Advancing stability analysis of mean-risk stochastic programs: Bilevel and two-stage models

Von der Fakultät für Mathematik der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades eines

Dr. rer. nat.

angenommene Dissertation

von

Matthias Claus

Erstgutachter: Prof. Dr. Rüdiger Schultz, Universität Duisburg-Essen

Zweitgutachter: Prof. i.R. Dr. Werner Römisch, Humboldt-Universität zu Berlin

Tag der mündlichen Prüfung: 08.11.2016

Acknowledgments

First and foremost, I deeply thank my supervisor Prof. Dr. Rüdiger Schultz for his constant support. Without his guidance and encouragement, this thesis would not have been possible.

Furthermore, I owe special thanks to Dr. Volker Krätschmer for our close collaboration. He always found time to discuss mathematical issues and offered invaluable advice.

All members of the working group *Optimization and Algorithmic Discrete Mathematics* have contributed to a very amicable working atmosphere. I am truly grateful to my colleagues for the countless constructive discussions we had.

Last but not least, I thank the *Deutsche Forschungsgemeinschaft* for their support within the Research Training Group GRK 1855 *Discrete optimization of technical systems under uncertainty.*

Abstract

Measuring and managing risk has become crucial in modern decision making under stochastic uncertainty. In two-stage stochastic programming, mean-risk models are essentially defined by a parametric recourse problem and a quantification of risk. This thesis addresses sufficient conditions for weak continuity of the resulting objective functions with respect to perturbations of the underlying probability measure. The approach is based on so called $\psi-$weak topologies that are finer than the topology of weak convergence and allows to unify and extend known results for a comprehensive class of risk measures and recourse problems. In particular, stability of mean-risk models with mixed-integer quadratic and general mixed-integer convex recourse problems is derived for any law-invariant, convex and nondecreasing quantification of risk.

From a conceptual point of view, two-stage stochastic programs and bilevel problems under stochastic uncertainty are closely related. Assuming that only the follower can observe the realization of the randomness, the optimistic and pessimistic setting give rise to two-stage problems where only optimal solutions of the lower level are feasible for the recourse problem. So far, stability in stochastic bilevel programming has only been examined for a specific model based on a quantile criterion. The novel approach allows to identify sufficient conditions for stability of stochastic bilevel problems with quadratic lower level and is applicable for a comprehensive class of risk measures.

Keywords: Mean-risk models, two-stage stochastic programming, stochastic bilevel problems, stability, risk functionals

Contents

1. Introduction

Uncertain data that evolves over time poses a major difficulty in many real-world decision problems. Optimization under uncertainty provides various approaches for resolving this issue. Robust models assume that a so called ambiguity set of possible realizations of the unknown data can be constructed. The worst possible outcome with respect to the ambiguity set is then optimized (see e.g. [32], [34], [35], [36], [37], [38], [39], [97], [98], [114]). While this approach can guarantee a certain quality of the outcome, it is known to produce highly conservative solutions. This has led to the investigation of various modifications that allow to lower the so called price of robustness (see e.g. [33], [44], [67], [112], [194]).

Stochastic programming can be applied if the underlying uncertainty is stochastic and a probability distribution of the unknown data is known. Based on the interplay between decision and observation, one can differ between one-stage, two-stage and multistage stochastic programming models. In one-stage models, all decisions have to be made without any information about the realization of the randomness (see e.g. [168]). Two-stage models have first been considered in [29], [76] and allow for a recourse action after observing the random parameters. More general, multistage models feature a alternating sequence of decision and observation (see e.g. [17], [30], [51], [72], [94], [103], [122], [123], [124], [165], [182], [211]).

In two-stage stochastic programming it is usually assumed that the stochasticity is purely exogenous, i.e. that the distribution of the random data does not depend on any of the decisions to be made (see [113] for a discussion of decision dependent uncertainty). Under this assumption, the problem can be understood as choosing an optimal random variable out of a given family (see e.g. [52], [92], [192]). The simplest risk-neutral model bases this decision on the expected outcome. More sophisticated mean-risk models allow to take various notions of risk aversion into account by punishing high-risk decisions in the objective function. Among many others, applications of mean-risk models include vehicle routing with uncertain demands ([43]), shape optimization ([73]), the management of flood and seismic risks ([100]), scheduling problems in production planning ([99]) and yacht racing ([167]). For specific models and classes of recourse problems, tailored solution algorithms are available (see e.g. [156], [207] (linear recourse), [61] (integer linear recourse), [145], [199] (mixed-integer recourse), [191]).

A different class of models is based on optimizing a utility function over a subset of feasible random variables having an acceptable risk. A popular way of specifying the set of variables with acceptable risk is the introduction of probabilistic constraints (see e.g. [125], [126], [127], [128], [179], [210]). More general models utilize stochastic dominance constraints with respect to a benchmark variable. The seminal paper in this field is [89]. Other works focus on applications (see e.g. [60], [73], [90], [115]), more general models (see e.g. [91]), solution methods (see e.g. [85], [87], [88], [93], [101], [116], [130], [150], [157], [187]) and stability (see e.g. [70], [84], [86], [117], [149]). The works [158], [159] and [160] point out links between stochastic dominance and certain mean-risks models.

All of the mentioned stochastic programming models depend on the distribution of the random data. In real-world applications, only an approximation of this distribution may be available, which motivates to examine stochastic problems from a parametric optimization point of view. Stability of the optimal value and the set of optimal solutions under perturbations of the underlying random vector (or rather the probability measure induced by it) are of particular interest (see e.g. [13], [14], [178], [183], [184], [196], [198], [202], [203]). Since the parameter space is infinite-dimensional, the choice of a topology becomes an issue. For qualitative stability analysis, equipping the space of Borel probability measures with the topology of weak convergence has proven to be instrumental (see e.g. [133], [135], [174]). Quantitative stability has been investigated based on suitable probability metrics (see e.g. [170], [180], [181], [185], [186], [197], [204]).

Stability of two-stage mean-risk models is closely related to the continuity of certain risk functionals that depend on the quantification of risk and the underlying deterministic problem. The present thesis provides a systematic approach for deriving weak continuity of a general class of functionals defined on subspaces of Borel probability measures satisfying certain moment conditions. Such functionals are continuous with respect to a finer, so called $\psi-$weak topology (see e.g. [107], [141], [143]). Since it is possible to exactly point out the subsets on which the relative $\psi-$weak topology and the relative topology of weak convergence coincide (see [142], [218]), the latter allows to derive weak continuity of suitable restrictions. The special case of restrictions to sets of measures with uniformly bounded moments of higher order is well established in stability analysis of stochastic programs (see e.g. [153], [196], [198], [200], [201]). In view of two-stage mean-risk models, the approach is applicable whenever the growth of the optimal value function of the recourse problem is polynomially bounded in the entering random parameter. Furthermore, it is assumed that the set of discontinuities is of measure zero with respect to the original probability measure and that the quantification of risk is law-invariant, convex and nondecreasing.

In qualitative robustness theory, convex (monetary) risk measures (see e.g. [106], [109]) provide a well established generalization of coherent measures of risk (see e.g. [1], [15]). Every law-invariant, convex risk measure is nondecreasing and hence meets the criteria described in the previous paragraph. Such functionals are of special interest due to their analytical traits (see e.g. [141], [144]), which have an immediate impact on statistical properties, e.g. in view of the sample average approximation method (see e.g. [31], [92], [166], [190]).

Stability analysis in two-stage stochastic programming often focuses on the case where the underlying deterministic problem is a mixed-integer linear program. For this special situation, stability results for various types of two-stage mean-risk models are available (see e.g. [153], [172], [200], [201]). However, the proofs differ greatly. For mixed-integer quadratic recourse, stability of a risk-neutral model has been investigated in [64]. The present thesis provides an umbrella for two-stage mean-risk models that allows to unify and extend the known results in various directions: Stability is derived for a comprehensive class of both risk measures and underlying deterministic problems (e.g. for mixed-integer quadratic problems and a fairly general class of mixed-integer problems where the continuous relaxation is convex). Furthermore, stability of stochastic bilevel problems is examined. Although not detailed in this thesis, the approach has also been applied to investigate stability of mean-risk formulations of stochastic complementarity problems (see [57]).

