most important work in this context is the axiomatic characterization of coherent risk measures [1], where the risk $\rho(z)$ is understood as the minimal amount of additional (risk-free) capital that is required to make the position z acceptable. Several generalizations of this paper followed, e.g. [5, 12, 9, 25], see also Chapter 4 in the monograph [10]. Further desirable properties, namely, the consistency of risk measures with stochastic dominance rules, were suggested in [13, 15, 16]. In addition, there are papers dealing with specific risk measures such as Value-at-Risk or Conditional-Value-at-Risk, e.g., [6, 24, 17], see also the volume [37]. Recently, an optimization theory of convex risk measures has been developed in [33].

Key words and phrases. Stochastic programming, risk measure, convexity, coherence, polyhedrality, quantitative stability, probability metrics, dual decomposition

2000 MSC 90C15

† Humboldt-University Berlin, Institute of Mathematics, Unter den Linden 6, 10099 Berlin, Germany

`<eichhorn / romisch@math.hu-berlin.de>`

* This research was supported by the DFG Research Center *Mathematics for Key Technologies* at Berlin and by a grant of EDF Electricité de France.

Currently, generalizations of one-period risk measures to different dynamic settings are discussed in the literature. Such generalizations become necessary when a sequence of random variables $z_1, ..., z_T$ is to be assessed with respect to its riskiness and when information is revealed gradually with the passing of time. In the literature, the settings as well as the postulated properties for risk functionals differ more than in the one-period case. Generally speaking, there are two classes of settings depending on whether liquidity risk over a time period is considered or intermediate monitoring by supervisors is to be anticipated. In some work an entire risk measure process $\rho_1, ..., \rho_T$ is defined, see [22, 39] and also [2, 3]. The more important case from the viewpoint of optimization is the case where one has one real number $\rho(z_1, ..., z_T)$ that represents the risk of the entire process. Such concepts are presented in [19, 34, 18] and again in [2, 3]. As in the one-period case, the number $\rho(z_1, ..., z_T)$ can be understood as minimal capital requirement for the overall time period so that the strategy corresponding to $z_1, ..., z_T$ is acceptable.

In the present paper, we consider (mixed-integer) multistage stochastic programs of the form

$$\min \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t\rangle\right] \;\middle|\; \begin{array}{ll} x_t \in X_t, & \\ x_t \text{ is } \mathcal{F}_t\text{-measurable}, & (t = 1, ..., T) \\ \sum_{\tau=0}^{t-1} A_{t,\tau}(\xi_t) x_{t-\tau} = h_t(\xi_t) & \end{array} \right\} \tag{1.1}$$

as starting point, where $(\xi_t)_{t=1}^T$ is a stochastic process and $\mathcal{F}_t = \sigma(\xi_1, ..., \xi_t)$, the sets X_t are closed and have polyhedral convex hulls, $b_t(\cdot)$ are cost coefficients, $h_t(\cdot)$ right-hand sides and $A_{t,\tau}(\cdot)$, $\tau = 0, \ldots, t-1$, matrices having appropriate dimensions and possibly depending on ξ_t for $t = 1, \ldots, T$.

Much is known for expectation-based stochastic programs, e.g., on optimality and duality, decomposition methods, statistical approximations and stability (cf. [32]). Most of these results are essentially based on the fact that $\mathbb{E}$ is a linear operator. As it will be seen below in Chapter 2, risk measures ρ are usually by no means linear. Hence, if we change from expectation to a risk measure in (1.1), many known results will be no longer valid. Nevertheless, there are results about incorporating certain risk functionals into (stochastic) optimization problems, e.g. [24, 17, 35, 36, 33]. In particular, the Conditional-Value-at-Risk turns out to behave very opportunely in stochastic programs. However, from an economic point of view not every risk measure is suitable for any application. In particular, for multistage stochastic programs it may become necessary to incorporate risk measures for processes, i.e., to minimize terms like $\rho(z_1, \ldots, z_T)$ with $z_t = -\sum_{\tau=1}^{t} \langle b_\tau(\xi_\tau), x_\tau\rangle$. Hence, it would be convenient to have an entire class of risk measures at hand such that every risk measure from this class behaves opportunely in stochastic programs.

Such a class will be introduced in Chapter 2, namely the class of polyhedral risk measures. Conditions implying that polyhedral risk measures are coherent and consistent with second order stochastic dominance are provided. In Chapter 3 this class will be extended to the multiperiod case. Briefly, polyhedral risk measures are defined as optimal values of certain simple linear stochastic programs. In Chapter 4 it will be shown that, indeed, several properties of expectation-based stochastic programs remain valid for stochastic programs with polyhedral risk measures as objectives. This is due to the fact that a problem of the form (1.1) with $\mathbb{E}$ replaced by a polyhedral risk measure ρ can easily be transformed into a stochastic program with an objective consisting of the expectation of a linear function where the original objective occurs in some additional dynamic constraints.

2. Polyhedral Risk Measures

Let $\mathcal{Z}$ denote a linear space of real random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that $\mathcal{Z}$ contains the constants, e.g. $\mathcal{Z} = L_p(\Omega, \mathcal{F}, \mathbb{P})$ with some $p \in [1, \infty]$. According to [10] a functional $\rho : \mathcal{Z} \to \bar{\mathbb{R}}$ is called a *risk measure* if it satisfies the following two conditions for all z, $\tilde{z} \in \mathcal{Z}$:
(i) If $z \leq \tilde{z}$, then $\rho(z) \geq \rho(\tilde{z})$ *(monotonicity)*.
(ii) For each $r \in \mathbb{R}$ we have $\rho(z + r) = \rho(z) - r$ *(translation invariance)*.
A risk measure ρ is called *convex* if it satisfies the condition

$$\rho(\mu z + (1 - \mu)\tilde{z}) \leq \mu\rho(z) + (1 - \mu)\rho(\tilde{z})$$

for all z, $\tilde{z} \in \mathcal{Z}$ and $\mu \in [0, 1]$. A convex risk measure is called *coherent* if it is *positively homogeneous*, i.e., $\rho(\mu z) = \mu\rho(z)$ for all $\mu \geq 0$ and $z \in \mathcal{Z}$.

There is a number of representation theorems for convex and especially for coherent risk measures in the literature emerging from convex duality. Next, we cite one of these representations adapted to our needs. Therefore, we set

$$\mathcal{D} := \{f \in L_1(\Omega, \mathcal{F}, \mathbb{P}) : f \geq 0, \mathbb{E}[f] = 1\},$$

the set of all density functions for $(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 2.1. *Let $\rho : L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \bar{\mathbb{R}}$ with $p \in [1, \infty]$. Assume that ρ satisfies the following continuity property:*

$$z_n \uparrow z_0 \text{ a.s.} \Rightarrow \lim_{n\to\infty} \rho(z_n) = \rho(z_0)$$

Then the following equivalence statement holds:

$$\rho \text{ is a coherent risk measure} \iff \exists \mathcal{P}_\rho \subseteq \mathcal{D} \text{ convex} : \rho(z) = \sup_{f \in \mathcal{P}_\rho} \mathbb{E}[-fz] \ \forall z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$$

Proof. Follows from [10], Corollary 4.14 + Proposition 4.17 + Lemma 4.25. See also [5, 25, 33]. □

Now we are ready to define the class of polyhedral risk measures.

Definition 2.2. A risk measure ρ on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with some $p \in [1, \infty]$ will be called *polyhedral* if there exist $k_1, k_2 \in \mathbb{N}$, $c_1, w_1 \in \mathbb{R}^{k_1}$, $c_2, w_2 \in \mathbb{R}^{k_2}$, a nonempty polyhedral set $Y_1 \subseteq \mathbb{R}^{k_1}$, and a polyhedral cone $Y_2 \subseteq \mathbb{R}^{k_2}$ such that

$$\rho(z) = \inf \left\{ \langle c_1, y_1 \rangle + \mathbb{E}[\langle c_2, y_2 \rangle] \;\middle|\; \begin{array}{l} y_1 \in Y_1, \\ y_2 \in L_p(\Omega, \mathcal{F}, \mathbb{P}),\, y_2 \in Y_2, \\ \langle w_1, y_1 \rangle + \langle w_2, y_2 \rangle = z \end{array} \right\} \tag{2.1}$$

for every $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$. Here, $\mathbb{E}$ denotes the expectation on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\langle \cdot, \cdot \rangle$ a scalar product on $\mathbb{R}^{k_1}$ or $\mathbb{R}^{k_2}$.

Hence, expressed in the language of stochastic programming, a polyhedral risk measure is given as the optimal value of a certain two-stage stochastic program with random right-hand side. We use the term *polyhedral* because, for $\#\Omega < \infty$, the space $L_p(\Omega, \mathcal{F}, \mathbb{P})$ can be identified with $\mathbb{R}^{\#\Omega}$ and in this case a risk measure defined by (2.1) is indeed a polyhedral function on $\mathbb{R}^{\#\Omega}$.

Remark 2.3. Of course, a convex combination of (negative) expectation and a polyhedral risk measure is again a polyhedral risk measure: Let $\mu \in [0, 1]$ and ρ be a polyhedral risk measure with dimensions k_t, vectors c_t and w_t $(t = 1, 2)$, and polyhedral set/cone Y_1 / Y_2. Then the risk measure $\hat{\rho} := \mu\rho - (1-\mu)\mathbb{E}$ is polyhedral with the same dimensions k_t and the same sets Y_t and vectors $\hat{w}_1 := w_1$, $\hat{w}_2 := w_2$, $\hat{c}_1 := \mu c_1 - (1-\mu)w_1$ and $\hat{c}_2 := \mu c_2 - (1-\mu)w_2$. Thus, so-called *mean-risk-models*, where expectation and risk are optimized simultaneously, do not need to be considered separately.

Next, we derive dual representations for (2.1). To this end, we do not need to assume that ρ is a risk measure in the sense of [9, 10], i.e., that it is monotone and translation invariant. We conclude in our first result that ρ is a convex functional. To state this result, we use the notation[1]

$$D_\rho := \{u \in \mathbb{R} : uw_1 - c_1 \in Y_1^*, uw_2 - c_2 \in Y_2^*\} \subseteq \bar{D}_\rho := \{u \in \mathbb{R} : uw_2 - c_2 \in Y_2^*\}$$

for the so-called *dual feasible sets.*

Theorem 2.4. *Let ρ be a functional of the form (2.1) on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty)$. Assume*
(i) complete recourse: $\langle w_2, Y_2 \rangle = \mathbb{R}$,
(ii) dual feasibility: $D_\rho \neq \emptyset$.
Then ρ is finite, Lipschitz continuous, and convex. Further, it holds that

$$\rho(z) = \inf_{y_1 \in Y_1} \left\{ \langle c_1, y_1 \rangle + \mathbb{E}\left[\max_{\ell=1,2} u_\ell \left(z - \langle w_1, y_1 \rangle\right) \right] \right\} \tag{2.2}$$

[1] Y_t^* is the polar cone of Y_t. For a nonempty set Y the polar cone Y^* is defined by $Y^* = \{y^* : \langle y, y^* \rangle \leq 0 \, \forall \, y \in Y\}$.

with two real numbers u_1 and u_2 that are the endpoints of $\bar{D}_\rho$ which is a compact interval in $\mathbb{R}$. Furthermore, if $p > 1$, then ρ admits the dual representation

$$\rho(z) = \sup\left\{-\mathbb{E}[\lambda z] + \inf_{y_1 \in Y_1} \langle c_1 + \mathbb{E}[\lambda] w_1, y_1\rangle : \lambda \in L_{p'}(\Omega, \mathcal{F}, \mathbb{P}), -(c_2 + \lambda w_2) \in Y_2^*\right\}, \tag{2.3}$$

where $\frac{1}{p} + \frac{1}{p'} = 1$. In particular, if Y_1 is a cone, then ρ is positively homogeneous and (2.3) becomes

$$\rho(z) = \sup\{-\mathbb{E}[\lambda z] : \lambda \in L_{p'}(\Omega, \mathcal{F}, \mathbb{P}), -(c_1 + \mathbb{E}[\lambda] w_1) \in Y_1^*, -(c_2 + \lambda w_2) \in Y_2^*\}. \tag{2.4}$$

Proof. Finiteness, Lipschitz continuity, and the representations (2.3) and (2.4) will be proved in a more general framework in Section 3, Theorem 3.9.
Representation (2.2) follows from LP duality applied to the second stage program. Namely, it holds for each $y_1 \in Y_1$ and each $z \in \mathbb{R}$ that

$$\min\{\langle c_2, y_2\rangle : y_2 \in Y_2, \langle w_1, y_1\rangle + \langle w_2, y_2\rangle = z\} = \max\{u(z - \langle w_1, y_1\rangle) : uw_2 - c_2 \in Y_2^*\}.$$

Due to complete recourse and dual feasibility the feasible sets of both problems are nonempty and the joint optimal value is finite for each $y_1 \in Y_1$ and each $z \in \mathbb{R}$. Since the expression $z - \langle w_1, y_1\rangle$ can reach any real number and the feasible set of the right problem $\bar{D}_\rho = \{u \in \mathbb{R} : uw_2 - c_2 \in Y_2^*\}$ does not depend on y_1 and z it is clear that the latter is bounded, i.e., a compact interval in $\mathbb{R}$. Of course, the maximum is attained for u being an endpoint of $\bar{D}_\rho$.
Convexity of ρ follows from the fact that the real-valued function

$$(y_1, z) \mapsto \langle c_1, y_1\rangle + \mathbb{E}\left[\max_{\ell=1,2} u_\ell (z - \langle w_1, y_1\rangle)\right]$$

is convex on $Y_1 \times L_p(\Omega, \mathcal{F}, \mathbb{P})$. Positive homogeneity of ρ holds if Y_1 is a cone. □

If a functional ρ on $L_1(\Omega, \mathcal{F}, \mathbb{P})$ is defined by formula (2.1), the question arises, for which choice of c_t,w_t and Y_t ($t = 1, 2$) this functional is a (convex) risk measure in the sense of [10]. Formula (2.4) provides a sufficient criterion for a functional of the form (2.1) to be a coherent risk measure:

Corollary 2.5. Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form (2.1) with Y_1 being a polyhedral cone and $1 < p < \infty$. Let the conditions of Theorem 2.4 be satisfied (complete recourse, dual feasibility) and assume that $\Lambda_\rho \subseteq \mathcal{D}$ where Λ_ρ is defined by

$$\Lambda_\rho := \{\lambda \in L_{p'}(\Omega, \mathcal{F}, \mathbb{P}) : -(c_1 + \mathbb{E}[\lambda] w_1) \in Y_1^*, -(c_2 + \lambda w_2) \in Y_2^*\}. \tag{2.5}$$

Then ρ is a coherent risk measure.