Bilevel problems arise from the interplay between two decision makers on different levels of a hierarchy. The *leader* decides first and passes the *upper level decision* on to the *follower*. Incorporating the leader's decision as a parameter, the follower then solves the *lower level problem* that reflects his or her own goals and returns an optimal solution back to the leader. The leader's objective function depends on both his or her decision and the solution that is fed back from the lower level. In bilevel optimization, it is assumed that the leader has full information about the influence of his or her decision on the lower level problem. As the latter may have more than one solution, one typically assumes that the follower returns either the best (*optimistic approach*) or the worst (*pessimistic approach*) solution with respect to the leader's objective. The bilevel optimization problem is to find an optimal upper level decision. Such problems have first been considered in economics ([208]). For a general discussion of bilevel programming, refer to [75] or [78]. Other works focus on applications (see e.g. [24], [27], [81]), more general multilevel models (see e.g.

[26], [58]), linear bilevel programming (see e.g. [25], [47], [59]), optimality conditions (see e.g. [23], [62], [79], [80], [82], [83], [213], [214], [215], [216]) solution methods (see e.g. [5], [6], [132]) or stability (see e.g. [129]).

In stochastic bilevel programming, the realization of some random vector whose distribution does not depend on the upper level decision enters the problem as an additional parameter. It is assumed that the leader has to make his or her decision without knowing the random parameter, while the follower decides under full information. Stochastic bilevel problems can be seen as an extension of classical two-stage stochastic programs, where upper and lower level mirror first and second stage, respectively. As in those problems, the upper level objective function gives rise to a random variable. However, this random variable now depends on *an* optimal solution rather than just on *the* optimal value of the lower level (or second stage) problem. This is a crucial difference that results in weaker analytical properties and a less stable behavior.

Nevertheless, stochastic bilevel problems are of great relevance for practical applications and have been discussed in the context of transportation ([9], [161]), the pricing of electricity swing options ([119], [140]), economics ([12]), supply chain planning ([16]), telecommunications ([212]), structural optimization ([66]) and general agency problems ([111]). Other works focus on solution methods ([50]), stochastic bilevel problems with Knapsack constraints ([139]), nonlinear bilevel programming under uncertainty ([162]) or stochastic equilibrium problems ([110]) and their stability ([148], [163]).

So far, the structure of risk-averse stochastic bilevel problems has only been addressed in the recent work [131], where problems based on a so called quantile criterion are considered. Using the optimistic approach and assumptions on the linearity of the upper and lower level problems, continuity of the objective function with respect to the leader's decision is shown. However, the underlying probability measure is assumed to be fixed and stability of optimal values and optimal solution sets is not examined. The focus in this thesis is on stability of more general mean-risk formulations of stochastic bilevel problems. The present analysis also applies to quadratic lower level problems and allows for a more general dependence on the random parameter.

Chapter 2 introduces the theoretical framework while paying special attention to the similarities and differences between the topology of weak convergence and $\psi-$weak topologies. A proof of the main continuity result is given in the final section. In chapter 3, sufficient conditions for qualitative stability of various two-stage models is examined based on the previous findings. Finally, chapter 4 extends the results to stochastic bilevel problems.

Parts of this chapter have also been submitted for publication (see [68] for a preprint).

2. Weak continuity of risk functionals

Mean-risk models in both two-stage and bilevel programming under stochastic uncertainty give rise to functionals defined on certain subspaces of Borel probability measures. This chapter examines their continuity with respect to the topology of weak convergence as well as a finer, so called ψ-weak topology and is organized in five sections: Imposing a growth condition and assumptions on the underlying quantification of risk, the considered functionals are characterized in section 2.1. Section 2.2 examines quantifications of risk that are suitable for the proposed setting, while sections 2.3 and 2.4 address selected properties of the relevant topologies as well as their relation. Finally, section 2.5 is devoted to proving the desired continuity.

Parts of this chapter have also been submitted for publication (see [69] for a preprint).

2.1. Setting and basic assumptions

Throughout this thesis, the focus will be on functionals defined on spaces of Borel probability measures. Hence, it seems appropriate to begin with recalling the following basic definitions:

Definition 2.1. *The* ***Borel*** $\sigma-$***algebra*** $\mathcal{B}(\mathbb{R}^s)$ *of* $\mathbb{R}^s$ *is the* $\sigma-$*algebra generated by the family of open sets.*

Properties of Borel $\sigma-$algebras are discussed in section 4.4 in [7]. Note that most of the concepts introduced in this chapter can be translated to the case where $\mathbb{R}^s$ is replaced with a general metric space S. However, it seems reasonable to confine the present discussion to the case relevant for the applications in chapters 3 and 4. For generalizations, refer to the corresponding chapters in [7] or section A.6 in [107].

Definition 2.2. *A* ***Borel probability measure*** *on* $\mathbb{R}^s$ *is a countably additive set function* $\mu : \mathcal{B}(\mathbb{R}^s) \to [0, \infty)$ *satisfying* $\mu[\mathbb{R}^s] = 1$. *The space of all such measures is denoted by* $\mathcal{P}(\mathbb{R}^s)$.

An important class of subsets of $\mathcal{P}(\mathbb{R}^s)$ can be defined via (generalized) moment conditions:

Definition 2.3. *For any continuous function $\psi : \mathbb{R}^s \to [0, \infty)$, define*

$$\mathcal{M}_s^{\psi} := \{\mu \in \mathcal{P}(\mathbb{R}^s) \mid \int_{\mathbb{R}^s} \psi(t)\ \mu(dt) < \infty\}.$$

Furthermore, set $\mathcal{M}_s^p := \mathcal{M}_s^{\|\cdot\|^p}$, where $p \in \mathbb{R}$ is a positive constant and $\|\cdot\|$ denotes the Euclidean norm.

Remark 2.4. *Due to the fact that every finite dimensional vector space admits a unique Hausdorff linear topology [7, Theorem 5.21], all norms on $\mathbb{R}^s$ are equivalent and the set $\mathcal{M}_s^p$ does not change if a norm other than the Euclidean one is considered.*

Remark 2.5. *The case where ψ is a so called gauge function (see Definition 2.51) will be of special interest in section 2.3).*

The functionals to be analyzed are induced by mappings f and ρ defined on $\mathbb{R}^n \times \mathbb{R}^s$ and some L^p-space, respectively (see chapter 13 in [7] for a discussion of L^p-spaces). All assumptions needed to derive the desired continuity will be imposed on these mappings. In view of the mean-risk models considered in chapters 3 and 4, assumptions on f correspond to assumptions on the underlying parametric problem, while ρ is directly related to the choice of the quantification of risk in the objective function.

The assumptions imposed on f can be formulated using the notion of locally bounded mappings:

Definition 2.6. *A mapping $\eta : \mathbb{R}^n \to \mathbb{R}$ is said to be* ***locally bounded*** *iff the convergence of $\{x_l\}_{l \in \mathbb{N}} \subset \mathbb{R}^n$ implies the boundedness of $\{\eta(x_l)\}_{l \in \mathbb{N}} \subset \mathbb{R}$.*

Remark 2.7. *In particular, all continuous and all bounded functions are locally bounded.*

Assumption 2.8 ($[\mathbb{A}_f]$: **Assumptions on** f**).**
$f : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ is Borel measurable and fulfills the following ***growth condition****: There is an exponent $\gamma \geq 0$ and some locally bounded mapping $\eta : \mathbb{R}^n \to \mathbb{R}$ such that*

$$|f(x, z)| \leq \eta(x)(\|z\|^{\gamma} + 1) \tag{2.1}$$

holds for any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$.

Remark 2.9. $[\mathbb{A}_f]$ *is especially fulfilled if there exist positive constants α and β such that*

$$|f(x, z) - f(x', z')| \leq \alpha \|(x, z) - (x', z')\|^{\beta}$$

holds for any $(x,z),(x',z') \in \mathbb{R}^n \times \mathbb{R}^s$. *In particular,* $[\mathbb{A}_f]$ *holds if* f *is jointly Hölder continuous with respect to* x *and* z.

The following lemma will prove useful:

Lemma 2.10. *Assume* $[\mathbb{A}_f]$, *then* $\int_{\mathbb{R}^s} |f(x,z)|^p \ \nu(dz) < \infty$ *holds for every* $p \geq 0$ *and* $(x,\nu) \in \mathbb{R}^n \times \mathcal{M}_s^{\gamma p}$.

Proof. By (2.1),

$$|f(x,z)|^p \leq \eta(x)^p (\|z\|^\gamma + 1)^p \leq \eta(x)^p 2^p \max\{\|z\|^{\gamma p}, 1\} \leq \eta(x)^p 2^p (\|z\|^{\gamma p} + 1)$$

holds for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$ and the statement of Lemma 2.10 follows immediately from the fact that $\int_{\mathbb{R}^s} 1 \ \nu(dz) = 1$. □

Lemma 2.10 can be restated in terms of finiteness of moments of a certain image measure under f:

Corollary 2.11. *Fix* $p \geq 0$ *as well as* $(x,\nu) \in \mathbb{R}^n \times \mathcal{M}_s^{\gamma p}$ *and let* $\delta_x \otimes \nu$ *denote the product probability measure of the Dirac measure at* x *and* ν. *Then, under assumption* $[\mathbb{A}_f]$, *the image measure of* $\delta_x \otimes \nu$ *under* f *has finite moments of order* p, *i.e.* $(\delta_x \otimes \nu) \circ f^{-1} \in \mathcal{M}_1^p$.

Proof. By the change-of-variable formula (see Theorem A.1), it holds that

$$\int_{\mathbb{R}} |t|^p \ ((\delta_x \otimes \nu) \circ f^{-1})(dt) = \int_{\mathbb{R}^n \times \mathbb{R}^s} |f(x',z)|^p \ (\delta_x \otimes \nu)(d(x',z)) = \int_{\mathbb{R}^s} |f(x,z)|^p \ \nu(dz)$$

and Lemma 2.1 completes the proof. □

The concept of atomless probability spaces plays an important role in the assumptions on ρ:

Definition 2.12. *A probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is said to be* ***atomless*** *or* ***nonatomic*** *iff for every* $A \in \mathcal{F}$ *with* $\mathbb{P}[A] > 0$ *there exists a set* $B \in \mathcal{F}$ *satisfying* $B \subsetneq A$ *and* $\mathbb{P}[A] > \mathbb{P}[B] > 0$.

In the field of monetary risk measures, it is common to confine the analysis to atomless probability spaces (see e.g. [11], [107] (chapter 4), [108] or [141] and refer to [77] for the more general case). The following result illustrates some properties of such spaces:

Proposition 2.13 (**[107, Proposition A.27]**).
For any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, *the following conditions are equivalent:*

(a) $(\Omega, \mathcal{F}, \mathbb{P})$ *is atomless.*

(b) There exists an independent and identically distributed sequence of random variables $Y_1, Y_2, \ldots$ *with Bernoulli distribution* $\mathbb{P}[Y_i = 0] = \mathbb{P}[Y_i = 1] = \frac{1}{2}$ $(i = 1, 2, \ldots)$.

(c) For every $\sigma \in \mathcal{P}(\mathbb{R})$*, there exist independent and identically distributed random variables* $Z_1, Z_2, \ldots$ *with common distribution* F_σ *defined by* $F_\sigma(t) := \sigma[(-\infty, t]]$.

(d) $(\Omega, \mathcal{F}, \mathbb{P})$ *supports a random variable with continuous distribution.*

Example 2.14. *Let* $\lambda^1_{[0,1]}$ *denote the restriction of the one-dimensional Lebesgue measure to the closed unit interval. By the equivalence of (a) and (d) in Proposition 2.13, the probability space* $([0, 1], \mathcal{B}(\mathbb{R}) \cap [0, 1], \lambda^1_{[0,1]})$ *is atomless.*

By Proposition 2.13, every atomless probability space supports a random variable U that is uniformly distributed on the open unit interval $(0, 1)$. This allows to explicitly associate Borel probability measures on $\mathbb{R}$ with random variables on the space via a so called quantile transformation. The approach is also used in the proof of the above proposition in [107] and justified by the following lemma:

Lemma 2.15 (Quantile transformation).
Fix a random variable U *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *that is uniformly distributed on* $(0, 1)$ *and let* $\sigma \in \mathcal{P}(\mathbb{R})$ *be a Borel probability measure. Then*

$$\omega \mapsto F_\sigma^{-1}(U(\omega)) := \inf\{t \in \mathbb{R} \mid F_\sigma(t) \geq U(\omega)\} \tag{2.2}$$

is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ *with distribution* F_σ.

Proof. F_σ is a normalized, nondecreasing, right-continuous function and the mapping F_σ^{-1} yields an inverse function in the sense of [107, Definition A.14]. Hence, [107, Lemma A.19] is applicable. □

Remark 2.16. *The function* F_σ^{-1} *is referred to as the (left-continuous)* ***quantile function*** *associated with* σ*, which motivates to call* (2.2) *the quantile transformation.*

By the previous results, a mapping defined on the space of random variables on some atomless probability space induces a function on $\mathcal{P}(\mathbb{R})$ if it only depends on the distribution of the entering random variables. Such mappings are called law-invariant:

Definition 2.17. *Let* $L^0(\Omega, \mathcal{F}, \mathbb{P})$ *denote the space of finite-valued random variables on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. A mapping* $\rho : L^0(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ *is called* ***law-invariant*** *iff* $\rho[Y] = \rho[Z]$ *holds whenever* Y *and* Z *have the same distribution under* $\mathbb{P}$.

Assumption 2.18 ($[\mathbb{A}_\rho]$: **Assumptions on** ρ)**.**
$\rho : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is a real-valued mapping defined on the L^p-space of some atomless probability space, where $p \geq 1$. In addition, ρ is law-invariant, convex, i.e.

$$\rho[mY + (1-m)Z] \leq m\rho[Y] + (1-m)\rho[Z] \quad \forall m \in [0,1] \quad \forall Y, Z \in L^p(\Omega, \mathcal{F}, \mathbb{P}),$$

and nondecreasing with respect to the $\mathbb{P}-$almost sure partial order, i.e.

$$\mathbb{P}[Y \leq Z] = 1 \Rightarrow \rho[Y] \leq \rho[Z] \quad \forall Y, Z \in L^p(\Omega, \mathcal{F}, \mathbb{P}).$$

The following simple observation will prove useful in the context of mean-risk models in chapter 3:

Remark 2.19. *For any atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and any constant $p \geq 1$, the space of mappings $\rho : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ satisfying assumption $[\mathbb{A}_\rho]$ is closed under conic combinations.*

Let assumptions $[\mathbb{A}_f]$ and $[\mathbb{A}_\rho]$ be fulfilled and consider a random variable U on the atomless probability space from $[\mathbb{A}_\rho]$ that is uniformly distributed on $(0, 1)$. By Lemma 2.15, the mapping $\mathcal{R}_\rho : \mathcal{M}_1^p \to \mathbb{R}$ given by

$$\mathcal{R}_\rho(\sigma) := \rho[F_\sigma^{-1}(U)] \tag{2.3}$$

is well defined. The subsequent analysis shall focus on functionals $Q : \mathbb{R}^n \times \mathcal{M}_s^{\gamma p} \to \mathbb{R}$ defined by

$$Q(x, \mu) := \mathcal{R}_\rho((\delta_x \otimes \mu) \circ f^{-1}). \tag{2.4}$$

Remark 2.20. *Q is well defined by Corollary 2.11.*

2.2. Suitable risk measures

This section points out links to the theory of monetary risk measures and provides relevant examples of mappings that satisfy assumptions $[\mathbb{A}_\rho]$. The notion of convex risk measures in the sense of mathematical finance will be of special interest:

Definition 2.21. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{X}$ a linear subspace of $L^0(\Omega, \mathcal{F}, \mathbb{P})$ containing the constants. A mapping $\varpi : \mathcal{X} \to \mathbb{R}$ is called a* ***convex (monetary) risk measure*** *if ϖ is convex, nondecreasing with respect to the $\mathbb{P}$-almost sure partial order and* *translation-equivariant, i.e.*

$$\varpi[Y+m] = \varpi[Y] + m \quad \forall m \in \mathbb{R} \quad \forall Y \in \mathcal{X}. \tag{2.5}$$

A convex risk measure ϖ is called ***coherent*** *if it is positively homogeneous, i.e.*

$$\varpi[mY] = m\varpi[Y] \quad \forall m \in [0,\infty) \quad \forall Y \in \mathcal{X}.$$

Remark 2.22. *$\mathcal{X}$ is often chosen as some L^p-space, where $p \in [0,\infty]$ (see e.g. [96]). In view of $[\mathbb{A}_\rho]$, only the case where $\mathcal{X} = L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $1 \leq p < \infty$ will be relevant for the analysis in section 2.5.*

Remark 2.23. *Monetary risk measures have also been considered in settings different from the one of the above definition (see e.g. [105], where no probability measure is fixed). Note that some definitions include a normalization, i.e. $\varpi[0] = 0$ (see e.g. section 4.1 in [107]) or work with possible gains rather than losses. In such a setting, ϖ is required to be nonincreasing and* (2.5) *is replaced with*

$$\varpi[Y+m] = \varpi[Y] - m \quad \forall m \in \mathbb{R} \quad \forall Y \in \mathcal{X}$$

(see e.g. [65], [96] or [141]). However, Definition 2.21 covers the framework needed in view of $[\mathbb{A}_\rho]$.