Proof. Regarding Theorem 2.1 and Theorem 2.4 it remains to show that Lipschitz continuity in $L_1(\Omega, \mathcal{F}, \mathbb{P})$ implies condition (2.1). This, however, follows by the monotone convergence theorem: Let $z_n \uparrow z_0$, thus $|z_0 - z_n| = z_0 - z_n \downarrow 0$. Hence, $|\rho(z_0) - \rho(z_n)| \leq L \cdot \mathbb{E}[|z_0 - z_n|] \to 0$. □

The following result provides another sufficient criterion for a functional of the form (2.1) to be a convex risk measure in case Y_1 is not a cone:

Proposition 2.6. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form (2.1) with $p \in [1, \infty)$. Assume that complete recourse and dual feasibility hold and that $\bar{D}_\rho \subseteq \mathbb{R}_-$ and let c_1, w_1 and Y_1 be of the form $c_1 = (\hat{c}_1, 1)$, $w_1 = (\hat{w}_1, -1)$ and $Y_1 = \hat{Y}_1 \times \mathbb{R}$, where $\hat{w}_1, \hat{c}_1 \in \mathbb{R}^{k_1 - 1}$ and $\hat{Y}_1 \subseteq \mathbb{R}^{k_1 - 1}$.
Then ρ is a (polyhedral) convex risk measure on $L_p(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. The monotonicity property (i) follows from the representation (2.2) and the fact that u_1 and u_2 are nonpositive. Indeed, let $z, \tilde{z} \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ be such that $z \leq \tilde{z}$. Then we have

$$\mathbb{E}\left[\max_{\ell=1,2} u_\ell (z - \langle w_1, y_1\rangle)\right] \geq \mathbb{E}\left[\max_{\ell=1,2} u_\ell (\tilde{z} - \langle w_1, y_1\rangle)\right], \quad \forall y_1 \in Y_1.$$

The translation invariance condition (ii) follows by setting $\tilde{y}_1^{(k_1)} := y_1^{(k_1)} + r$ as a consequence of the identity

$$\begin{aligned}\rho(z+r) &= \inf\left\{\langle \hat{c}_1, \hat{y}_1\rangle + y_1^{(k_1)} + \mathbb{E}\left[\max_{\ell=1,2} u_\ell\left(z + r - \langle \hat{w}_1, \hat{y}_1\rangle + y_1^{(k_1)}\right)\right],\ \hat{y}_1 \in \hat{Y}_1,\ y_1^{(k_1)} \in \mathbb{R}\right\}\\ &= \inf\left\{\langle \hat{c}_1, \hat{y}_1\rangle + \tilde{y}_1^{(k_1)} + \mathbb{E}\left[\max_{\ell=1,2} u_\ell\left(z - \langle \hat{w}_1, \hat{y}_1\rangle + \tilde{y}_1^{(k_1)}\right)\right],\ \hat{y}_1 \in \hat{Y}_1, \tilde{y}_1^{(k_1)} \in \mathbb{R}\right\} - r\\ &= \rho(z) - r\end{aligned}$$

for each $r \in \mathbb{R}$ and $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$. Finiteness and convexity of ρ follow from Theorem 2.4. □

The assumptions of Proposition 2.6 guarantee even a stronger type of monotonicity than imposed earlier for risk measures. Such stronger monotonicity properties are based on so-called *integral stochastic orders* or *stochastic dominance rules* (see [13] for a recent survey). For real random variables z and $\tilde{z}$ in $L_1(\Omega, \mathcal{F}, \mathbb{P})$, stochastic dominance rules are defined by classes $\mathcal{F}$ of measurable real-valued functions on $\mathbb{R}$. A stochastic dominance rule is defined by

$$z \preceq_{\mathcal{F}} \tilde{z} \quad \text{if} \quad \mathbb{E}[f(z)] \leq \mathbb{E}[f(\tilde{z})]$$

for each $f \in \mathcal{F}$ such that the expectations exist. Important special cases are the classes $\mathcal{F}_{nd}$ and $\mathcal{F}_{ndc}$ of nondecreasing and of nondecreasing concave functions, respectively. In these cases the rules are also called *first order stochastic dominance* and *second order stochastic dominance* and denoted by $\preceq_{FSD}$ and $\preceq_{SSD}$, respectively. Clearly, $z \preceq_{FSD} \tilde{z}$ implies $z \preceq_{SSD} \tilde{z}$. The relation $z \preceq_{FSD} \tilde{z}$ is equivalent to $\mathbb{P}(z > t) \leq \mathbb{P}(\tilde{z} > t)$ for each $t \in \mathbb{R}$. Furthermore, $z \preceq_{SSD} \tilde{z}$ is equivalent to the condition $\mathbb{E}[\min\{z, t\}] \leq \mathbb{E}[\min\{\tilde{z}, t\}]$ for each $t \in \mathbb{R}$ (cf. [13, Section 8]).
In [15, 16] the consistency of risk measures ρ with certain stochastic dominance rules $\preceq_{\mathcal{F}}$ is studied. In particular, it is said that ρ is *consistent with second order stochastic dominance* if $z \preceq_{SSD} \tilde{z}$ implies $\rho(z) \geq \rho(\tilde{z})$.

Proposition 2.7. *Let ρ be a functional on $L_p(\Omega, \mathcal{F}, \mathbb{P})$ of the form (2.1) with $p \in [1, \infty)$. Assume that complete recourse and dual feasibility hold and that $\bar{D}_\rho \subseteq \mathbb{R}_-$. Then ρ is consistent with second order stochastic dominance.*

Proof. Due to Theorem 2.4 the representation (2.2) holds with $u_1, u_2 \in \mathbb{R}_-$. Define for $y_1 \in Y_1$ the real-valued function g_{y_1} given by $g_{y_1}(t) := \langle c_1, y_1\rangle + \max_{\ell=1,2} u_\ell\,(t - \langle w_1, y_1\rangle)$ for $t \in \mathbb{R}$. Note that g_{y_1} is convex and, because of $u_1, u_2 \leq 0$, nonincreasing. Let $z \preceq_{SSD} \tilde{z}$. Then $\mathbb{E}\left[-g_{y_1}(z)\right] \leq \mathbb{E}\left[-g_{y_1}(\tilde{z})\right]$ for all $y_1 \in Y_1$ and, thus, $\rho(z) = \inf_{y_1 \in Y_1} \mathbb{E}\left[g_{y_1}(z)\right] \geq \inf_{y_1 \in Y_1} \mathbb{E}\left[g_{y_1}(\tilde{z})\right] = \rho(\tilde{z})$. □

Remark 2.8. For a risk measure ρ the *acceptance set* $\mathcal{A}_\rho$ is given by $\mathcal{A}_\rho = \{z \in \mathcal{Z} : \rho(z) \leq 0\}$ [2, 10]. Let the conditions of the Theorem 2.4 be satisfied. Then, since ρ is a convex functional, $\mathcal{A}_\rho$ is a convex set. If, in addition, Y_1 is a cone then $\mathcal{A}_\rho$ is a convex cone. Regarding (2.5) it is obvious that $\mathcal{A}_\rho = \{z \in L_p(\Omega, \mathcal{F}, \mathbb{P}) \,|\, \forall\, \lambda \in \Lambda_\rho : \mathbb{E}\left[\lambda z\right] \geq 0\} = -\Lambda_\rho^*$ in this case. Of course, if $\Omega = \{\omega_1, ..., \omega_S\}$, then Λ_ρ is a polyhedron in $\mathbb{R}^S$, thus $\mathcal{A}_\rho = -\Lambda_\rho^*$ is a polyhedral cone.

Example 2.9. We consider the *Conditional-* or *Average-Value-at-Risk* at level α ($CVaR_\alpha$ or $AVaR_\alpha$) defined by

$$CVaR_\alpha(z) := \frac{1}{\alpha}\int_0^\alpha VaR_\gamma(z)d\gamma = \inf_{r \in \mathbb{R}}\left\{r + \frac{1}{\alpha}\mathbb{E}\left[(r+z)^-\right]\right\},$$

where $VaR_\alpha(z) := \inf\{r \in \mathbb{R} : \mathbb{P}(z + r < 0) \leq \alpha\}$ is the Value-at-Risk at level $\alpha \in (0,1)$ (see [10, Section 4.4] and [24]) and $a^- = \max\{0, -a\}$ denotes the negative part of a real number a. Note that

$$\inf_{r \in \mathbb{R}}\left\{r + \frac{1}{\alpha}\mathbb{E}\left[(r+z)^-\right]\right\} = \inf\left\{y_1 + \frac{1}{\alpha}\mathbb{E}\left[y_2^{(1)}\right] \;\middle|\; \begin{array}{c} y_1 \in \mathbb{R},\ y_2 \in \mathbb{R}_+ \times \mathbb{R}_+,\\ -y_2^{(1)} + y_2^{(2)} = z + y_1 \end{array}\right\}, \tag{2.6}$$

thus, $CVaR_\alpha$ is of the form (2.1) by setting $k_1 = 1$, $k_2 = 2$, $w_1 = -1$, $c_1 = 1$, $c_2 = \left(\frac{1}{\alpha}, 0\right)$, $w_2 = (-1, 1)$, $Y_1 = \mathbb{R}$ and $Y_2 = \mathbb{R}_+^2$, and, hence, it is polyhedral. Moreover, $\langle w_2, Y_2\rangle = \mathbb{R}$, $D_\rho = \{-1\}$ and $\bar{D}_\rho = \left[-\frac{1}{\alpha}, 0\right] \subseteq \mathbb{R}_-$, thus the dual representation (2.4) holds and $CVaR_\alpha$ is consistent with second order stochastic dominance. The representation (2.2) holds with $u_1 = -\frac{1}{\alpha}$ and $u_2 = 0$. The condition $-(c_2 + \lambda w_2) \in Y_2^*$ in the dual representation (2.4) is equivalent to $\lambda \in \left[0, \frac{1}{\alpha}\right]$. Hence, (2.4) becomes

$$CVaR_\alpha(z) = \sup\left\{-\mathbb{E}\left[\lambda z\right] : \lambda \in L_{p'}(\Omega, \mathcal{F}, \mathbb{P}),\ \lambda \in \left[0, \tfrac{1}{\alpha}\right],\ \mathbb{E}\left[\lambda\right] = 1\right\} \quad (2.7)$$

for each $z \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, $1 < p < +\infty$. Corollary 2.5 applies, thus, $CVaR$ is a coherent risk measure, too.

Example 2.10. Consider the *expected regret* or *expected loss* defined by

$$\rho(z) = \mathbb{E}\left[(z - \gamma)^-\right]$$

with some fixed threshold $\gamma \in \mathbb{R}$. This functional, too, can be written in the the form (2.1) with $k_1 = 1$, $k_2 = 2$, $w_1 = 1$, $c_1 = 0$, $c_2 = (1, 0)$, $w_2 = (-1, 1)$, $Y_1 = \{\gamma\}$, $Y_2 = \mathbb{R}_+ \times \mathbb{R}_+$. Note that, actually, Y_1 is *not* a cone here. Further, $\langle w_2, Y_2 \rangle = \mathbb{R}$, $D_\rho \neq \emptyset$ and $\bar{D}_\rho = [-1, 0] \subseteq \mathbb{R}_-$, thus the dual representations (2.2) and (2.3) hold and ρ is consistent with second order stochastic dominance. However, ρ is not translation invariant, i.e. not a risk measure in the sense of [9, 10]. Nevertheless, it is used as a risk measure in some applications.

3. Multiperiod Risk

When random variables $z_1, ..., z_T$ with $z_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P})$, $p \geq 1$, are considered and the available information is revealed with the passing of time, it may become necessary to use multiperiod risk measures (see [1, 19]). We assume that a filtration of σ-fields $\mathcal{F}_t$, $t = 1, ..., T$, is given, i.e., $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$, and that $\mathcal{F}_1 = \{\emptyset, \Omega\}$, i.e., that z_1 is always deterministic. We will now generalize the concepts of the previous section to this multiperiod framework.

Remark 3.1. When dealing with multiperiod risk measures one has to determine whether the random variables represent (potentially financial) *incomes* or *payments* as, e.g., in [19, 34, 39], or if they have to be understood in a cumulative sense, i.e., as a *wealth* or *value process* as in [2, 3]. Of course, the one can easily be transformed into the other: If Z_t is an income, then one can consider accumulation $z_t = Z_1 + ... + Z_t$, if z_t is an accumulated value, then the income is given by $Z_t = z_t - z_{t-1}$. Throughout this paper we consider $z = (z_1, ..., z_T)$ to be a value process.

We give the definition of coherence in the multiperiod case as introduced[2] in [2, 3].

Definition 3.2. A functional ρ on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ is called *multiperiod coherent risk measure* if:
(i) If $z_t \leq \tilde{z}_t$ a.s., $t = 1, ..., T$, then $\rho(z_1, ..., z_T) \geq \rho(\tilde{z}_1, ..., \tilde{z}_T)$ *(monotonicity)*
(ii) for each $r \in \mathbb{R}$ we have $\rho(z_1 + r, ..., z_T + r) = \rho(z) - r$ *(translation invariance)*
(iii) $\rho(\mu z_1 + (1-\mu)\tilde{z}_1, ..., \mu z_T + (1-\mu)\tilde{z}_T) \leq \mu\rho(z_1, ..., z_T) + (1-\mu)\rho(\tilde{z}_1, ..., \tilde{z}_1)$ for $\mu \in [0, 1]$ *(convexity)*
(iv) for $\mu \geq 0$ we have $\rho(\mu z_1, ..., \mu z_T) = \mu\rho(z_1, ..., z_T)$ *(positive homogeneity)*.

Remark 3.3. How translation invariance is to be defined in the multiperiod case is still subject to discussion in the ongoing research in financial mathematics. Different suggestions were made, e.g., in [34, 22, 39] such that nonrandom amounts can be shifted in time by means of credits. However, from the viewpoint of capital requirement and optimization it appears reasonable to keep with [2, 3].

Example 3.4. In [2], Example 3, it was shown that

$$\rho(z) := -\mathbb{E}\left[\min_{1 \leq t \leq T} z_t\right]$$

is a multiperiod coherent risk measure on $\times_{t=1}^T L_\infty(\Omega, \mathcal{F}_t, \mathbb{P})$.