Convex risk measures have been introduced independently in [106] and [109] and generalize the concept of coherent risk measures that originates from [15]. Their relevance for the present work is pointed out in the following remark:

Remark 2.24. *Assumption $[\mathbb{A}_\rho]$ is fulfilled for every law-invariant, convex risk measure (in the sense of Definition 2.21) that is defined on $L^p(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless and $1 \leq p \leq \infty$.*

Law-invariant, convex risk measure are well established in the field of mathematical finance (see e.g. [1], [31], [65], [104], [108] and [141]) and highly relevant for stochastic programming. By [96, Corollary 2.5], polyhedral risk measures are coherent (and hence convex) under mild standard assumptions. The following examples discuss various risk measures of importance in stochastic programming:

Example 2.25. *The* ***expectation*** *or* ***mean*** *$\mathbb{E} : L^1([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ defined by*

$$\mathbb{E}[Y] := \int_0^1 Y(t)\ \lambda^1_{[0,1]}(dt)$$

is a law-invariant, coherent risk measure in the sense of Definition 2.21.

Example 2.26. *Consider the* ***variance*** $\mathrm{Var} : L^2([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ *defined by*

$$\mathrm{Var}[Y] := \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2.$$

Var *is law-invariant and convex (see e.g. [56]): Let* $\mathrm{Cov}[Y,Z] := \mathbb{E}[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])]$ *denote the covariance of Y and Z. By* $\mathrm{Cov}[Y,Z]^2 \leq \mathrm{Var}[Y]\mathrm{Var}[Z]$ *and the convexity of the mapping $x \mapsto x^2$,*

$$\begin{aligned}\mathrm{Var}[mY + (1-m)Z] &= m^2\mathrm{Var}[Y] + (1-m)^2\mathrm{Var}[Z] + 2m(1-m)\mathrm{Cov}[Y,Z] \\ &\leq (m\sqrt{\mathrm{Var}[Y]} + (1-m)\sqrt{\mathrm{Var}[Z]})^2 \\ &\leq m\mathrm{Var}[Y] + (1-m)\mathrm{Var}[Z]\end{aligned}$$

holds for any $Y, Z \in L^2([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]})$ and $m \in [0,1]$.

Since $\mathrm{Var}[Y+m] = \mathrm{Var}[Y]$ *holds for any $m \in \mathbb{R}$ and $Y \in L^2([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]})$,* Var *is not translation-equivariant. In addition, the variance is not nondecreasing: Consider the random variables $Y, Z : [0,1] \to \mathbb{R}$ defined by $Y(\omega) \equiv 1$ and $Z(\omega) = \chi_{[\frac{1}{2},1]}(\omega)$, where $\chi_{[\frac{1}{2},1]}$ denotes the indicator function of the interval $[\frac{1}{2}, 1]$. Although $Y \geq Z$ almost surely, it holds that* $\mathrm{Var}[Y] = 0 < \frac{1}{4} = \mathrm{Var}[Z]$*. Finally, the variance is not positively homogenous:* $\mathrm{Var}[2Z] = 1 \neq \frac{1}{2} = 2\mathrm{Var}[Z]$*.*

In stochastic programming, mean-variance models are seldom considered as they are known to be computationally intractable even for simple stochastic problems (see e.g. [4]).

Example 2.27. *The* ***expected excess*** $\rho^{EE}_\alpha : L^1([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ *of a predefined target level $\alpha \in \mathbb{R}$ given by*

$$\rho^{EE}_\alpha[Y] := \mathbb{E}[\max\{Y - \alpha, 0\}]$$

is law-invariant and nondecreasing. Furthermore, it is jointly convex with respect to Y and α, i.e.

$$\rho^{EE}_{m\alpha_1 + (1-m)\alpha_2}[mY_1 + (1-m)Y_2] \leq m\rho^{EE}_{\alpha_1}[Y_1] + (1-m)\rho^{EE}_{\alpha_2}[Y_2] \tag{2.6}$$

holds for any $Y_1, Y_2 \in L^1([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]})$, $\alpha_1, \alpha_2 \in \mathbb{R}$ and $m \in [0,1]$. Consequently, assumption $[\mathbb{A}_\rho]$ is fulfilled for the expected excess. However, the expected excess is not translation-equivariant and hence no monetary risk measure: The random variable

$Y : [0,1] \to \mathbb{R}$, $Y \equiv 1$ yields the counterexample $\rho_2^{EE}[Y] = 0 = \rho_2^{EE}[Y+1]$. Finally, the expected excess is not positively homogenous: $\rho_2^{EE}[3Y] = 1 \neq 0 = 3\rho_2^{EE}[Y]$.

*The expected excess has been examined in the context of two-stage stochastic programming problems with mixed-integer linear recourse in [153]. For any $q \geq 1$, the present analysis extends to the q-**th order expected excess** $\rho_\alpha^{EE,q} : L^q([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ of the target level $\alpha \in \mathbb{R}$ defined by*

$$\rho_\alpha^{EE,q}[Y] := \mathbb{E}[\max\{Y - \alpha, 0\}^q]^{\frac{1}{q}}.$$

Assumption $[\mathbb{A}_\rho]$ is fulfilled for $\rho_\alpha^{EE,q}$, since it is law-invariant, convex and nondecreasing.

Example 2.28. *The **semideviation** $\rho^{SD} : L^1([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ given by*

$$\rho^{SD}[Y] := \rho^{EE}_{\mathbb{E}[Y]}[Y]$$

is law-invariant, positively homogenous and convex. The latter is a direct conclusion from the linearity of the expectation and the joint convexity of the expected excess (2.6)*. However, the semideviation is not translation-equivariant, as $\rho^{SD}[Y+m] = \rho^{SD}[Y]$ for any $Y \in L^1([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]})$ and $m \in \mathbb{R}$. In addition, ρ^{SD} is not nondecreasing: Consider the random variables $Y, Z : [0,1] \to \mathbb{R}$ given by $Y(t) = \chi_{[0,\frac{1}{2})}(t) - \chi_{[\frac{1}{2},1]}(t)$ and $Z \equiv 1$. Although $Y \leq Z$ almost surely, it holds that $\rho^{SD}[Y] = \frac{1}{2} > 0 = \rho^{SD}[Z]$.*

One can easily compensate for the lacking monotonicity by considering a weighted sum with the expectation: For any $m \in [0,1]$, $\mathbb{E} + m\rho^{SD} : L^1([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ is a law-invariant, coherent risk measure in the sense of Definition 2.21 (see e.g. section 6.3.2 in [92]).

For a discussion of the semideviation in the context of two-stage stochastic programming with mixed-integer linear recourse, refer to [153].

Example 2.29. *The **excess probability** $\rho_\alpha^P : L^0([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ of a predefined target level $\alpha \in \mathbb{R}$ given by*

$$\rho_\alpha^P[Y] := \lambda^1_{[0,1]}[\{t \in [0,1] \mid Y(t) > \alpha\}]$$

is nondecreasing and law-invariant. However, it lacks convexity, translation-equivariance and positive homogeneity: Consider the random variables $Y, Z : [0,1] \to \mathbb{R}$ given by $Y \equiv 1$ and $Z \equiv 0$. Since

$$\rho^P_{\frac{1}{2}}[2Y] = \rho^P_{\frac{1}{2}}[Y+1] = 1 \neq 2 = \rho^P_{\frac{1}{2}}[Y] + 1 = 2\rho^P_{\frac{1}{2}}[Y],$$

the excess probability is neither positively homogenous nor translation-equivariant. Furthermore, the calculation

$$\rho^P_{\frac{1}{4}}[\frac{1}{2}Y + \frac{1}{2}Z] = 1 > \frac{1}{2} = \frac{1}{2}\rho^P_{\frac{1}{4}}[Y] + \frac{1}{2}\rho^P_{\frac{1}{4}}[Z]$$

shows that ρ^P_α *is nonconvex in general.*

In the context of two-stage stochastic programming, the excess probability has been investigated for example in [40], [172] and [201].

Example 2.30. *The* ***value-at-risk*** $\mathrm{VaR}_\alpha : L^\infty([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ *for a predefined level* $\alpha \in (0,1)$ *given by*

$$\mathrm{VaR}_\alpha[Y] := \inf\{m \in \mathbb{R} \mid \lambda^1_{[0,1]}[\{t \in [0,1] \mid Y(t) \leq m\}] \geq \alpha\}$$

is law-invariant, nondecreasing, translation-equivariant and positively homogenous. However, it is not convex: Consider the random variables $Y, Z : [0,1] \to \mathbb{R}$ *defined by* $Y(t) = \chi_{[0,\frac{1}{2})}(t) + 3\chi_{[\frac{1}{2},1]}(t)$ *and* $Z(t) = 3\chi_{[0,\frac{1}{2})}(t) + \chi_{[\frac{1}{2},1]}(t)$*. It holds that*

$$\mathrm{VaR}_{\frac{1}{4}}[\frac{1}{2}Y + \frac{1}{2}Z] = 2 > 1 = \frac{1}{2}\mathrm{VaR}_{\frac{1}{4}}[Y] + \frac{1}{2}\mathrm{VaR}_{\frac{1}{4}}[Y].$$

A detailed discussion of the value-at-risk is provided in [164].

The lack of convexity of the value-at-risk has inspired the investigation of coherent alternatives like the conditional value-at-risk (see e.g. [2], [3], [10], [164], [175] and [176]). [200] provides a discussion in the context of two-stage stochastic programming problems with mixed-integer recourse.

Example 2.31. *The* ***conditional value-at-risk****, also known as* ***average value-at-risk*** *or* ***expected shortfall****,* $\mathrm{CVaR}_\alpha : L^1([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \lambda^1_{[0,1]}) \to \mathbb{R}$ *for a predefined level* $\alpha \in (0,1)$ *is given by*

$$\mathrm{CVaR}_\alpha[Y] := \inf\{m + \frac{1}{1-\alpha}\rho^{EE}_m[Y] \mid m \in \mathbb{R}\}.$$

CVaR_α *is a law-invariant, coherent risk measure (see e.g. [164, Proposition 2]).*

2.3. The topology of weak convergence

The present chapter aims at verifying continuity of the mapping Q with respect to the topology of weak convergence. Dating back to a least 1978 (see [135]), the use of this topology has proven to be instrumental in stability analysis for stochastic programming models. While coarse enough to be relevant for a large spectrum of applications (see e.g. Remarks 2.44 and 2.50), the topology possesses very desirable mathematical properties (see e.g. Proposition 2.37, Theorems 2.41, 2.48 and 2.49). Discussing selected results and characterizations, this section aims to make the case for utilizing the topology of weak convergence.

Definition 2.32. *Let $C_b^0(\mathbb{R}^s)$ denote the linear space of all bounded and continuous functions $h : \mathbb{R}^s \to \mathbb{R}$. The **topology of weak convergence**, denoted by τ_w^s, is the coarsest topology on $\mathcal{P}(\mathbb{R}^s)$ for which all mappings $g_h : \mathcal{P}(\mathbb{R}^s) \to \mathbb{R}$ defined by*

$$g_h(\mu) = \int_{\mathbb{R}^s} h(t)\ \mu(dt), \quad h \in C_b^0(\mathbb{R}^s) \tag{2.7}$$

*are continuous. A sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^s)$ is said to **converge weakly** to $\mu \in \mathcal{P}(\mathbb{R}^s)$, written $\mu_l \overset{w}{\to} \mu$, iff it converges with respect to τ_w^s.*

By the following result, a Borel probability measure μ is completely determined by the integrals in (2.7):

Proposition 2.33 (**[49, Theorem 1.2]**).
For $\mu, \nu \in \mathcal{P}(\mathbb{R}^s)$, it holds that $\mu = \nu$ iff $g_h(\mu) = g_h(\nu)$ for all $h \in C_b^0(\mathbb{R}^s)$.

Remark 2.34. *In particular, Proposition 2.33 yields that weak limits are unique.*

By Definition 2.32,

$$\lim_{l\to\infty} \int_{\mathbb{R}^s} h(t)\ \mu_l(dt) = \int_{\mathbb{R}^s} h(t)\ \mu(dt) \quad \forall h \in C_b^0(\mathbb{R}^s)$$

holds whenever $\mu_l \overset{w}{\to} \mu$, as continuity implies sequential continuity. Moreover, since τ_w^s is metrizable by the Prokhorov metric (see Proposition 2.37), the notions of continuity and sequential continuity coincide and the converse statement is also true.

Definition 2.35. *The **Prokhorov metric** $\pi : \mathcal{P}(\mathbb{R}^s) \times \mathcal{P}(\mathbb{R}^s) \to [0, \infty)$ is defined by*

$$\pi(\mu, \nu) := \inf\{\epsilon > 0 \mid \mu[A] \leq \nu[A + B_\epsilon(0)] + \epsilon, \nu[A] \leq \mu[A + B_\epsilon(0)] + \epsilon \ \ \forall A \in \mathcal{B}(\mathbb{R}^s)\},$$

where $A + B_\epsilon(0) \subseteq \mathbb{R}^s$ denotes the Minkowski sum of A and the open ball of radius ϵ centered at 0 (with respect to the Euclidean norm).

Remark 2.36. *π is indeed a metric (see fact (i) on page 72 in [49]).*

Proposition 2.37. *τ_w^s coincides with the topology induced by the Prokhorov metric π on $\mathcal{P}(\mathbb{R}^s)$.*

Proof. Combine facts (iii) and (iv) on page 72 in [49] and invoke the separability of $\mathbb{R}^s$. □

It is also possible to characterize the topology of weak convergence in terms of functional analysis: Endow $C_b^0(\mathbb{R}^s)$ with the supremum norm

$$\|h\|_\infty = \sup_{t\in\mathbb{R}^s} |h(t)| \tag{2.8}$$

and let $(C_b^0(\mathbb{R}^s))^*$ denote its dual space equipped with the norm given by

$$\|\zeta\|_\infty^* = \sup\{|\zeta(h)| \mid h \in C_b^0(\mathbb{R}^s),\ \|h\|_\infty \le 1\}. \tag{2.9}$$

[7, Theorem 14.10] allows to identify $(C_b^0(\mathbb{R}^s))^*$ with the AL-space (abstract Lebesgue space) of normal signed Borel charges of bounded variation on $\mathcal{B}(\mathbb{R}^s)$ (see section 10.10 as well as Definitions 10.2, 12.2 and 12.11 in [7]).

Theorem 2.38 (Riesz representation theorem, see e.g. Theorem 2.14 in [189]**).** *Let $\mathcal{Z}$ denote the set of all $\zeta \in (C_b^0(\mathbb{R}^s))^*$ satisfying $\|\zeta\|_\infty^* = 1$ and $\zeta(h) \ge 0$ for any nonnegative $h \in C_b^0(\mathbb{R}^s)$. Then for any $\zeta \in \mathcal{Z}$, there exists a unique $\mu_\zeta \in \mathcal{P}(\mathbb{R}^s)$ satisfying*

$$\zeta(h) = \int_{\mathbb{R}^s} h(t)\ \mu_\zeta(dt) \quad \forall h \in C_b^0(\mathbb{R}^s).$$

Moreover, the mapping $\Lambda : \mathcal{Z} \to \mathcal{P}(\mathbb{R}^s)$ defined by $\Lambda(\zeta) := \mu_\zeta$ is bijective.