[2] In [2, 3] the definition is slightly different since another framework was considered: The first time stage (i.e. the deterministic stage) was denoted by index 0. Here, the formulation is adapted to our framework with index 1 for the deterministic time stage (i.e., $\mathcal{F}_1 = \{\emptyset, \Omega\}$).

Remark 3.5. Let ρ_t be (one-period) coherent risk measures on $L_p(\Omega, \mathcal{F}_t, \mathbb{P})$, $t = 1, ..., T$. Let further $\emptyset \neq S \subseteq \{1, ..., T\}$. Then

$$\rho(z_1, ..., z_T) = \max_{t \in S} \rho_t(z_t)$$

is multiperiod coherent. Let further $\mu_t \in \mathbb{R}_+$, $t = 1, ..., T$ and $\sum_{t=1}^T \mu_t = 1$. Then also

$$\rho(z_1, ..., z_T) = \sum_{t=1}^{T} \mu_t \rho_t(z_t)$$

is a multiperiod coherent risk measure. This can easily be verified by checking the four properties of Definition 3.2.

As shown in [2, 3], the representation result for (one-period) risk measures (Theorem 2.1) can be carried over to the multiperiod case:

Theorem 3.6. *Let* $\rho : \times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P}) \to \bar{\mathbb{R}}$ *and set*

$$\mathcal{D}_T := \left\{ f \in \times_{t=1}^T L_1(\Omega, \mathcal{F}_t, \mathbb{P}) : f_t \geq 0 \ (t = 1, \ldots, T), \ \textstyle\sum_{t=1}^T \mathbb{E}\left[f_t\right] = 1 \right\}.$$

Assume that ρ *satisfies the following continuity property:*

$$z_{t,n} \uparrow z_{t,0} \ a.s. \, (t = 1, ..., T) \ \Rightarrow \ \lim_{n \to \infty} \rho(z_{1,n}, ..., z_{T,n}) = \rho(z_{1,0}, ..., z_{T,0}).$$

Then, the following equivalence holds:

$$\rho \text{ multiperiod coherent risk measure } \iff \exists \, \mathcal{P}_\rho \subseteq \mathcal{D}_T \text{ convex} : \rho(z) = \sup \left\{ -\textstyle\sum_{t=1}^T \mathbb{E}\left[f_t z_t\right] : f \in \mathcal{P}_\rho \right\}$$

Proof. We follow the ideas of [2, 3], but in reverse order. Obviously, ρ is coherent if and only if the corresponding one-period risk measure ρ' on $L_p(\Omega', \mathcal{F}', \mathbb{P}')$ is coherent in the usual sense, where $(\Omega', \mathcal{F}', \mathbb{P}')$ and ρ' are defined as follows:

$$\begin{aligned} &\Omega' := \Omega \times \{1, ..., T\} \\ &\mathcal{F}' := \left\{ \textstyle\bigcup_{t=1}^T \left(A_t \times \{t\}\right) : A_t \in \mathcal{F}_t \right\} \\ &\mathbb{P}' \left(\textstyle\bigcup_{t=1}^T \left(A_t \times \{t\}\right) \right) := \tfrac{1}{T} \textstyle\sum_{t=1}^T \mathbb{P}(A_t) \\ &\rho'(z') := \rho\left(z(z')\right) \end{aligned}$$

and $z(z')$ is defined by $z(z')(\omega) := (z'(\omega, 1), z'(\omega, 2), ..., z'(\omega, T))$. Theorem 2.1 says that there exists a convex set of density functions $\mathcal{P}'_\rho \subseteq \mathcal{D}$ such that, for $z \in \times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$,

$$\rho(z) = \rho'(z'(z)) = \sup\left\{ -\mathbb{E}'\left[f' z'\right] : f' \in \mathcal{P}' \right\}$$

with $z'(z)(\omega, t) := z_t(\omega)$. Note that also the conditions from Definition 3.2 are equivalent to those from Theorem 2.1 for $(\Omega', \mathcal{F}', \mathbb{P}')$ here. By setting

$$\mathcal{P}_\rho := \left\{ f = \left(\frac{1}{T} f'(.,1), \frac{1}{T} f'(.,2), ..., \frac{1}{T} f'(.,T) \right) : f' \in \mathcal{P}'_\rho \right\}$$

the assertion follows. □

Now we are ready to extend Definition 2.2 to the multiperiod case.

Definition 3.7. A multiperiod risk measure ρ on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ with $p \in [1, \infty]$ is called *multiperiod polyhedral* if there are $k_t \in \mathbb{N}$, $c_t \in \mathbb{R}^{k_t}$, $t = 1, \ldots, T$, $w_{t\tau} \in \mathbb{R}^{k_{t-\tau}}$, $t = 1, \ldots, T$, $\tau = 0, ..., t-1$, a polyhedral set $Y_1 \subseteq \mathbb{R}^{k_1}$, and polyhedral cones $Y_t \subseteq \mathbb{R}^{k_t}$, $t = 2, \ldots, T$, such that

$$\rho(z) = \inf \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \langle c_t, y_t \rangle \right] \;\middle|\; \begin{array}{ll} y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), & \\ y_t \in Y_t, & (t = 1, \ldots, T) \\ \sum_{\tau=0}^{t-1} \langle w_{t,\tau}, y_{t-\tau} \rangle = z_t & \end{array} \right\}. \tag{3.1}$$

Remark 3.8. The reader might wonder why, for $T = 2$, this definition does not precisely coincide with the Definition 2.2 for the one-period case. This is due to the fact that, in the literature, the risk of a process $z_1, ..., z_T$ is allowed to depend also on z_1 although this value is constant, i.e., deterministic (see [2, 3, 22]), whereas one-period risk depends on one scalar random variable only. Nevertheless, the one-period case can be regarded as a special case of Definition 3.7 because for $T = 2$ the parameters Y_1, c_1 and $w_{1,0}$ can easily be set such that z_1 does not contribute to the optimal value of (3.1).

Theorem 3.9. *Let ρ be a functional of the form (3.1) on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$. Assume*
(i) complete recourse: $\langle w_{t,0}, Y_t\rangle = \mathbb{R}$ $(t = 1, ..., T)$,
(ii) dual feasibility: $\left\{u \in \mathbb{R}^T : -\left(c_t + \sum_{\nu=t}^{T} u_\nu w_{\nu,\nu-t}\right) \in Y_t^* \ (t = 1, ..., T)\right\} \neq \emptyset$.
Then ρ is Lipschitz continuous on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ and the following dual representation holds whenever $p \in (1, \infty)$ and $\frac{1}{p} + \frac{1}{p'} = 1$:

(3.2)
$$\rho(z) = \sup \left\{ \begin{array}{l} -\mathbb{E}\left[\sum_{t=1}^{T} \lambda_t z_t\right] \\ + \inf\limits_{y_1 \in Y_1} \left\langle c_1 + \sum_{\nu=1}^{T} \mathbb{E}\left[\lambda_\nu\right] w_{\nu,\nu-1}, y_1 \right\rangle \end{array} \middle| \begin{array}{l} \lambda_t \in L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \ (t = 1, \dots, T), \\ -\left(c_t + \sum_{\nu=t}^{T} \mathbb{E}\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t}\right) \in Y_t^* \\ (t = 2, ..., T) \end{array} \right\}.$$

If, in addition, Y_1 is a polyhedral cone, (3.2) simplifies to

$$\rho(z) = \sup \left\{ -\mathbb{E}\left[\sum_{t=1}^{T} \lambda_t z_t\right] \middle| \begin{array}{l} \lambda_t \in L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}), \\ -\left(c_t + \sum_{\nu=t}^{T} \mathbb{E}\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t}\right) \in Y_t^* \end{array} \ (t = 1, ..., T) \right\}. \tag{3.3}$$

Proof. a) We first prove the Lipschitz continuity of ρ. Let $\bar{z}, \tilde{z} \in \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ and $\epsilon > 0$. Due to complete recourse and dual feasibility $\rho(\bar{z})$ and $\rho(\tilde{z})$ are finite (see part b) of the proof) and there exists $\bar{y} \in \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t})$ such that $\bar{y}_t \in Y_t$ and $\sum_{\tau=0}^{t-1} \langle w_{t,\tau}, \bar{y}_{t-\tau}\rangle = \bar{z}_t$ for all $t = 1, ..., T$, i.e., $\bar{y}$ is feasible, and $f(\bar{y}) \leq \rho(\bar{z}) + \epsilon$, i.e., $\bar{y}$ is ϵ-optimal. Here, f is defined by

$$f : \times_{t=1}^{T} L_p\left(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}\right) \to \mathbb{R}, \quad f(y) := \mathbb{E}\left[\sum_{t=1}^{T} \langle c_t, y_t\rangle\right].$$

Next we show that there exist a constant $L_M > 0$ and an element $\tilde{y} \in \times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t})$ such that $\|\bar{y} - \tilde{y}\|_{\mathbb{R}^K} \leq L_M \|\bar{z} - \tilde{z}\|_{\mathbb{R}^T}$ and that $\tilde{y}_t \in Y_t$ and $\sum_{\tau=0}^{t-1} \langle w_{t,\tau}, \tilde{y}_{t-\tau}\rangle = \tilde{z}_t$ for $t = 1, ..., T$.
To prove this we first note that, for each $t = 1, \dots, T$, the graph of the set-valued mapping M_t from $\mathbb{R}$ to $\mathbb{R}^{k_t}$ given by

$$M_t(u) := \{y_t \in Y_t : \langle w_{t,0}, y_t\rangle = u\} \quad (u \in \mathbb{R})$$

is polyhedral. Hence, M_t is Lipschitz continuous with respect to the Hausdorff distance d_H on the closed subsets of $\mathbb{R}^{k_t}$ (see [38], [27, Example 9.35]), i.e., there exist constants $L_{M_t} > 0$ such that

$$d_H\left(M_t(\bar{u}), M_t(\tilde{u})\right) = \max\left\{ \sup_{y \in M_t(\bar{u})} d\left(y, M_t(\tilde{u})\right), \sup_{y \in M_t(\tilde{u})} d\left(y, M_t(\bar{u})\right) \right\} \leq L_{M_t} |\bar{u} - \tilde{u}|$$

holds for all $\bar{u}$ and $\tilde{u}$ in $\mathbb{R}$. Thus, for all $\bar{u}, \tilde{u} \in \mathbb{R}$ and $\bar{y} \in M_t(\bar{u})$, we have that

$$L_{M_t} |\bar{u} - \tilde{u}| \geq \sup_{y \in M_t(\bar{u})} d\left(y, M_t(\tilde{u})\right) \geq d\left(\bar{y}, M_t(\tilde{u})\right) = \min_{y \in M_t(\tilde{u})} \|\bar{y} - y\|. \tag{3.4}$$

We prove the existence of $\tilde{y}$ by verifying the following assertion by induction with respect to t:

$$\begin{array}{l} \forall\, t \in \{1, ..., T\} \ \exists\, L_t > 0 \ \exists\, \tilde{y}_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}) : \\ 1.\ \|\bar{y}_t - \tilde{y}_t\| \leq L_t \left\|(\bar{z}_1, ..., \bar{z}_t) - (\tilde{z}_1, ..., \tilde{z}_t)\right\|, \ \mathbb{P}-\text{a.s.}, \\ 2.\ \tilde{y}_t \in Y_t, \ \sum_{\tau=0}^{t-1} \langle w_{t,\tau}, \tilde{y}_{t-\tau}\rangle = \tilde{z}_t\,, \ \mathbb{P}-\text{a.s.}. \end{array}$$

$t = 1$: Set $\bar{u} := \bar{z}_1$, $\tilde{u} := \tilde{z}_1$, then, due to (3.4), $\exists\, \tilde{y}_1 \in M_1(\tilde{u}) : \|\bar{y}_1 - \tilde{y}_1\| \leq L_{M_1} |\bar{u} - \tilde{u}| = L_{M_1} |\bar{z}_1 - \tilde{z}_1|$
$t - 1 \curvearrowright t$: We set $\bar{u} := \bar{z}_t - \sum_{\tau=1}^{t-1} \langle w_{t,\tau}, \bar{y}_{t-\tau}\rangle$, $\tilde{u} := \tilde{z}_t - \sum_{\tau=1}^{t-1} \langle w_{t,\tau}, \tilde{y}_{t-\tau}\rangle$ and consider the following set-valued mappings from Ω to $\mathbb{R}^{k_t}$ given by

$$\omega \mapsto M_t\left(\tilde{u}(\omega)\right) \qquad \text{and} \qquad \omega \mapsto \min_{y \in M_t(\tilde{u}(\omega))} \|\bar{y}_t(\omega) - y\|.$$

Both are measurable with respect to the σ-field $\mathcal{F}_t$ due to the induction hypothesis and well known measurability results for set-valued mappings (e.g., [27, Theorem 14.36]). Hence, due to [27, Theorem 14.37] there exists a measurable selection $\tilde{y}_t$ of the second mapping. From (3.4) and the induction hypothesis we obtain the estimate

$$\begin{aligned} \|\bar{y}_t(\omega) - \tilde{y}_t(\omega)\| &\leq L_{M_t} \left|\bar{u}(\omega) - \tilde{u}(\omega)\right| \\ &= L_{M_t} \left|\bar{z}_t(\omega) - \tilde{z}_t(\omega) - \textstyle\sum_{\tau=1}^{t-1} \langle w_{t,\tau}, \bar{y}_{t-\tau}(\omega) - \tilde{y}_{t-\tau}(\omega)\rangle\right| \\ &\leq L_{M_t} \left(|\bar{z}_t(\omega) - \tilde{z}_t(\omega)| + \textstyle\sum_{\tau=1}^{t-1} \|w_{t,\tau}\| \, \|\bar{y}_{t-\tau}(\omega) - \tilde{y}_{t-\tau}(\omega)\|\right) \\ &\leq L_{M_t} \hat{L}_t \left\|(\bar{z}_1(\omega), ..., \bar{z}_t(\omega)) - (\tilde{z}_1(\omega), ..., \tilde{z}_t(\omega))\right\| \end{aligned}$$

with some constant $\hat{L}_t$. Moreover, $\tilde{y}_t(.)$ belongs to L_p since $\bar{y}$, $\bar{z}$ and $\tilde{z}$ do so. Thus, the induction step is proved and there exists a constant L_M (only depending on T and $L_1, \dots, L_T$) such that

$$\|\bar{y}(\omega) - \tilde{y}(\omega)\| \leq L_M \|\bar{z}(\omega) - \tilde{z}(\omega)\| \quad (\omega \in \Omega).$$