The following result is an immediate conclusion from the Riesz representation theorem and the definition of weak* convergence:

Corollary 2.39. *For any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^s)$ the following statements are equivalent:*

(a) $\mu_l \overset{w}{\to} \mu_1$.

(b) $\Lambda^{-1}(\mu_l) \overset{w^*}{\to} \Lambda^{-1}(\mu_1)$, *where* $\overset{w^*}{\to}$ *denotes the weak* convergence on $\mathcal{Z}$, i.e.* $\zeta_l \overset{w^*}{\to} \zeta$ *iff* $\lim_{l\to\infty} \zeta_l(h) = \zeta(h)$ *for any* $h \in C_b^0(\mathbb{R}^s)$.

Another way to characterize a topology is to point out a base:

Proposition 2.40. *The sets*

$$N_\epsilon(\alpha, \mu, h_1, \dots, h_l) := \bigcap_{i=1}^{l} \left\{ \nu \in \mathcal{P}(\mathbb{R}^s) \mid \left| \int_{\mathbb{R}^s} h_i(t)\ (\alpha\mu)(dt) - \int_{\mathbb{R}^s} h_i(t)\ \nu(dt) \right| < \epsilon \right\}$$

for $\epsilon, \alpha > 0$, $\mu \in \mathcal{P}(\mathbb{R}^s)$, $l \in \mathbb{N}$ and $h_1, \dots, h_l \in C_b^0(\mathbb{R}^s)$ form a base of the topology of weak convergence on $\mathcal{P}(\mathbb{R}^s)$.

Proof. Endow the space $\{\alpha\mu \mid \alpha \geq 0,\ \mu \in \mathcal{P}(\mathbb{R}^s)\}$ of nonnegative, finite Borel measures on $\mathbb{R}^s$ with the coarsest topology for which all mappings $\mu \mapsto \int_{\mathbb{R}^s} h(t)\ \mu(dt)$, $h \in C_b^0(\mathbb{R}^s)$ are continuous (as done in [107, Definition A.36]). A base of this topology is given by formula (A.20) in [107] and yields a base of the topology of weak convergence on $\mathcal{P}(\mathbb{R}^s)$, since τ_w^s arises as a subspace topology. □

Apart from the representations via weak* convergence on the dual space of $C_b^0(\mathbb{R}^s)$, a base or the Prokhorov metric, the topology of weak convergences admits various equivalent characterizations that are typically summarized in the so called Portmanteau theorem:

Theorem 2.41 (**Portmanteau theorem**, see e.g. Theorem 2.1 in [49]).
For any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^s)$ and any measure $\mu \in \mathcal{P}(\mathbb{R}^s)$, the following statements are equivalent:

(a) $\mu_l \xrightarrow{w} \mu$.

(b) $\limsup_{l\to\infty} \int_{\mathbb{R}^s} h(t)\ \mu_l(dt) \leq \int_{\mathbb{R}^s} h(t)\ \mu(dt)$ *holds for any upper semicontinuous mapping $h : \mathbb{R}^s \to \mathbb{R}$ that is bounded from above.*

(c) $\liminf_{l\to\infty} \int_{\mathbb{R}^s} h(t)\ \mu_l(dt) \geq \int_{\mathbb{R}^s} h(t)\ \mu(dt)$ *holds for any lower semicontinuous mapping $h : \mathbb{R}^s \to \mathbb{R}$ that is bounded from below.*

(d) $\lim_{l\to\infty} \int_{\mathbb{R}^s} h(t)\ \mu_l(dt) = \int_{\mathbb{R}^s} h(t)\ \mu(dt)$ *holds for any bounded, uniformly continuous mapping $h : \mathbb{R}^s \to \mathbb{R}$.*

(e) $\limsup_{l\to\infty} \mu_l[B] \leq \mu[B]$ *holds for any closed set $B \subseteq \mathbb{R}^s$.*

(f) $\liminf_{l\to\infty} \mu_l[B] \geq \mu[B]$ *holds for any open set $B \subseteq \mathbb{R}^s$.*

(g) $\lim_{l\to\infty} \mu_l[B] = \mu[B]$ *holds whenever $B \in \mathcal{B}(\mathbb{R}^s)$ is a $\mu-$continuity set, i.e. $\mu[\partial B] = 0$, where ∂B denotes the topological boundary of B.*

In many practical applications of stochastic programming, it may seem more natural to work with random vectors instead of Borel probability measures. However, any sequence

$\{Y_l\}_{l\in\mathbb{N}}$ of random vectors $Y_l : \Omega_l \to \mathbb{R}^s$ on probability spaces $(\Omega_l, \mathcal{F}_l, \mathbb{P}_l)$ induces a sequence of Borel probability measures $\{\mu^{Y_l}\}_{l\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^s)$ via

$$\mu^{Y_l} := \mathbb{P}_l \circ Y_l^{-1}. \tag{2.10}$$

For every $l \in \mathbb{N}$, μ^{Y_l} is the law of Y_l (in the sense of section 3 in [49]), which provides a link between weak convergence and the so called convergence in distribution.

Definition 2.42. *The sequence $\{Y_l\}_{l\in\mathbb{N}}$ is said to* ***converge in distribution*** *to a random vector $Y : \Omega \to \mathbb{R}^s$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, written $Y_l \overset{d}{\to} Y$, iff*

$$\lim_{l\to\infty} \mathbb{P}_l[Y_l \in B] = \mathbb{P}[Y \in B]$$

holds whenever $B \in \mathcal{B}(\mathbb{R}^s)$ is such that $\mathbb{P}[Y \in \partial B] = 0$.

Lemma 2.43. *$Y_l \overset{d}{\to} Y$ iff $\mu^{Y_l} \overset{w}{\to} \mu^Y$.*

Proof. Combine the equivalence of (a) and (g) in Theorem 2.41 with the fact that $\mathbb{P}_l[Y_l \in B] = \mu^{Y_l}[B]$ by formula (2.10). □

Remark 2.44. *Convergence in distribution is highly relevant in practice, since it arises from the central limit theorem and its generalizations (see chapter 27 in [48] for a discussion). Nevertheless, the assumption of convergence in distribution is rather weak, as the random vectors in a converging sequence are not even required to be defined on a common probability space. Furthermore, convergence in distribution is implied by convergence in probability (see [48, Theorem 25.2]) and hence in particular implied by almost sure convergence. The converse statements do not hold true in general.*

For any sequence of random vectors $\{Y_l\}_{l\in\mathbb{N}}$ that converges in distribution to Y_1, there exists a sequence of Borel probability measures $\{\mu_l\}_{l\in\mathbb{N}}$ such that μ_l is the law of Y_l and $\mu_l \overset{w}{\to} \mu_1$. An even stronger version of the converse statement is given by the following result:

Theorem 2.45 (Skorohod representation theorem).
For any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^s)$, $\mu_l \overset{w}{\to} \mu_1$ holds iff there exists a sequence of random vectors $\{Y_l\}_{l\in\mathbb{N}}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for any $l \in \mathbb{N}$, μ_l is the law of Y_l and

$$\lim_{l\to\infty} Y_l(\omega) = Y_1(\omega) \quad \forall \omega \in \Omega. \tag{2.11}$$

Proof. Every closed subset of $\mathbb{R}^s$ is separable with respect to the relative topology induced by the Euclidean norm by [7, Lemma 2.9]. Hence, the support of every $\mu \in \mathcal{P}(\mathbb{R}^s)$ is separable and [49, Theorem 6.7] is applicable. Lemma 2.43 completes the proof. □

The probability space in 2.45 may depend on the specific sequence of Borel probability measures that is considered. However, weakening the sure convergence in (2.11) to almost sure convergence, the following result allows represent all weakly converging sequences in $\mathcal{P}(\mathbb{R}^s)$ with random variables on a fixed atomless space:

Theorem 2.46 (**Skorohod representation on a fixed probability space**). *Fix any atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then for any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^s)$, $\mu_l \xrightarrow{w} \mu_1$ holds iff there exists a sequence of random vectors $\{Y_l\}_{l\in\mathbb{N}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that for any $l \in \mathbb{N}$, μ_l is the law of Y_l and*

$$\lim_{l\to\infty} Y_l(\omega) = Y_1(\omega) \quad \textit{for } \mathbb{P} - \textit{almost all } \omega \in \Omega. \tag{2.12}$$

Proof. Combine [42, Theorem 3.2 (i)] with the fact that the support of every $\mu \in \mathcal{P}(\mathbb{R}^s)$ is separable and employ Lemma 2.43. □

Remark 2.47. *In particular,* (2.12) *implies $Y_l \xrightarrow{d} Y_1$ by [48, Theorem 25.2].*

Weak convergence of a sequence in $\mathcal{P}(\mathbb{R}^s)$ translates to weak convergence of the sequence of image measures under a fixed mapping that is continuous almost everywhere with respect to the weak limit of the original sequence:

Theorem 2.48 (**Continuous mapping theorem**, see e.g. Theorem 2.7 in [49]). *Fix any Borel measurable mapping $T : \mathbb{R}^s \to \mathbb{R}^k$ and let $D_T \subseteq \mathbb{R}^s$ denote the set of its discontinuities. Then for any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^s)$ satisfying $\mu_l \xrightarrow{w} \mu \in \mathcal{P}(\mathbb{R}^s)$ and $\mu[D_T] = 0$, it holds that $\mu_l \circ T^{-1} \xrightarrow{w} \mu \circ T^{-1}$.*

The following theorem states that $\mathcal{P}(\mathbb{R}^s)$ endowed with τ_w^s is separable and explicitly points out a countable dense subset:

Theorem 2.49. *$(\mathcal{P}(\mathbb{R}^s), \tau_w^s)$ is a Polish space, i.e. a separable, completely metrizable topological space. In particular, the countable set*

$$\mathcal{D}^s = \{\sum_{i=1}^{l} \alpha_i \delta_{x_i} \mid l \in \mathbb{N},\ 0 \le \alpha_1, \ldots, \alpha_l \in \mathbb{Q},\ \sum_{i=1}^{l} \alpha_i = 1,\ x_1, \ldots, x_l \in \mathbb{Q}^s\} \tag{2.13}$$

is dense in $\mathcal{P}(\mathbb{R}^s)$ with respect to the topology of weak convergence.

Proof. Combine Theorems 15.10 and 15.15 in [7]. □

Remark 2.50. *By the above theorem, every element of $\mathcal{P}(\mathbb{R}^s)$ is the weak limit of a sequence of convex combinations of Dirac measures. Stochastic programming problems are usually more traceable if the support of the underlying measure is finite. In some cases, such problems even admit equivalent reformulations as mixed-integer linear programs (see e.g. [153], [200], [201]). Consider a problem where the dependence of the optimal value and the optimal solution set on the underlying measure is continuous with respect to the topology of weak convergence. By the above result, it is possible to obtain approximate solutions of arbitrary precision by approximating the underlying measure with an element of $\mathcal{D}^s$ and solving the resulting (easier) problem.*

2.4. $\psi-$weak topologies

While continuity of Q with respect to the topology of weak convergence is the ultimate goal of this chapter, the proof in section 2.5 also employs finer topologies that shall be introduced in the present section. Enclosing the topology of weak convergence (see Remark 2.54), $\psi-$weak topologies are topologies on subsets of $\mathcal{P}(\mathbb{R}^s)$ that are defined by generalized moment conditions. In the following, basic properties of $\psi-$weak topologies and their relation to the topology of weak convergence shall be examined. In view of the desired continuity, subsets where the relative topology of weak convergence coincides with a certain relative $\psi-$weak topology are of special interest. A characterization of such sets is given by Lemma 2.66.

$\psi-$weak topologies are induced by so called gauge functions:

Definition 2.51. *A continuous function $\psi : \mathbb{R}^s \to [0, \infty)$ is called a* ***gauge function****, if $\psi \geq 1$ holds outside a compact set.*

Remark 2.52. *In section 2.5, the case where ψ is chosen to be the Euclidean norm to a positive power will be of special interest.*

Definition 2.53. *Let $\psi : \mathbb{R}^s \to [0, \infty)$ be a gauge function and denote by C_s^ψ the linear space of all continuous functions $h : \mathbb{R}^s \to \mathbb{R}$ for which there exists a real constant $c \geq 0$ such that*

$$|h(t)| \leq c(\psi(t) + 1)$$

*holds for any $t \in \mathbb{R}^s$. The $\psi-$**weak topology**, denoted by τ_ψ, is the coarsest topology on*

$\mathcal{M}_s^\psi$ for which all mappings $g_h : \mathcal{M}_s^\psi \to \mathbb{R}$ defined by

$$g_h(\mu) = \int_{\mathbb{R}^s} h(t)\ \mu(dt), \quad h \in C_s^\psi$$

*are continuous. A sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{M}_s^\psi$ is said to **converge $\psi-$weakly** to $\mu \in \mathcal{M}_s^\psi$, written $\mu_l \xrightarrow{\psi} \mu$, iff it converges with respect to τ_ψ.*

Remark 2.54. *Consider the constant gauge function $\psi_1 \equiv 1$ on $\mathbb{R}^s$. Then $C_s^{\psi_1} = C_b^0(\mathbb{R}^s)$ and $\mathcal{M}_s^{\psi_1} = \mathcal{P}(\mathbb{R}^s)$, since $\mu[\mathbb{R}^s] = 1$ holds for any Borel probability measure μ. Consequently, the topology of weak convergence coincides with the ψ_1-weak topology.*

Lemma 2.55. *For any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$ and any real constants $d, e \geq 0$, $(1+e)\psi + d$ is a gauge function and the topologies τ_ψ and $\tau_{(1+e)\psi+d}$ coincide.*

Proof. By definition, it holds that $\mathcal{M}_s^\psi = \mathcal{M}_s^{(1+e)\psi+d}$ and $C_s^\psi = C_s^{(1+e)\psi+d}$. □

Remark 2.56. *Note that [107] and [143] use a more restrictive definition of gauge functions by demanding $\psi \geq 1$ to hold on the whole space. However, for every gauge function ψ in the sense of Definition 2.51, the function $\psi + 1$ is a gauge in function in the sense of [107] that yields the same topology by Lemma 2.55.*

The following characterization of τ_ψ by a base is a generalization of Proposition 2.40:

Proposition 2.57. *For any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$, the sets*

$$N_\epsilon^\psi(\alpha, \mu, h_1, \ldots, h_l) := \bigcap_{i=1}^{l} \left\{ \nu \in \mathcal{M}_s^\psi \mid \left| \int_{\mathbb{R}^s} h_i(t)\ (\alpha\mu)(dt) - \int_{\mathbb{R}^s} h_i(t)\ \nu(dt) \right| < \epsilon \right\}$$

for $\epsilon, \alpha > 0$, $\mu \in \mathcal{M}_s^\psi$, $l \in \mathbb{N}$ and $h_1, \ldots, h_l \in C_s^\psi$ form a base of the $\psi-$weak topology on $\mathcal{M}_s^\psi$.

Proof. Endow the space $\{\alpha\mu \mid \alpha \geq 0,\ \mu \in \mathcal{M}_s^\psi\}$ with the coarsest topology for which all mappings $\mu \mapsto \int_{\mathbb{R}^s} h(t)\ \mu(dt)$, $h \in C_s^\psi$ are continuous (as done in [107, Definition A.44]). τ_ψ arises as a subspace topology, hence the base given on page 502 in [107] yields a base of the $\psi-$topology on $\mathcal{M}_s^\psi$. □

Remark 2.58. *For any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$, the mapping*

$$\Psi : \{\alpha\mu \mid \alpha \geq 0,\ \mu \in \mathcal{M}_s^\psi\} \to \{\alpha\mu \mid \alpha \geq 0,\ \mu \in \mathcal{P}(\mathbb{R}^s)\}$$

given by

$$(\Psi(\mu))(dt) := \psi(t)\mu(dt)$$

yields a homeomorphism between the topological spaces considered in the proofs of Propositions 2.40 and 2.57 (see page 502 in [107]).

The above remark allows to generalize Theorem 2.49:

Theorem 2.59. *For any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$, $(\mathcal{M}_s^\psi, \tau_\psi)$ is a Polish space and the countable set $\mathcal{D}^s$ defined in* (2.13) *is dense in $\mathcal{M}_s^\psi$ with respect to τ_ψ.*

Proof. Combine [107, Theorem A.45] with the fact that closed subspaces of Polish spaces are Polish spaces by [71, Proposition 8.1.2]). □

Remark 2.60. *In particular, $\psi-$weak topologies are metrizable and the notions of continuity and sequential continuity coincide.*

The following characterization of the relation between weakly and $\psi-$weakly converging sequences is a combination of [69, Lemma 4.1] and [143, Lemma 3.4]:

Lemma 2.61. *For any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$ and any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{M}_s^\psi$, the following statements are equivalent:*

(a) $\mu_l \overset{\psi}{\to} \mu_1$.