Of course, the objective f is Lipschitz continuous, say, with modulus $L_f > 0$. Hence, we obtain

$$\begin{aligned} \rho(\tilde{z}) - \rho(\bar{z}) &\leq \rho(\tilde{z}) - f(\bar{y}) + \epsilon \leq f(\tilde{y}) - f(\bar{y}) + \epsilon \leq |f(\tilde{y}) - f(\bar{y})| + \epsilon \\ &\leq L_f \mathbb{E}\left[\|\tilde{y} - \bar{y}\|_{\mathbb{R}^K}\right] + \epsilon \leq L_f \left(L_M \mathbb{E}\left[\|\bar{z} - \tilde{z}\|_{\mathbb{R}^T}\right]\right) + \epsilon \\ &= L_f \left(L_M \|\bar{z} - \tilde{z}\|_{\times_{t=1}^T L_1(\Omega, \mathcal{F}_t, \mathbb{P})}\right) + \epsilon. \end{aligned}$$

Since ϵ was chosen arbitrarily, we have $\rho(\tilde{z}) - \rho(\bar{z}) \leq L_f L_M \|\bar{z} - \tilde{z}\|$. Changing the role of $\bar{z}$ and $\tilde{z}$ leads to $|\rho(\tilde{z}) - \rho(\bar{z})| \leq L_f L_M \|\bar{z} - \tilde{z}\|$, i.e., Lipschitz continuity holds on $\times_{t=1}^T L_1(\Omega, \mathcal{F}_t, \mathbb{P})$ with modulus $L = L_f L_M$. In turn, this implies Lipschitz continuity on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$.

b) Now we prove the dual representation (3.3) of ρ. We make use of the results of [8] and [23], see also [26]. The stochastic program (3.1) is of the form[3] (SP) in [8, Section 3], i.e.,

$$\min \left\{ \langle y, c^+ \rangle \,|\, y \in K_E,\, Ty - b \in K_F \right\}$$

with the spaces $E = L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^K)$, $K := \sum_{t=1}^T k_t$, $F = L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^T)$, dual spaces $E^+ = L_{p'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^K)$, $F^+ = L_{p'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^T)$, cones $\Omega_E = \{y \in E : y_t \in Y_t \text{ a.s. } (t = 1, ..., T)\}$, $\Omega_F = \{0\} \subseteq F$, $K_E = P_E E \cap \Omega_E = P_E \Omega_E$ and $K_F = P_F^{-1} \Omega_F = \{v \in F : \mathbb{E}\left[v_t | \mathcal{F}_t\right] = 0 \, (t = 1, ..., T)\}$. Here, P_E is the projection given by $P_E y = \left(\mathbb{E}\left[y_1|\mathcal{F}_1\right], \mathbb{E}\left[y_2|\mathcal{F}_2\right], \dots, \mathbb{E}\left[y_T|\mathcal{F}_T\right]\right)'$ and P_F, P_{E^+}, P_{F^+} are defined analogously, $c^+ = (c_1', ..., c_T')' \in E^+$ (constant on Ω), $b = (z_1, ..., z_T)' \in F$ and $T : E \to F$ is the mapping defined by $(Tu)(\cdot) = A\left(u(\cdot)\right)$ with the matrix

$$A = \begin{pmatrix} w'_{1,0} & 0 & 0 & \cdots & 0 \\ w'_{2,1} & w'_{2,0} & 0 & \cdots & 0 \\ w'_{3,2} & w'_{3,1} & w'_{3,0} & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ w'_{T,T-1} & w'_{T,T-2} & w'_{T,T-3} & \cdots & w'_{T,0} \end{pmatrix} \in \mathbb{R}^{T \times K}.$$

Note that A is a lower block triangular matrix that is constant on Ω. The corresponding dual program (SP*) in [8] is of the form

$$\max \left\{ \langle b, v^+ \rangle \,|\, c^+ - T^+ v^+ \in K_E^+, v^+ \in K_F^+ \right\}.$$

Further, it is shown in [8] that the dual cones are of the form $K_F^+ = P_{F^+} \Omega_F^+ = P_{F^+} F^+ \cap \Omega_F^+$ and $K_E^+ = (P_{E^+})^{-1} \Omega_E^+ = \left\{y^+ \in E^+ : \left(\mathbb{E}\left[y_1^+|\mathcal{F}_1\right], \dots, \mathbb{E}\left[y_T^+|\mathcal{F}_T\right]\right) \in \Omega_E^+\right\}$.
Translated back into our notation this becomes (setting $\lambda = -v^+$)

$$\max \left\{ -\mathbb{E}\left[\textstyle\sum_{t=1}^T \lambda_t z_t\right] \;\middle|\; \begin{array}{l} c^+ + A'\lambda \in K_E^+ \\ \lambda_t \in L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}),\ t = 1, \dots, T \end{array} \right\} =: \bar{\rho}(z).$$

[3]We use the superscript $+$ here to indicate the dual objects, i.e., dual space, vectors in dual spaces, adjoint operator, dual cone, whereas the superscript $*$ denotes the polar cone and A', w' are the transpose of matrix A and vector w, respectively. Note that for a cone C it holds that $C^+ = -C^*$ since $C^+ = \left\{c^+ : \langle c, c^+ \rangle \geq 0 \,\forall\, c \in C\right\}$.

This is exactly (3.3) since A and c^+ are constant on Ω (thus $(T^+v^+)(\omega) = A'v^+(\omega)$) and it holds that $\Omega_E^+ = \{y^+ \in E^+ : -y_t^+ \in Y_t^* \ \mathbb{P} - \text{a.s.}\}$.
It remains to prove that (3.1) and (3.3) have the same optimal values, i.e., strong duality, $\bar{\rho}(z) = \rho(z)$. To this end, note that (3.3) is the (concave) dual in E^+ of the convex program (3.1) in the reflexive real Banach space E. Complete recourse, dual feasibility, and Lemma 1 in [23] ensure that

$$-\infty < \bar{\rho}(z) \le \rho(z) < +\infty$$

(weak duality). Due to Theorem 7 in [23], normality implies strong duality. Normality means (cf. [23]) that for the perturbed problem given by

$$\min\left\{\langle y, c^+\rangle \,|\, y \in K_E,\, Ty - b - a \in K_F\right\} = \rho\,(z + P_E a)$$

with perturbation $a \in E$ it holds that $\liminf_{a\to 0} \rho\,(z + P_E a) = \rho\,(z)$. The latter condition, however, is satisfied since ρ is even Lipschitz continuous. Thus it holds indeed $\bar{\rho}(z) = \rho(z)$.

c) Finally, we briefly sketch the extension of the proof for the dual representation (3.2), i.e., for the case that $Y_1 \subseteq \mathbb{R}^{k_1}$ is *not* a cone but a polyhedral set. Let Y_1 be given by a matrix $D \in \mathbb{R}^{m\times k_1}$ and a vector $a \in \mathbb{R}^m$ with some $m \in \mathbb{N}$, i.e., $Y_1 = \{y_1 \in \mathbb{R}^{k_1} : Dy_1 \ge a\}$. The condition $y_1 \in Y_1$ has to be integrated into the condition $Ty - b \in K_F$ in part b) of the proof. Therefore, the following definitions have to replace the earlier ones:

$$F = L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m+T}),\ \Omega_E = \left\{y \in E : y_1 \in \mathbb{R}^{k_1},\, y_t \in Y_t \text{ a.s. } (t = 2, ..., T)\right\},\ \Omega_F = \mathbb{R}_+^m \times \{0\}^T$$

and the matrix A and and the vector b have to be replaced by

$$\tilde{A} := \begin{pmatrix} D & 0\cdots 0 \\ & A \end{pmatrix} \in \mathbb{R}^{(m+T)\times K}, \quad \text{and} \quad \tilde{b} := \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{with} \quad b = (z_1, \ldots, z_T)',$$

respectively. This leads to m additional (scalar, non-random) dual variables $v_1, ..., v_m$ and the dual program reads

$$\sup\left\{ \langle a, v\rangle - \mathbb{E}\left[\sum_{t=1}^T \lambda_t z_t\right] \;\middle|\; \begin{array}{l} v \in \mathbb{R}_+^m,\ \ \lambda_t \in L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P})\ (t = 1, \ldots, T), \\ -\left(c_1 + \sum_{\nu=1}^T \mathbb{E}\left[\lambda_\nu\right] w_{\nu,\nu-1}\right) + D'v = 0, \\ -\left(c_t + \sum_{\nu=t}^T \mathbb{E}\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t}\right) \in Y_t^*\ (t = 2, ..., T) \end{array} \right\}.$$

Indeed, this is equivalent to (3.2) since LP duality lead to

$$\inf_{y_1\in Y_1} \left\langle c_1 + \textstyle\sum_{\nu=1}^T \mathbb{E}\left[\lambda_\nu\right] w_{\nu,\nu-1}, y_1 \right\rangle = \min\left\{\langle h, y_1\rangle : Dy_1 \le a\right\} = \max\left\{\langle a, v\rangle : D'v = h,\ v \in \mathbb{R}_+^m\right\},$$

where $h := c_1 + \sum_{\nu=1}^T \mathbb{E}\left[\lambda_\nu\right] w_{\nu,\nu-1}$. This completes the proof. □

Corollary 3.10. Let ρ be a functional on $\times_{t=1}^T L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ of the form (3.1) with Y_1 being a polyhedral cone. Let the conditions of Theorem 3.9 be satisfied (complete recourse, dual feasibility) and assume

$$\text{(3.5)} \qquad \Lambda_\rho := \left\{\lambda \in \times_{t=1}^T L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \,\middle|\, -\left(c_t + \textstyle\sum_{\nu=t}^T \mathbb{E}\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t}\right) \in Y_t^*\ (t = 1, ..., T)\right\} \subseteq \mathcal{D}_T\,.$$

Then ρ is a multiperiod coherent risk measure.

Proof. As for Corollary 2.5, only the continuity condition from Theorem 3.6 needs to be verified here, since $\mathcal{P}_\rho := \Lambda_\rho$ does the job. This continuity condition, however, is again a consequence of the Lipschitz continuity of ρ and the monotone convergence theorem. □

Example 3.11. A straightforward approach to incorporate risk in terms of the Conditional-Value-at-Risk at all time stages consists in considering a weighted sum

$$\rho_1(z) := \sum_{t=2}^T \gamma_t CVaR_{\alpha_t}(z_t)$$

with some weights $\gamma_t \geq 0$ (e.g. $\gamma_t = \frac{1}{T-1}$) and some confidence levels $\alpha_2, \alpha_3, ..., \alpha_T \in (0,1)$. Note that

$$\begin{aligned} \rho_1(z) &= \sum_{t=2}^T \gamma_t \inf_{r_t \in \mathbb{R}} \left\{ r_t + \tfrac{1}{\alpha_t}\mathbb{E}\left[(z_t + r_t)^-\right]\right\} \\ &= \inf_{(r_2,...,r_T)\in\mathbb{R}^{T-1}} \left\{ \sum_{t=2}^T \gamma_t \left(r_t + \tfrac{1}{\alpha_t}\mathbb{E}\left[(z_t + r_t)^-\right]\right)\right\} \\ &= \inf \left\{ \sum_{t=2}^T \gamma_t \left(r_t + \frac{1}{\alpha_t}\mathbb{E}\left[y_t^{(2)}\right]\right) \;\middle|\; \begin{array}{l} y_t^{(1)} - y_t^{(2)} = z_t + r_t \ (t = 2, ..., T), \\ y_t^{(1)}, y_t^{(2)} \geq 0 \ \mathcal{F}_t\text{-measurable } (t = 2, ..., T), \\ (r_2, ..., r_T) \in \mathbb{R}^{T-1} \end{array} \right\} \end{aligned}$$

i.e., ρ_1 is of the form (3.1) with $k_1 = T$, $k_t = 2$ $(t = 2, ..., T)$, $c_1 = (0, \gamma_2, ..., \gamma_T)$, $c_t = (0, \frac{\gamma_t}{\alpha_t})$ $(t = 2, ..., T)$, $w_{1,0} = e_1$, $w_{t,0} = (1, -1)$ $(t = 2, ..., T)$, $w_{t,t-1} = -e_t$ $(t = 2, ..., T)$, $w_{t,\tau} = 0$ $(\tau = 1, ..., t-2,\ t = 3, ..., T)$, $Y_1 = \mathbb{R}^T$, $Y_t = \mathbb{R}_+ \times \mathbb{R}_+$ $(t = 2, ..., T)$ (with e_t denoting the t-th standard basis vector in $\mathbb{R}^T$).

Thus, the risk measure ρ defined in this way is *multiperiod polyhedral.* Due to Remark 3.5 it is *multiperiod coherent*, too, if $\sum_{t=2}^T \gamma_t = 1$. This can also be seen by means of Corollary 3.10. The set of feasible multipliers is given here by

$$\Lambda_{\rho_1} = \left\{ \lambda \in \times_{t=1}^T L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \;\middle|\; \begin{array}{l} \lambda_1 = 0, \\ 0 \leq \lambda_t \leq \frac{\gamma_t}{\alpha_t} \ (t = 2, ..., T), \\ \mathbb{E}\left[\lambda_t\right] = \gamma_t \end{array} \right\} \tag{3.6}$$

and, of course, $\Lambda_{\rho_1} \subseteq \mathcal{D}_T$. Moreover, the conditions from Theorem 3.9 are satisfied: Complete recourse and dual feasibility (take $u = (0, \gamma_2, ..., \gamma_T)'$).

Next we present more involved examples, which extend the Conditional-Value-at-Risk to the multiperiod situation. The characteristic thing about $CVaR$ is that, in the dual representation, the density functions resp. Lagrangian multipliers are bounded pointwise from above (cf. Example 2.9). This idea will be found somehow in all of the following examples.

Example 3.12. In this example, we define a multiperiod coherent risk measure where not every timestep contributes with a fixed weight. When looking at the dual representation (3.2) and at condition (3.5), it becomes obvious that each of the dual constraints $c_t + \sum_{\nu=t}^T \mathbb{E}\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t} \in -Y_t^*$ has to imply $\lambda_t \geq 0$ for $t = 1, \ldots, T$. A natural candidate for implying $\sum_{\nu=1}^T \mathbb{E}[\lambda_\nu] = 1$ is the corresponding constraint for $t = 1$, which reads $c_1 + \sum_{\nu=1}^T \mathbb{E}[\lambda_\nu] w_{\nu,\nu-1} \in -Y_1^*$.

Now, setting $k_t = 2$ $(t = 1, ..., T)$, $c_1 = (0, 1)$, $c_t = (0, \beta_t)$ with some $\beta_t > 0$ $(t = 2, ..., T)$ such that $\sum_{t=2}^T \beta_t \geq 1$, $w_{t,0} = (1, -1)$ $(t = 1, ..., T)$, $w_{t,t-1} = (0, -1)$ $(t = 2, ..., T)$ and $w_{t,\tau} = 0$ $(\tau = 1, ..., t-2,\ t = 3, ..., T)$, $Y_1 = \mathbb{R} \times \mathbb{R}$, $Y_t = \mathbb{R}_+ \times \mathbb{R}_+$ $(t = 2, \ldots, T)$ leads to

$$\begin{array}{lcl} -\left(c_1 + \sum_{\nu=1}^T \mathbb{E}\left[\lambda_\nu\right] w_{\nu,\nu-1}\right) \in Y_1^* & \iff & \lambda_1 = 0 \quad \text{and} \quad \sum_{\nu=1}^T \mathbb{E}\left[\lambda_\nu\right] = 1, \\ -\left(c_t + \sum_{\nu=t}^T \mathbb{E}\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t}\right) \in Y_t^* & \iff & 0 \leq \lambda_t \quad \text{and} \quad \lambda_t \leq \beta_t \ \ (t = 2, ..., T) \end{array}$$

since $Y_1^* = \{0\} \times \{0\}$ and $Y_t^* = \mathbb{R}_- \times \mathbb{R}_-$ $(t = 2, \ldots, T)$. Hence, the dual set Λ_{ρ_2} is of the form

$$\Lambda_{\rho_2} = \left\{ \lambda \in \times_{t=1}^T L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \;\middle|\; \begin{array}{l} \lambda_1 = 0, \\ 0 \leq \lambda_t \leq \beta_t \ \ (t = 2, ..., T), \\ \sum_{t=1}^T \mathbb{E}[\lambda_t] = 1 \end{array} \right\}. \tag{3.7}$$

Note that complete recourse and dual feasibility hold. Thus, Corollary 3.10 implies that the functional

$$\rho_2(z) := \inf \left\{ y_1^{(2)} + \sum_{t=2}^T \beta_t \mathbb{E}\left[y_t^{(2)}\right] \;\middle|\; \begin{array}{l} y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^2) \ (t = 1, ..., T), \\ y_1 \in \mathbb{R} \times \mathbb{R},\ y_t \in \mathbb{R}_+ \times \mathbb{R}_+ \ (t = 2, ..., T), \\ z_1 = y_1^{(1)}, \\ z_t + y_1^{(2)} = y_t^{(1)} - y_t^{(2)} \ (t = 2, ..., T) \end{array} \right\}$$

or simply $\rho_2(z) = \inf_{r \in \mathbb{R}} \left\{ r + \sum_{t=2}^T \beta_t \mathbb{E}\left[(z_t + r)^-\right]\right\}$ is a *multiperiod polyhedral* and *coherent risk measure* for $1 < p < \infty$. Variants of this risk measure result from other choices of $w_{t,\tau}$, $\tau = 2, \ldots, t-2$, $t = 2, \ldots, T$.

The remaining examples present multiperiod polyhedral coherent risk measures that depend on the filtration $\{\mathcal{F}_t\}_{t=1}^T$, i.e., on the information flow over time.

Example 3.13. To incorporate the information structure we adapt the previous example in such a manner that successive timesteps are associated. Hence, we choose everything as before, only the assignment $w_{t,\tau} = 0$ $(\tau = 1, ..., t-2,\ t = 3, ..., T)$ has to be replaced by $w_{t,1} = (0, -1)$ $(t = 2, ..., T)$, $w_{t,\tau} = 0$ $(\tau = 2, \ldots, t-2,\ t = 4, ..., T)$. In addition, we set $c_t = (0, \delta_t)$ with $\delta_t > 0$ for $t = 2, ..., T$. Hence, the dual set Λ_{ρ_3} is of the form

$$\Lambda_{\rho_3} = \left\{ \lambda \in \times_{t=1}^T L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \;\middle|\; \begin{array}{l} \lambda_1 = 0\,,\ \sum_{t=1}^T \mathbb{E}[\lambda_t] = 1, \\ 0 \leq \lambda_t\,,\ \lambda_t + \mathbb{E}[\lambda_{t+1}|\mathcal{F}_t] \leq \delta_t \ \ (t = 2, ..., T-1), \\ 0 \leq \lambda_T \leq \delta_T. \end{array} \right\}. \tag{3.8}$$

Again, the complete recourse condition is satisfied and dual feasibility holds if the parameters δ_t are chosen sufficiently large. Altogether, Corollary 3.10 implies that the functional

$$\rho_3(z) := \inf \left\{ y_1^{(2)} + \sum_{t=2}^T \delta_t \mathbb{E}\left[y_t^{(2)}\right] \;\middle|\; \begin{array}{l} y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^2) \ (t = 1, ..., T) \\ y_1 \in \mathbb{R} \times \mathbb{R},\ y_t \in \mathbb{R}_+ \times \mathbb{R}_+ \ (t = 2, ..., T), \\ z_1 = y_1^{(1)}, \\ z_2 + y_1^{(2)} = y_2^{(1)} - y_2^{(2)}, \\ z_t + y_1^{(2)} + y_{t-1}^{(2)} = y_t^{(1)} - y_t^{(2)} \ (t = 3, ..., T) \end{array} \right\}$$

is a *multiperiod polyhedral* and *coherent risk measure.*

Example 3.14. In this approach, the concatenation of the timesteps is even stronger than in the previous example. We set $k_t = 2$ $(t = 1, ..., T)$, $c_1 = \left(0, \frac{1}{T-1}\right)$, $c_t = (\beta_t, 0)$ $(t = 2, ..., T)$ with some numbers $\frac{1}{T-1} < \beta_2 \leq \beta_3 \leq ... \leq \beta_T$, $w_{1,0} = (-1, 0)$, $w_{t,0} = (-1, 1)$ $(t = 2, ..., T)$, $w_{t,1} = (0, -1)$ $(t = 2, ..., T)$, $w_{t,\tau} = 0$ for $\tau > 1$, $Y_1 = \mathbb{R} \times \mathbb{R}$, $Y_t = \mathbb{R}_+ \times \mathbb{R}$ $(t = 2, ..., T-1)$, $Y_T = \mathbb{R}_+ \times \mathbb{R}_+$.

The dual constraints $-\left(c_t + \sum_{\nu=t}^T \mathbb{E}\left[\lambda_\nu|\mathcal{F}_t\right] w_{\nu,\nu-t}\right) \in Y_t^*$ imply that λ has to be a *martingale* with respect to the filtration $(\mathcal{F}_t)_{t=1}^T$. This implies $\mathbb{E}\left[\lambda_2\right] = ... = \mathbb{E}\left[\lambda_T\right]$ and $\lambda_t \geq 0$ since $\lambda_T \geq 0$. Together with (3.5) we obtain:

$$\Lambda_{\rho_4} = \left\{ \lambda \in \times_{t=1}^T L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \;\middle|\; \begin{array}{l} \lambda_1 = 0, \\ 0 \leq \lambda_t \leq \beta_t \ (t = 2, ..., T), \\ \lambda_t = \mathbb{E}\left[\lambda_{t+1}|\mathcal{F}_t\right] \ (t = 2, ..., T-1), \\ \mathbb{E}\left[\lambda_2\right] = ... = \mathbb{E}\left[\lambda_T\right] = \frac{1}{T-1} \end{array} \right\}. \tag{3.9}$$

Complete recourse is satisfied and dual feasibility holds since the vector $u \in \mathbb{R}^T$ with $u_1 = 0$ and $u_t = \frac{1}{T-1}$ for $t = 2, ..., T$ defines a (constant) element of Λ_{ρ_4}. Hence, Corollary 3.10 applies and the resulting functional

$$\rho_4(z) := \inf \left\{ \frac{1}{T-1} y_1^{(2)} + \sum_{t=2}^T \beta_t \mathbb{E}\left[y_t^{(1)}\right] \;\middle|\; \begin{array}{l} y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^2) \ (t = 1, ..., T), \\ y_1 \in \mathbb{R} \times \mathbb{R},\ y_T \in \mathbb{R}_+ \times \mathbb{R}_+, \\ y_t \in \mathbb{R}_+ \times \mathbb{R} \ (t = 2, ..., T-1), \\ z_1 = -y_1^{(1)}, \\ z_t = -y_t^{(1)} + y_t^{(2)} - y_{t-1}^{(2)} \ (t = 2, ..., T) \end{array} \right\}$$

is a *multiperiod coherent risk measure* (and, due to its definition, multiperiod polyhedral).

Comparing (3.9) for $\beta_t = \frac{1}{(T-1)\alpha}$ with the dual representation (2.7) of the Conditional-Value-at-Risk it turns out that the multiperiod risk measure ρ_4 define in this way is a kind of canonical extension of the Conditional-Value-at-Risk in terms of [2, Sections 4 and 5] and of [22][4].

The next example is motivated from the viewpoint of the value of information (cf. [18, 19]):

[4]The framework in these papers assumes that the multiperiod risk measure is determined only by a set of (scalar) density functions $\mathcal{P}_\rho \subseteq L_1(\Omega, \mathcal{F}, \mathbb{P})$ rather than $\subseteq \times_{t=1}^T L_1(\Omega, \mathcal{F}_t, \mathbb{P})$. Then, the risk $\rho(z)$ is given by expressions like $\sup\{-\frac{1}{T}\sum_{t=1}^T \mathbb{E}\left[f z_t\right] : f \in \mathcal{P}_\rho\}$ or $\sup\{-\mathbb{E}\left[f z_\tau\right] : f \in \mathcal{P}_\rho,\ \tau$ stopping time$\}$. Indeed, Λ_{ρ_4} is nothing else but the set of densities for the Conditional-Value-at-Risk (2.7), i.e., all density functions bounded by $\frac{1}{\alpha}$.

Example 3.15. In [19], the following multiperiod risk measure was suggested. Given some constants $0 \le d \le b_{T-1} \le ... \le b_2 \le b_1$ and $b_{t-1} \le q_t$ for $t = 2, ..., T$ this risk measure is defined[5] on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ by

$$\rho_5(Z) = -\sup \left\{ \begin{array}{l} \mathbb{E}\left[b_1 A_1 + \sum_{t=2}^{T-1} (b_t A_t - q_t M_t) + d K_T - q_T M_T \right] \Bigg| \\ A_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}) \quad (t = 1, ..., T), \\ K_t = [K_{t-1} + Z_t - A_{t-1}]^+ \quad (t = 2, ..., T), \\ M_t = [K_{t-1} + Z_t - A_{t-1}]^- \quad (t = 2, ..., T) \end{array} \right\}$$

with $K_1 := 0$. However, in [19] $Z = (Z_1, ..., Z_T)$ is understood as income process with $Z_1 = 0$, thus this definition does not fit in our framework.

Therefore, we rewrite this definition taking the value processes $z = (z_1, ..., z_T)$ with $z_1 = Z_1 = 0$, $z_t = \sum_{\tau=1}^{T} Z_\tau$, i.e., $Z_t = z_t - z_{t-1}$ for $t > 2$. This reformulation leads to the representation (3.1) with $k_t = 3$ $(t = 1, ..., T)$, $Y_1 = \mathbb{R} \times \mathbb{R} \times \{0\}$, $Y_t = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$ $(t = 2, ..., T)$, $y_t = (A_t, M_t, K_t)$, $w_{t,0} = (0, -1, 1)$ $(t = 1, ..., T)$, $w_{t,\tau} = (1, -1, 0)$ $(\tau = 1, ..., t-2,\ t = 3, ..., T)$, $w_{t,t-1} = (1, 0, 0)$ $(t = 2, ..., T)$, $c_1 = (-b_1, 0, 0)$, $c_t = (-b_t, q_t, 0)$ $(t = 2, ..., T-1)$, $c_T = (0, q_T, -d)$.

To understand this reformulation note that $w_{1,0} = (0, -1, 1)$ implies $M_1 = -z_1 = 0$ and that for $t = 2, ..., T$ the recursion $K_t - M_t = K_{t-1} + Z_t - A_{t-1}$ with $K_t \ge 0$ and $M_t \ge 0$ must hold. This recursion can be transformed into a recursion of the type of the definition of multiperiod polyhedrality:

$$z_t = K_t + \textstyle\sum_{\tau=1}^{t-1} A_\tau - \sum_{\tau=2}^{t} M_\tau \quad (t = 2, ..., T)$$

with $K_1 = 0$. Thus, this risk measure fits into the framework of *multiperiod polyhedral* risk measures.

Furthermore, it is *multiperiod coherent* if $b_1 = 1$. This can be shown by Corollary 3.10. Note that

$$-\left(c_1 + \textstyle\sum_{\nu=1}^{T} \mathbb{E}[\lambda_\nu]\, w_{\nu,\nu-1}\right) \in Y_1^* \iff \textstyle\sum_{\nu=2}^{T} \mathbb{E}[\lambda_\nu] = b_1 \text{ and } \lambda_1 = 0$$

and

$$\begin{array}{l} -\left(c_t + \sum_{\nu=t}^{T} \mathbb{E}[\lambda_\nu | \mathcal{F}_t]\, w_{\nu,\nu-t}\right) \in Y_t^* \ (t = 2, ..., T) \iff \\ d \le \lambda_T \le q_T,\ 0 \le \lambda_t \le q_t - b_t,\ \sum_{\nu=t+1}^{T} \mathbb{E}[\lambda_\nu | \mathcal{F}_t] = b_t \ (t = 2, ..., T-1), \end{array}$$

thus

$$\Lambda_{\rho_5} = \left\{ \lambda \in \times_{t=1}^{T} L_{p'}(\Omega, \mathcal{F}_t, P) \;\middle|\; \begin{array}{l} \lambda_1 = 0, \\ 0 \le \lambda_t \le q_t - b_t \ (t = 2, ..., T-1), \\ d \le \lambda_T \le q_T, \\ \mathbb{E}[\lambda_T | \mathcal{F}_{T-1}] = b_{T-1}, \\ \mathbb{E}[\lambda_t | \mathcal{F}_{t-1}] = b_{t-1} - b_t,\ (t = 2, ..., T-1) \end{array} \right\}.$$

Further, complete recourse is obviously satisfied and dual feasibility holds since the vector $u \in \mathbb{R}^T$ with $u_1 = 0$ and $u_T = b_{T-1}$ and $u_t = b_{t-1} - b_t$ for $t = 2, ..., T-1$ defines a (constant) element of Λ_{ρ_5}. Furthermore, $\sum_{t=1}^{T} \mathbb{E}[\lambda_t] = b_1$ for $\lambda \in \Lambda_{\rho_5}$, thus the inclusion $\Lambda_{\rho_5} \subseteq \mathcal{D}_T$ holds indeed if $b_1 = 1$.

An interesting specific case appears for $d = 0$, $q_t = b_{t-1}$ $(t = 1, \ldots, T)$ and with $\alpha_t \in (0, 1)$, where $\frac{1}{(T-1)\alpha_t} = b_{t-1} - b_t$ $(t = 2, \ldots, T)$ and $0 = b_T < b_{T-1} < \ldots < b_2 < b_1 = 1$. Then we obtain

$$\Lambda_{\rho_5} = \left\{ \lambda \in \times_{t=1}^{T} L_{p'}(\Omega, \mathcal{F}_t, P) \;\middle|\; \begin{array}{l} \lambda_1 = 0, \\ 0 \le \lambda_t \le \frac{1}{(T-1)\alpha_t} \ (t = 2, ..., T), \\ \mathbb{E}[\lambda_t | \mathcal{F}_{t-1}] = \frac{1}{(T-1)\alpha_t} \ (t = 2, ..., T) \end{array} \right\}$$

and the risk measure ρ_5 on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, P)$ takes the form

$$\rho_5(z) = \mathbb{E}\left[\tfrac{1}{T-1} \textstyle\sum_{t=2}^{T} \inf \left\{ u_{t-1} + \tfrac{1}{\alpha_t}(z_t + u_{t-1})^- \;\middle|\; u_{t-1} \in L_p(\Omega, \mathcal{F}_{t-1}, P) \right\} \right]. \tag{3.10}$$

[5]In [19], ρ_5 is called a (negative) utility measure rather than a risk measure. Moreover, the first time stage (i.e., the deterministic stage) is denoted by index 0 there. Here, the formulation is adapted to our framework with index 1 for the deterministic time stage (i.e., $\mathcal{F}_1 = \{\emptyset, \Omega\}$). In addition, the notations c_t and a_t were replaced by the definitions $b_t := c_{t+1}$ and $A_t := a_{t+1}$.

The t-th summand can be interpreted as the Conditional-Value-at-Risk of z_t conditioned with respect to the σ-field $\mathcal{F}_{t-1}$. Clearly, (3.10) boils down to the one-period $CVaR$ (2.6) for $T = 2$.

Remark 3.16. Of course, it is interesting to compare these examples. To this end, it is useful to consider the dual representations, i.e., the Lagrange multiplier sets Λ_{ρ_j} $(j = 1, ..., 4)$. Hence, regarding formulas (3.7), (3.8), and (3.9), it is obvious that for $\delta_t = \beta_t$ it holds that $\Lambda_{\rho_4} \subseteq \Lambda_{\rho_2} \supseteq \Lambda_{\rho_3}$, thus, since

$$\rho_j(z) = \sup\left\{-\sum_{t=1}^{T}\mathbb{E}\left[\lambda_t z_t\right] : \lambda \in \Lambda_{\rho_j}\right\}, \tag{3.11}$$

the relation $\rho_4 \leq \rho_2 \geq \rho_3$ is valid. On the other hand, comparing ρ_3 and ρ_4 for the case $\delta_t = 2\beta_t$ leads to $\Lambda_{\rho_4} \subseteq \Lambda_{\rho_3}$, thus $\rho_4 \leq \rho_3$. Hence, ρ_3 is more cautious than ρ_4 in this case. Moreover, if we set $\gamma_t = \frac{1}{T-1}$ and $\beta_t = \frac{1}{(T-1)\alpha_t}$, formula (3.6) shows $\Lambda_{\rho_4} \subseteq \Lambda_{\rho_1} \subseteq \Lambda_{\rho_2}$, hence, $\rho_4 \leq \rho_1 \leq \rho_2$. Thus, ρ_2 is the most cautious or most pessimistic of these risk measures.

More precisely: For a fixed random variable z let $\lambda^j = \lambda^j(z) \in \Lambda_{\rho_j}$ be a maximizer for the dual representations (3.11) of ρ_j, respectively. Then, roughly speaking, λ^j is big where z is small in compliance with the respective restrictions. For $j = 1$ and $j = 4$, the weighting of the time steps is fixed in advance since $\mathbb{E}[\lambda_t^j]$ is fixed. For $j = 2$ the weighting of the time steps is variable, hence, the available probability mass of λ^2 is concentrated at time steps at which z is low. Thus, ρ_2 is a kind of worst time step risk measure. This might be desirable or not, depending on the application.

Comparing ρ_1 with ρ_4, one sees that in the first case λ_t^1 is big where z_t is small, independent of the other time steps. In the second case, λ^4 is completely determined by λ_T^4 since $\lambda_t^4 = \mathbb{E}\left[\lambda_T^4|\mathcal{F}_t\right]$ because of the martingale property. This means that the maximization (3.11) takes all time steps into account simultaneously, i.e., the maximization occurs along the paths of the treelike information structure given by the filtration $(\mathcal{F}_t)_{t=1}^T$. This latter approach seems to be more efficient in case the risk of paths is of interest. Then, ρ_1 may be more pessimistic than necessary. Furthermore, it does not incorporate the information structure of the problem. On the other hand, the martingale property of ρ_4 seems very restrictive.

Comparing ρ_3 and ρ_4 for the case $\delta_t = 2\beta_t$ leads to $\Lambda_{\rho_4} \subseteq \Lambda_{\rho_3}$, thus $\rho_4 \leq \rho_3$. Hence, ρ_3 is more cautious than ρ_4 in this case. Regarding the dual sets for ρ_5, one obtains $\Lambda_{\rho_5} \subseteq \Lambda_{\rho_1}$ for $\gamma_t = b_{t-1} - b_t$ and $\alpha_t = (b_{t-1} - b_t)/(q_t - b_t)$, and $\Lambda_{\rho_5} \subseteq \Lambda_{\rho_3}$ for $\delta_t = q_t - b_{t+1}$. Hence, $\rho_1 \geq \rho_5 \leq \rho_3$, i.e., ρ_5 is less cautious for this choice of the coefficients.

However, cautiousness is not necessarily a desirable property, because in applications one usually has to pay a price for being cautious. Which risk measure to take depends highly on the intention of the application. It seems that ρ_3 may be a good compromise, since the information structure is taken into account and there is no fixed weighting of the time steps.

4. Risk Measures in Stochastic Programs

In this section we study the effect of replacing expectation-based objectives of stochastic programming problems by polyhedral risk measures. In particular, we are interested in consequences for structural and stability properties of the resulting models. We assume that randomness occurs as a (possibly multivariate) stochastic data process $(\xi_t)_{t=1}^T$ and set $\mathcal{F}_t = \sigma(\xi_1, ..., \xi_t)$, $t = 1, \ldots, T$. We consider multistage stochastic programs of the form

$$\min\left\{\mathbb{E}\left[\sum_{t=1}^{T}\langle b_t(\xi_t), x_t\rangle\right] \;\middle|\; \begin{array}{ll} x_t \in X_t, & \\ H_t(x_t) = 0, & \\ B_t(\xi_t)x_t \leq d_t(\xi_t), & (t = 1, ..., T) \\ \sum_{\tau=0}^{t-1} A_{t,\tau}(\xi_t)x_{t-\tau} = h(\xi_t) & \end{array}\right\} \tag{4.1}$$

with closed sets X_t having the property that their convex hull is polyhedral, cost coefficients $b_t(\cdot)$, right-hand sides $d_t(\cdot)$ and $h_t(\cdot)$, and matrices $A_{t,\tau}(\cdot)$, $\tau = 0, \ldots, t-1$, and $B_t(\cdot)$ all having suitable dimensions and possibly depending affine linearly on ξ_t for $t = 1, \ldots, T$. The constraints consist of four groups, where the first $x_t \in X_t$ models simple fixed constraints, the second $H_t(z) := z - \mathbb{E}\left[z|\mathcal{F}_t\right] = 0$ ensures the non-anticipativity of the decisions x_t, and the third and fourth are the coupling and the dynamic constraints, respectively. By $\mathcal{X}(\xi)$ we denote the set of decisions satisfying all constraints of (4.1).

When replacing the expectation of the stochastic costs $\sum_{t=1}^{T}\langle b_t(\xi_t), x_t\rangle$ by some polyhedral multi-period risk measure ρ applied to the random vector $(-\sum_{\tau=1}^{t}\langle b_\tau(\xi_\tau), x_\tau\rangle)_{t=1}^{T}$ of negative intermediate costs, we arrive at the following risk averse alternative to problem (4.1):

$$\min\left\{\rho\left(-\langle b_1(\xi_1), x_1\rangle, -\langle b_1(\xi_1), x_1\rangle - \langle b_2(\xi_2), x_2\rangle, \ldots, -\textstyle\sum_{\tau=1}^{T}\langle b_\tau(\xi_\tau), x_\tau\rangle\right)\,\middle|\, x \in \mathcal{X}(\xi)\right\}. \tag{4.2}$$

The polyhedral risk measure ρ is defined by the minimization problem

$$\rho(z) = \inf\left\{\mathbb{E}\left[\textstyle\sum_{t=1}^{T}\langle c_t, y_t\rangle\right] \,\middle|\, \begin{array}{l} H_t(y_t) = 0,\ y_t \in Y_t, \\ \sum_{\tau=0}^{t-1}\langle w_{t,\tau}, y_{t-\tau}\rangle = z_t \end{array} \quad (t = 1, \ldots, T)\right\}.$$

This gives rise to the question whether (4.2) is equivalent to the optimization model

$$\min\left\{\mathbb{E}\left[\sum_{t=1}^{T}\langle c_t, y_t\rangle\right] \,\middle|\, \begin{array}{l} x \in \mathcal{X}(\xi), \\ H_t(y_t) = 0,\ y_t \in Y_t\ (t = 1, \ldots, T), \\ \sum_{\tau=0}^{t-1}\langle w_{t,\tau}, y_{t-\tau}\rangle + \sum_{\tau=1}^{t}\langle b_\tau(\xi_\tau), x_\tau\rangle = 0\ (t = 1, \ldots, T) \end{array}\right\}, \tag{4.3}$$

where the minimization with respect to the original decision x and the variable y defining ρ is carried out simultaneously. Of course, the answer is positive.

Proposition 4.1. *Minimizing (4.2) with respect to x is equivalent to minimizing (4.3) with respect to all pairs (x, y) in the following sense: The optimal values of (4.2) and (4.3) coincide and a pair (x^*, y^*) is a solution of (4.3) iff x^* solves (4.2) and y^* is a solution of the minimization problem defining $\rho\left(\left(-\sum_{\tau=1}^{t}\langle b_\tau(\xi_\tau), x_\tau^*\rangle\right)_{t=1}^{T}\right)$.*

Proof. The minimization with respect to all feasible pairs (x, y) of (4.3) can be carried out by minimizing with respect to y, thus arriving at $\rho\left(\left(-\sum_{\tau=1}^{t}\langle b_\tau(\xi_\tau), x_\tau\rangle\right)_{t=1}^{T}\right)$, and then by minimizing the latter residual with respect to $x \in \mathcal{X}(\xi)$. Hence, the optimal values coincide and, if the pair (x^*, y^*) solves (4.3), its first component x^* is a solution of (4.2) and y^* is a solution of the problem

$$\min\left\{\mathbb{E}\left[\textstyle\sum_{t=1}^{T}\langle c_t, y_t\rangle\right] \,\middle|\, \begin{array}{l} H_t(y_t) = 0,\ y_t \in Y_t, \\ \sum_{\tau=0}^{t-1}\langle w_{t,\tau}, y_{t-\tau}\rangle + \sum_{\tau=1}^{t}\langle b_\tau(\xi_\tau), x_\tau^*\rangle = 0 \end{array} \quad (t = 1, \ldots, T)\right\}, \tag{4.4}$$

whose optimal value is just $\rho\left(\left(-\sum_{\tau=1}^{t}\langle b_\tau(\xi_\tau), x_\tau^*\rangle\right)_{t=1}^{T}\right)$. Conversely, if x^* is a solution of (4.2) and y^* a solution of (4.4), the pair (x^*, y^*) has to be a solution of (4.3). □

Thus, minimizing a stochastic program with a polyhedral risk measure in the objective leads to a "traditional" stochastic program with linear expectation-based objective and with additional variables y and constraints, respectively. Both, the variables and the constraints are convenient for stochastic programs since the variables are nicely constrained by polyhedral sets (no integer requirements). Thus, if the original expectation-based stochastic program (4.1) has convenient properties, there is good reason to expect that these properties are maintained when using a polyhedral risk measure for risk aversion.

4.1. **Stability of Stochastic Programs.** Stability of solutions and optimal values of stochastic programs with respect to the perturbation of the underlying probability measure is an important issue since in applications the true measure $\mathbb{P}$ is usually unknown and has to be approximated by some other measure $\mathbb{Q}$. Such an approximation may be gained by sampling techniques.

In [28] various stability results involving distances $d(\mathbb{P}, \mathbb{Q})$ of probability measures are developed for different types of (mainly) expectation-based stochastic programs. It is shown there that certain ideal probability metrics (see [20] for an exposition) may be associated with classes of stochastic programs. Here, we briefly show that these stability results remain valid for important classes if the expectation is replaced by a polyhedral risk measure. We restrict ourselves to the two-stage case here since stability properties are best understood for such programs. In the context of distances of probability measures it turns out to be useful to assume that $\Omega = \Xi \subseteq \mathbb{R}^n$ and $\mathcal{F} = \mathcal{B}(\Xi)$.

4.1.1. *Linear two-stage programs.* In [21, Theorem 3.3] and [28] it is shown that two-stage stochastic programs with fixed recourse of the form

$$\min\left\{ -\mathbb{E}_{\mathbb{P}}\left[\langle b, x_1\rangle + \langle p(\cdot), x_2(\cdot)\rangle\right] \,\middle|\, \begin{array}{l} Wx_2(\xi) = h(\xi) - T(\xi)x_1, \\ x_1 \in X_1,\, x_2(\xi) \in X_2 \end{array} \right\}, \tag{4.5}$$

with X_1 and Ξ being polyhedral sets, X_2 a polyhedral cone and $p(\cdot)$, $h(\cdot)$, $T(\cdot)$ being affine linear functions (of $\xi \in \Xi$), are known to be stable[6] with respect to the probability metric ζ_2 given by

$$\zeta_2(\mathbb{P},\mathbb{Q}) = \sup\left\{ |\mathbb{E}_{\mathbb{P}}[F] - \mathbb{E}_{\mathbb{Q}}[F]| : F \in \mathbb{R}^{\Xi}, |F(\xi) - F(\xi')| \le \max\{1, \|\xi\|, \|\xi'\|\}\, \|\xi - \xi'\|, \forall \xi, \xi' \in \Xi \right\}$$

if the following three conditions hold:
(i) $\forall\,(x_1,\xi) \in X_1 \times \Xi\ \exists\, x_2 \in X_2 : Wx_2 = h(\xi) - T(\xi)x_1$ (relatively complete recourse),
(ii) $\forall\,\xi \in \Xi\ \exists\, z : W'z + p(\xi) \in X_2^*$ (dual feasibility),
(iii) $\mathbb{E}_{\mathbb{P}}\left[\|\xi\|^2\right] < \infty$ (finite second moments).
If we exchange from (negative) expectation to a (one-period) polyhedral risk measure $\rho = \rho_{\mathbb{P}}$ according to (2.1), we obtain the problem

$$\min\left\{ \rho\left[\langle b, x_1\rangle + \langle p, x_2(.)\rangle\right] \,\middle|\, \begin{array}{l} Wx_2(\xi) = h(\xi) - T(\xi)x_1, \\ x_1 \in X_1,\, x_2(\xi) \in X_2 \end{array} \right\},$$

which is equivalent to

$$\min\left\{ \langle c_1, y_1\rangle + \mathbb{E}_{\mathbb{P}}\left[\langle c_2, y_2(.)\rangle\right] \,\middle|\, \begin{array}{l} Wx_2(\xi) = h(\xi) - T(\xi)x_1, \\ \langle w_2, y_2(\xi)\rangle - \langle p(\xi), x_2(\xi)\rangle = \langle b, x_1\rangle - \langle w_1, y_1\rangle, \\ x_1 \in X_1,\, x_2(\xi) \in X_2, \\ y_1 \in Y_1,\, y_2(\xi) \in Y_2 \end{array} \right\}. \tag{4.6}$$

The latter program has almost the same structure as (4.5) with
$\hat{x}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\hat{x}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$, $\hat{X}_1 = X_1 \times Y_1$, $\hat{X}_2 = X_2 \times Y_2$,
$\hat{W}(\xi) = \begin{pmatrix} W & 0 \\ -p(\xi)' & w_2' \end{pmatrix}$, $\hat{T}(\xi) = \begin{pmatrix} T(\xi) & 0 \\ -b' & w_1' \end{pmatrix}$, $\hat{h}(\xi) = \begin{pmatrix} h(\xi) \\ 0 \end{pmatrix}$, $\hat{b} = \begin{pmatrix} 0 \\ -c_1 \end{pmatrix}$, $\hat{p} = \begin{pmatrix} 0 \\ -c_2 \end{pmatrix}$,
but now the recourse matrix $\hat{W}$ is random while the cost coefficient $\hat{p}$ is non-random.

Moreover, if we also impose complete recourse and dual feasibility for the polyhedral risk measure ρ in the sense of Section 2, i.e., (i) $\langle w_2, Y_2\rangle = \mathbb{R}$ and (ii) $\emptyset \neq \{u \in \mathbb{R} : uw_2 - c_2 \in Y_2^*\} \subseteq \mathbb{R}_-$, then we can conclude both, relatively complete recourse and dual feasibility, for the risk aversive alternative (4.6):
(i) Relatively complete recourse:
Let $(x_1, y_1, \xi) \in X_1 \times Y_1 \times \Xi$, then $\exists\, x_2 \in X_2 : Wx_2 = h(\xi) - T(\xi)x_1$ and $y_2 \in Y_2$ can be chosen such that $\langle w_2, y_2\rangle - \langle p(\xi), x_2\rangle = \langle b, x_1\rangle - \langle w_1, y_1\rangle$, thus $\hat{W}(\xi)\hat{x}_2 = \hat{h}(\xi) - \hat{T}(\xi)\hat{x}_1$.
(ii) Dual feasibility:
Let $\xi \in \Xi$. Choose z such that $W'z + p(\xi) \in X_2^*$ and $v \in \{u \in \mathbb{R} : +uw_2 - c_2 \in Y_2^*\} \subseteq \mathbb{R}_-$, set $\hat{z} = (-vz, v)'$, then one obtains

$$\hat{W}(\xi)'\hat{z} + \hat{p} = \begin{pmatrix} -v(W'z + p(\xi)) \\ vw_2 - c_2 \end{pmatrix} \in X_2^* \times Y_2^* = \hat{X}_2^*,$$

by making use of the fact that $-v \in \mathbb{R}_+$ and that X_2 is a cone.

Since the randomness enters only the last row of $\hat{W}(\xi)$ except for the coefficient in the main diagonal, the stability results from [30] for the random recourse situation with only lower diagonal randomness apply. The model (4.6) with non-random costs, however, is again stable with respect to the same metric ζ_2 as for (4.5).

[6]We do not give a precise definition of stability here, see [28] for this. Briefly, stability means that optimal values and solution sets behave (quantitatively) continuous at the original measure $\mathbb{P}$ with respect to a distance $d(\mathbb{P},\mathbb{Q})$.

4.1.2. *Linear mixed-integer two-stage programs.* In [28, Theorem 35], it is shown that programs of the form

$$\min\left\{ -\mathbb{E}_{\mathbb{P}}\left[\langle b, x_1\rangle + \langle p, x_2(.)\rangle + \langle \bar{p}, \bar{x}_2(.)\rangle\right] \,\middle|\, \begin{array}{l} Wx_2(\xi) + \bar{W}\bar{x}_2(\xi) = h(\xi) - T(\xi)x_1, \\ x_1 \in X_1,\ x_2(\xi) \in X_2 \cap \mathbb{Z}^m,\ \bar{x}_2(\xi) \in \bar{X}_2 \end{array} \right\} \tag{4.7}$$

with a closed Euclidean set X_1, a polyhedral set Ξ, polyhedral cones X_2 and $\bar{X}_2$ are known to be stable with respect to the probability metric ζ_{1,ph_k} with some $k \in \mathbb{N}$ if the following four conditions are satisfied:
(i) $\forall\, (x_1, \xi) \in X_1 \times \Xi\ \exists\, x_2 \in X_2 \cap \mathbb{Z}^m,\ \bar{x}_2 \in \bar{X}_2 : Wx_2 + \bar{W}\bar{x}_2 = h(\xi) - T(\xi)x_1$ (rel. complete recourse),
(ii) $\exists\, z \in \mathbb{R}^r : W'z + p \in X_2^*$ and $\bar{W}'z + \bar{p} \in \bar{X}_2^*$ (dual feasibility),
(iii) $\mathbb{E}_{\mathbb{P}}\left[\|\xi\|\right] < \infty$ (finite first moments),
(iv) W and $\bar{W}$ have rational coefficients only (rational recourse).
The metric ζ_{1,ph_k} is given by

$$\zeta_{1,ph_k}(\mathbb{P},\mathbb{Q}) = \sup\left\{ \left|\mathbb{E}_{\mathbb{P}}\left[F\chi_B\right] - \mathbb{E}_{\mathbb{Q}}\left[F\chi_B\right]\right| : B \in \mathcal{B}_{ph_k}(\Xi), F \in \mathbb{R}^{\Xi}, |F(\xi) - F(\xi')| \le \|\xi - \xi'\|\, \forall \xi, \xi' \in \Xi \right\}$$

where $\mathcal{B}_{ph_k}(\Xi)$ is the set of polyhedra contained in Ξ with at most k faces and χ denotes the characteristic function, i.e., $\chi_B(\xi) = 1$ if $\xi \in B$ and $= 0$ otherwise.

If we exchange from (negative) expectation to a polyhedral risk measure $\rho = \rho_{\mathbb{P}}$ according to (2.1) we obtain the problem

$$\min\left\{ \langle c_1, y_1\rangle + \mathbb{E}_{\mathbb{P}}\left[\langle c_2, y_2(.)\rangle\right] \,\middle|\, \begin{array}{l} Wx_2(\xi) + \bar{W}\bar{x}_2(\xi) = h(\xi) - T(\xi)x_1, \\ \langle w_2, y_2(\xi)\rangle - \langle p, x_2(\xi)\rangle - \langle \bar{p}, \bar{x}_2(\xi)\rangle = \langle b, x_1\rangle - \langle w_1, y_1\rangle, \\ x_1 \in X_1,\ x_2(\xi) \in X_2 \cap \mathbb{Z}^m,\ \bar{x}_2(\xi) \in \bar{X}_2, \\ y_1 \in Y_1,\ y_2(\xi) \in Y_2 \end{array} \right\}.$$

The latter program has the same structure as (4.7) with
$\hat{x}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\hat{x}_2 = x_2$, $\hat{\bar{x}}_2 = \begin{pmatrix} \bar{x}_2 \\ y_2 \end{pmatrix}$, $\hat{X}_1 = X_1 \times Y_1$, $\hat{X}_2 = X_2$, $\hat{\bar{X}}_2 = \bar{X}_2 \times Y_2$,
$\hat{W} = \begin{pmatrix} W & 0 \\ -q' & w_2' \end{pmatrix}$, $\hat{\bar{W}} = \begin{pmatrix} \bar{W} \\ -\bar{q}' \end{pmatrix}$, $\hat{T}(\xi) = \begin{pmatrix} T(\xi) & 0 \\ -b' & w_1' \end{pmatrix}$, $\hat{h}(\xi) = \begin{pmatrix} h(\xi) \\ 0 \end{pmatrix}$,
$\hat{b} = \begin{pmatrix} 0 \\ -c_1 \end{pmatrix}$, $\hat{q} = \begin{pmatrix} 0 \\ -c_2 \end{pmatrix}$, $\hat{\bar{q}} = 0$.

As in the previous paragraph, this combined program here satisfies relatively complete recourse and dual feasibility if both, (4.7) and ρ, do so. To have all the conditions (i) to (iv) satisfied one has to impose additionally that also q, $\bar{q}$ and w_2 have only rational coefficients. Then, however, the same stability (with respect to the metric ζ_{1,ph_k}) as for the original program is guaranteed.

4.2. **Lagrangian Relaxation and Decomposition.** We consider again the multistage stochastic program (4.1) and its risk averse alternative (4.2), which, according to Proposition 4.1, is of the form

$$\min\left\{ \mathbb{E}\left[\sum_{t=1}^{T} \langle c_t, y_t\rangle\right] \,\middle|\, \begin{array}{ll} x_t \in X_t,\ y_t \in Y_t, & \\ H_t(x_t) = 0,\ H_t(y_t) = 0, & \\ B_t(\xi_t)x_t \le d_t(\xi_t), & (t = 1, ..., T) \\ \sum_{\tau=0}^{t-1} A_{t,\tau}(\xi_t)x_{t-\tau} = h(\xi_t), & \\ \sum_{\tau=0}^{t-1}(\langle w_{t,\tau}, y_{t-\tau}\rangle + \langle b_{\tau+1}(\xi_{\tau+1}), x_{\tau+1}\rangle) = 0 & \end{array} \right\}. \tag{4.8}$$

Obviously, (4.8) has the same structure as (4.1), but additionally with T vector valued random variables and T dynamic (equality) constraints. Thus, decomposition methods that work for (4.1) are likely to work for (4.8), too. We exemplify this here by two important dual decomposition methods.

4.2.1. *Scenario Decomposition.* When solving problems like (4.1) or (4.8) one usually has to approximate $\mathbb{P}$ or, equivalently, ξ by a finite number of scenarios (more precisely: by a finite scenario tree). This can be expressed by $\infty > \#\Omega =: S$ and one can assume without loss of generality $\Omega = \left\{\xi^1, ..., \xi^S\right\}$ and $\mathcal{F} = \wp(\Omega)$. Then the problem is no longer infinite-dimensional and can be solved by standard mixed-integer linear programming techniques, but it is very large scale in most cases. Thus, specialized decomposition techniques are of great interest (cf. [7, 31, 29]).

Scenario Decomposition means Lagrange-dualizing the non-anticipativity constraints of (4.8) and solving the dual scenario-wise. Setting $m_t := \dim x_t$ we obtain the dual problem

$$\max\left\{D(\lambda_1,\lambda_2) : \lambda_{1t} \in L_1(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^{m_t}),\ \lambda_{2t} \in L_1(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^{k_t})\right\}$$

with the dual function $D(\lambda_1,\lambda_2)$ given by

$$D(\lambda_1,\lambda_2) = \min\left\{ \mathbb{E}\left[\sum_{t=1}^{T}(\langle c_t, y_t\rangle + \langle\lambda_{1t}, H_t(x_t)\rangle + \langle\lambda_{2t}, H_t(y_t)\rangle)\right] \;\middle|\; \begin{array}{l} x_t \in X_t,\ y_t \in Y_t, \\ B_t(\xi_t)x_t \le d_t(\xi_t), \\ \sum_{\tau=0}^{t-1} A_{t,\tau}(\xi_t)x_{t-\tau} = h(\xi_t), \\ \sum_{\tau=0}^{t-1}(\langle w_{t,\tau}, y_{t-\tau}\rangle, \\ +\langle b_{\tau+1}(\xi_{\tau+1}), x_{\tau+1}\rangle) = 0 \end{array} \right\}.$$

Solving this problem is an iterative process: $D(\lambda_1,\lambda_2)$ has to be computed for a fixed pair (λ_1,λ_2) and then (λ_1,λ_2) has to be updated via subgradient-type methods and so on. If the sets X_t are non-convex, this procedure only leads to lower bounds of the optimal value of (4.1) and suitable globalization techniques based on these lower bounds have to be used in addition.

With $p_s := \mathbb{P}(\{\xi^s\})$, $x_t^s := x_t(\xi^s)$, $y_t^s := y_t(\xi^s)$ and $\lambda_{jt}^s = \lambda_{jt}^s(\xi^s)$ the dual function reads

$$D(\lambda_1,\lambda_2) = \sum_{s=1}^{S} p_s \min\left\{ \sum_{t=1}^{T}(\langle c_t, y_t^s\rangle + \langle H_t^s(\lambda_{1t}), x_t^s\rangle + \langle H_t^s(\lambda_{2t}), y_t^s\rangle) \;\middle|\; \begin{array}{l} x_t^s \in X_t,\ y_t^s \in Y_t, \\ B_t(\xi_t^s)x_t^s \le d_t(\xi_t^s), \\ \sum_{\tau=0}^{t-1} A_{t,\tau}(\xi_t^s)x_{t-\tau}^s = h(\xi_t^s), \\ \sum_{\tau=0}^{t-1}(\langle w_{t,\tau}, y_{t-\tau}^s\rangle + \\ \langle b_{\tau+1}(\xi_{\tau+1}^s), x_{\tau+1}^s\rangle) = 0 \end{array} \right\}$$

and, thus, decomposes across scenarios. To derive the above form of D, the identities $\mathbb{E}[\langle\lambda_{1t}, H_t(x_t)\rangle] = \mathbb{E}[\langle H_t(\lambda_{1t}), x_t\rangle]$ and $\mathbb{E}[\langle\lambda_{2t}, H_t(y_t)\rangle] = \mathbb{E}[\langle H_t(\lambda_{2t}), y_t\rangle]$ were used. Hence, instead of one problem with $S\cdot\sum_{t=1}^{T}(m_t+k_t)$ variables one only has to solve S subproblems each with $\sum_{t=1}^{T}(m_t+k_t)$ variables to update the multipliers. In comparison with the (dualized form of the) purely expectation based problem (4.1) one has T additional equality constraints and $\sum_{t=1}^{T} k_t$ additional variables in each subproblem. Note that the dimensions k_t of y_t are typically small compared to the dimensions m_t of x_t.

4.2.2. *Geographical Decomposition.* In many practical applications problem (4.1) shows the following kind of separability with respect to blocks $x_i = (x_{i1},\dots,x_{iT})$, $i = 1,\dots,I$, of components of x:

$$\min\left\{ \mathbb{E}\left[\sum_{i=1}^{I}\sum_{t=1}^{T}\langle b_{it}(\xi_t), x_{it}\rangle\right] \;\middle|\; \begin{array}{l} x_{it} \in X_{it}, \\ H_t(x_{it}) = 0, \\ \sum_{i=1}^{I} B_{it}(\xi_t)x_{it} \le d_t(\xi_t), \\ \sum_{\tau=0}^{t-1} A_{it,\tau}(\xi_t)x_{i,t-\tau} = h_{it}(\xi_t) \end{array} \right\}. \tag{4.9}$$

Hence, the I blocks of x are only coupled by the sum $\sum_{i=1}^{I} B_{it}(\xi_t)x_{it}$. By exchanging from $\mathbb{E}$ to a multiperiod polyhedral risk measure this property is maintained, but an additional block consisting of the y_t variables and T additional (dynamic) coupling constraints appear:

$$\min\left\{ \mathbb{E}\left[\sum_{t=1}^{T}\langle c_t, y_t\rangle\right] \;\middle|\; \begin{array}{l} x_{it} \in X_{it},\ y_t \in Y_t, \\ H_t(x_{it}) = 0,\ H_t(y_t) = 0, \\ \sum_{i=1}^{I} B_{it}(\xi_t)x_{it} \le d_t(\xi_t), \\ \sum_{\tau=0}^{t-1} A_{it,\tau}(\xi_t)x_{i,t-\tau} = h_{it}(\xi_t), \\ \sum_{\tau=0}^{t-1}(\langle w_{t,\tau}, y_{t-\tau}\rangle + \sum_{i=1}^{I}\langle b_{i,\tau+1}(\xi_t), x_{i,\tau+1}\rangle) = 0 \end{array} \right\} \tag{4.10}$$

For such block-separable programs, *geographical* or *component decomposition* may lead to efficient algorithms for computing lower bounds. Geographical decomposition is just another notion for *Lagrange relaxation of coupling constraints.* The latter means (cf. [7, 29]) to assign $\mathcal{F}_t$-measurable Lagrange multipliers λ_{1t} and λ_{2t} to the third and fifth constraint in (4.10), respectively, and to arrive at the dual problem

$$\max\left\{D(\lambda_1,\lambda_2) : \lambda_{1t} \in L_{p'}(\Omega,\mathcal{F}_t,\mathbb{P};\mathbb{R}_+^{n_t}),\ \lambda_{2t} \in L_{p'}(\Omega,\mathcal{F}_t,\mathbb{P})\right\}$$

The dual function $D(\lambda_1, \lambda_2)$ given by

$$D(\lambda_1,\lambda_2) = \min \left\{ \begin{array}{l} \mathbb{E}\Big[\sum_{t=1}^T \Big(\langle c_t, y_t\rangle + \Big\langle \lambda_{1t}, \sum_{i=1}^I B_{it}(\xi_t)x_{it} - d_t(\xi_t)\Big\rangle + \\ \lambda_{2t}\sum_{\tau=0}^{t-1}\Big(\langle w_{t,\tau}, y_{t-\tau}\rangle + \sum_{i=1}^I \langle b_{i,\tau+1}(\xi_{\tau+1}), x_{i,\tau+1}\rangle\Big)\Big)\Big] \end{array} \;\middle|\; \begin{array}{l} x_{it}\in X_{it},\ y_t \in Y_t, \\ H_t(x_{it}) = 0,\ H_t(y_t) = 0, \\ \sum_{\tau=0}^{t-1} A_{it,\tau}(\xi_t)x_{i,t-\tau} = h_{it}(\xi_t) \end{array} \right\}$$

and by rearranging with respect to blocks in the objective, the dual function D decomposes into $I+1$ minimization subproblems and is then of the form

$$D(\lambda_1,\lambda_2) = \sum_{i=1}^I D_i(\lambda_1,\lambda_2) + D_R(\lambda_2) - \mathbb{E}\left[\sum_{t=1}^T \langle \lambda_{1t}, d_t(\xi_t)\rangle\right].$$

The functions D_i correspond to I geographical subproblems and D_R to the risk subproblem:

$$D_i(\lambda_1,\lambda_2) = \min\left\{ \mathbb{E}\Big[\sum_{t=1}^T \Big\langle B_{it}(\xi_t)'\lambda_{1t} + b_{it}(\xi_t)\sum_{\tau=t}^T \lambda_{2\tau}, x_{it}\Big\rangle\Big] \;\middle|\; \begin{array}{l} x_{it} \in X_{it}, \\ H_t(x_{it}) = 0, \\ \sum_{\tau=0}^{t-1} A_{it,\tau}(\xi_t)x_{i,t-\tau} = h_{it}(\xi_t) \end{array}\right\}$$

$$D_R(\lambda_2) = \min\left\{ \mathbb{E}\Big[\sum_{t=1}^T \Big\langle c_t + \sum_{\tau=t}^T \lambda_{2\tau} w_{\tau,\tau-t}, y_t\Big\rangle\Big] \;\middle|\; \begin{array}{l} y_t \in Y_t, \\ H_t(y_t) = 0 \end{array}\right\}.$$

Compared to the (dualized form of the) purely expectation-based problem (4.9), the subproblems for the x_i-blocks have the same structure, therefore the same solution methods can be applied. The only change consists in the additional factors $\sum_{\tau=t}^T \lambda_{2\tau}$ of $b_{it}(\xi_t)$ in the objective. If Y_1 is a cone the subproblem for the additional y-block represents a cone constrained linear stochastic program and can be solved explicitly, namely, it holds

$$D_R(\lambda_2) = \left\{ \begin{array}{ll} 0 & \text{, if } -\Big(c_t + \sum_{\tau=t}^T \lambda_{2\tau} w_{\tau,\tau-t}\Big) \in Y_t^* \ (t=1,\ldots,T), \\ -\infty & \text{, otherwise.} \end{array}\right.$$

Hence, the dual problem reads

$$\max\left\{ \sum_{i=1}^I D_i(\lambda_1,\lambda_2) - \mathbb{E}\left[\sum_{t=1}^T \langle \lambda_{1t}, d_t(\xi_t)\rangle\right] \;\middle|\; \begin{array}{l} \lambda_{1t} \in L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}_+^{n_t}),\ \lambda_{2t} \in L_{p'}(\Omega,\mathcal{F}_t,\mathbb{P}), \\ -\Big(c_t + \sum_{\tau=t}^T \lambda_{2\tau} w_{\tau,\tau-t}\Big) \in Y_t^* \ (t=1,\ldots,T) \end{array}\right\}$$

and the whole Lagrangian decomposition strategy has the same favorable features for the risk averse model (4.10) as for the expectation-based one (4.9). For example, the known Lagrangian relaxation based algorithms for electricity portfolio optimization (e.g., [4, 11, 14]) apply to risk averse models after only minor modifications.

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk, *Mathematical Finance* 9 (1999), 203–228.

[2] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., Ku, H.: Coherent multiperiod risk adjusted values and Bellman's principle, manuscript, 2001, downloadable from <www.math.ethz.ch/~delbaen>.

[3] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., Ku, H.: Coherent multiperiod risk measurement, manuscript, 2002, downloadable from <www.math.ethz.ch/~delbaen>.

[4] Bacaud, L., Lemaréchal, C., Renaud, A., Sagastizábal, C.: Bundle methods in stochastic optimal power management: A disaggregated approach using preconditioners. *Computational Optimization and Applications* 20 (2001), 227–244.

[5] Delbaen, F.: Coherent risk measures on general probability spaces, in *Advances in Finance and Stochastics* (K. Sandmann, P. Schönbucher eds.), Springer 2002, 1–37.

[6] Dempster, M.A.H. (ed.): *Risk Management: Value at Risk and Beyond*, Cambridge University Press, 2002.

[7] Dentcheva, D., Römisch, W.: Duality gaps in nonconvex stochastic optimization, *Math. Program.* Ser. A (2004) (to appear).

[8] Eisner, M.J., Olsen, P.: Duality for stochastic programming interpreted as L.P. in L_p-space, *SIAM J. Appl. Math.* 28 (1975), 779–792.

[9] Föllmer, H., Schied, A.: Convex risk measures and trading constraints, *Finance and Stochastics* 6 (2002), 429-447.

[10] Föllmer, H., Schied, A.: *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin 2002.

[11] Gröwe-Kuska, N., Kiwiel, K.C., Nowak, M.P., Römisch, W., Wegner, I.: Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation, in: *Decision Making under Uncertainty: Energy and Power* (Greengard, C., Ruszczyński, A. eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer-Verlag, New York 2002, 39–70.
[12] Kusuoka, S.: On law invariant coherent risk measures, in *Advances in Mathematical Economics* Vol. 3 (S. Kusuoka et al. ed.), Springer, Tokyo 2001, 83-95.
[13] Müller, A., Stoyan, D.: *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester 2002.
[14] Nowak, M.P., Römisch, W.: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty, *Annals of Operations Research* 100 (2000), 251–272.
[15] Ogryczak, W., Ruszczyński, A.: From stochastic dominance to mean-risk models: Semideviations as risk measures, *Eur. J. Oper. Res.* 116 (1999), 33-50.
[16] Ogryczak, W., Ruszczyński, A.: On consistency of stochastic dominance and mean-semideviation models.*Math. Program.* Ser. B, 89 (2001), 217-232.
[17] Pflug, G.: Some remarks on the value-at-risk and the conditional value-at-risk, in *Probabilistic constrained optimization* (Uryasev, S.P. ed.), Kluwer, Dordrecht 2000, 272-281.
[18] Pflug, G.: The value of perfect information as a risk measure, in: *Dynamic Stochastic Optimization* (Marti, K., Ermoliev, Y., Pflug, G. eds.), Lecture Notes in Economics and Mathematical Systems Vol. 532, Springer, Berlin 2003, 275–292.
[19] Pflug, G., Ruszczyński, A.: Risk measures for income streams, in *Risk Measures for the 21st Century* (Szegö, G. ed.), Wiley, 2004 (to appear).
[20] Rachev, S.T.: *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester 1991.
[21] Rachev, S.T., Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27 (2002), 792–818.
[22] Riedel, F.: Dynamic coherent risk measures, manuscript, 2003, downloadable from `<www.uni-bonn.de/~friedel>`.
[23] Rockafellar, R.T.: Duality and stability in extremum problems involving convex functions, *Pac. J. Math.* 21 (1967), 167–187.
[24] Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* 26 (2002), 1443–1471.
[25] Rockafellar, R.T., Uryasev, S., Zarabankin, M.: Deviation Measures in Risk Analysis and Optimization, Research Report 2002-7, Department of Industrial and Systems Engineering, University of Florida, 2002.
[26] Rockafellar, R.T., Wets, R.J-B.: The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints. *SIAM J. Control Optimization* 16 (1978), 16-36.
[27] Rockafellar, R.T., Wets, R.J-B: *Variational Analysis*, Springer, Berlin, 1998.
[28] Römisch, W.: Stability of Stochastic Programming Problems, in *Stochastic Programming* (Ruszczyński A., Shapiro, A. eds.), Handbooks in Operations Research and Management Science Vol. 10, Elsevier, Amsterdam 2003, 483–554.
[29] Römisch, W., Schultz, R.: Multistage stochastic integer programs: An introduction, in *Online Optimization of Large Scale Systems* (Grötschel, M., Krumke, S.O., Rambau, J. eds.), Springer-Verlag, Berlin 2001, 579–598.
[30] Römisch, W., Wets, R.J-B: Stability of ε-approximate solutions to convex stochastic programs, in preparation.
[31] Ruszczyński, A.: Decomposition methods in stochastic programming, *Math. Program.* Ser. B, 79 (1997), 333–353.
[32] Ruszczyński, A., Shapiro, A. (eds.): *Stochastic Programming*, Handbooks of Operations Research and Management Science Vol. 10, Elsevier, Amsterdam 2003.
[33] Ruszczyński, A., Shapiro, A.: Optimization of convex risk functions, manuscript, 2004, downloadable from `<http://www.optimization-online.org/DB_HTML/2004/02/822.html>`.
[34] Scandolo, G.: Risk measures for processes and capital requirement, Preprint, Università degli Studi di Milano, 2003.
[35] Schultz, R.: Stochastic programming with integer variables, *Math. Program.* Ser. B, 97 (2003), 285–309.
[36] Schultz, R., Tiedemann, S.: Risk aversion via excess probabilities in stochastic programs with mixed-integer recourse, *SIAM J. Optim.* (to appear).
[37] Szegö, G.P. (ed.): *Statistical and Computational Problems in Risk Management: VaR and Beyond VaR*, special issue of *Journal of Banking & Finance* 26 (2002) No. 7.
[38] Walkup, D.W., Wets, R.J-B: A Lipschitzian characterization of convex polyhedra, *Proc. Am. Math. Soc.* 23, 167-173 (1969).
[39] Weber, S.: Distribution-invariant dynamic risk measures, manuscript, 2003, downloadable from `<http://www.math.hu-berlin.de/~finance/papers/>`.