(b) $\mu_l \overset{w}{\to} \mu_1$ *and* $\lim_{l\to\infty} \int_{\mathbb{R}^s} \psi(t)\ \mu_l(dt) = \int_{\mathbb{R}^s} \psi(t)\ \mu_1(dt)$.

(c) $\lim_{l\to\infty} \int_{\mathbb{R}^s} h(t)\ \mu_l(dt) = \int_{\mathbb{R}^s} h(t)\ \mu_1(dt)$ *holds for any continuous function* $h : \mathbb{R}^s \to \mathbb{R}$ *with compact support and for* $h = \psi$.

(d) $\mu_l \overset{w}{\to} \mu_1$ *and*

$$\lim_{a\to\infty} \sup_{l\in\mathbb{N}} \int_{\mathbb{R}^s} \psi(t) \cdot \chi_{(a,\infty)}(\psi(t))\ \mu_l(dt) = 0.$$

Proof. ((a) ⇒ (b)) follows directly from Definition 2.53 and the fact that $\psi \in C_s^\psi$.

((b) ⇒ (c)): Let $h : \mathbb{R}^s \to \mathbb{R}$ be continuous and have a compact support. Then $h \in C_b^0(\mathbb{R}^s)$ and (c) follows from Definition 2.32.

((c) ⇒ (a)): Let Ψ denote the homeomorphism defined in Remark 2.58. By the equivalence of (i) and (iii) in [28, Theorem 30.8], (c) implies $\Psi(\mu_n) \overset{w}{\to} \Psi(\mu_1)$. Invoking Remark 2.58, the latter yields $\mu_n \overset{\psi}{\to} \mu_1$, i.e. (a).

((a) ⇔ (d)) is a direct conclusion from [206, Theorem 2.20]. □

The equivalence of (a) and (b) in the above Lemma allows to explicitly point out a metric that generates τ_ψ:

Corollary 2.62. *For any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$, the $\psi-$weak topology on $\mathcal{M}_s^\psi$ is generated by the metric $d_\psi : \mathcal{M}_s^\psi \times \mathcal{M}_s^\psi \to \mathbb{R}$ defined by*

$$d_\psi(\mu, \nu) := \pi(\mu, \nu) + \left| \int_{\mathbb{R}^s} \psi(t)\ \mu(dt) - \int_{\mathbb{R}^s} \psi(t)\ \nu(dt) \right|.$$

While the above corollary applies to any gauge function, the case where ψ is a positive power of the Euclidean norm is of special interest in view of section 2.5. The following result characterizes $\tau_{\|\cdot\|^q}$ for $q \geq 1$:

Proposition 2.63. *For any $q \geq 1$, the $\| \cdot \|^q-$weak topology on $\mathcal{M}_s^q$ is generated by the Wasserstein metric $d_{W,s,q} : \mathcal{M}_s^q \times \mathcal{M}_s^q \to \mathbb{R}$ of order q:*

$$d_{W,s,q}(\mu, \nu) := \inf \left\{ \left(\int_{\mathbb{R}^s \times \mathbb{R}^s} \|x - y\|^q\ \sigma(dx, dy) \right)^{\frac{1}{q}}\ \mid \sigma \in \mathbb{M}(\mu, \nu) \right\},$$

where $\mathbb{M}(\mu, \nu)$ denotes the set of all Borel probability measures on $\mathbb{R}^s \times \mathbb{R}^s$ with μ as the first $s-$dimensional marginal and ν as the second one. For $s = 1$, $\tau_{|\cdot|^q}$ is also generated by the $q-$th order Fortet-Mourier metric $d_{FM,q} : \mathcal{M}_1^q \times \mathcal{M}_1^q \to \mathbb{R}$. The latter is given by

$$d_{FM,q}(\mu, \nu) := \inf \left\{ \int_{\mathbb{R} \times \mathbb{R}} |x - y| \max\{1, |x|^{q-1}, |y|^{q-1}\}\ \sigma(dx, dy) \mid \sigma \in \mathbb{L}(\mu, \nu) \right\},$$

where $\mathbb{L}(\mu, \nu)$ denotes the set of all finite Borel measures on $\mathbb{R} \times \mathbb{R}$ satisfying

$$\sigma[A \times \mathbb{R}] - \sigma[\mathbb{R} \times A] = \mu[A] - \nu[A] \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

Proof. Combine [169, Theorem 6.3.1] with Lemma 2.61. □

For details on the metrics in the above proposition, refer to [169]. By the equivalence of (a) and (b) in Lemma 2.61, $\psi-$weak convergence implies weak convergence. The converse statement does not hold true in general:

Example 2.64. *Consider the sequence $\{\mu_l\}_{l \in \mathbb{N}} \subseteq \mathcal{P}(\mathbb{R})$ defined by $\mu_l := (1 - \frac{1}{l})\delta_0 + \frac{1}{l}\delta_l$ (a similar sequence is used in [107, Example A.43]). By*

$$\int_{\mathbb{R}} |t|\ \mu_l(dt) = (1 - \frac{1}{l})|0| + \frac{1}{l}|l| = 1 < \infty \quad \forall\, l \in \mathbb{N}$$

it holds that $\{\mu_l\}_{l\in\mathbb{N}} \in \mathcal{M}_1^1$. *Fix an arbitrary bounded and continuous function* $h : \mathbb{R} \to \mathbb{R}$. *Then*

$$\lim_{l\to\infty} \int_{\mathbb{R}} h(t)\ \mu_l(dt) = \lim_{l\to\infty} (1 - \frac{1}{l})h(0) + \frac{1}{l}h(l) = h(0) = \int_{\mathbb{R}} h(t)\ \delta_0(dt)$$

and hence $\mu_n \overset{w}{\to} \delta_0$. *However, since*

$$\lim_{l\to\infty} \int_{\mathbb{R}} |t|\ \mu_l(dt) = 1 \neq 0 = \int_{\mathbb{R}} |t|\ \delta_0(dt),$$

the sequence $\{\mu_l\}_{l\in\mathbb{N}}$ *does not converge with respect to the* $|\cdot|-$*weak topology by Lemma 2.61.*

Nevertheless, there are subspaces of $\mathcal{M}_s^\psi$ on which the relative topology of weak convergence coincides with the relative $\psi-$weak topology.

Definition 2.65. *Let* $\psi : \mathbb{R}^s \to [0,\infty)$ *be a gauge function. A set* $\mathcal{M} \subseteq \mathcal{M}_s^\psi$ *is said to be* ***locally uniformly*** ψ***−integrating*** *iff for any* $\mu \in \mathcal{M}$ *and any* $\epsilon > 0$ *there exists some open neighborhood* $\mathcal{N}$ *of* μ *with respect to the topology of weak convergence such that*

$$\lim_{a\to\infty} \sup_{\nu\in\mathcal{M}\cap\mathcal{N}} \int_{\mathbb{R}^s} \psi(t) \cdot \chi_{(a,\infty)}(\psi(t))\ \nu(dt) \leq \epsilon.$$

The relevance of locally uniformly $\psi-$integrating subsets in the present context is given by the following result:

Lemma 2.66 (**[218, Lemma 3.4]**).
For any gauge function $\psi : \mathbb{R}^s \to [0,\infty)$ *and any* $\mathcal{M} \subseteq \mathcal{M}_s^\psi$, *the following statements are equivalent:*

(a) The relative $\psi-$*weak topology on* $\mathcal{M}$ *coincides with the relative topology of weak convergence on* $\mathcal{M}$.

(b) $\mathcal{M}$ *is locally uniformly* $\psi-$*integrating.*

A detailed discussion of locally uniformly $\psi-$integrating subsets and generalizations is provided in [142], where various equivalent characterizations are established. The following result arises from the application of [142, Lemma 3.1] to a constant sequence of gauge functions and yields a whole class of locally uniformly $\psi-$integrating sets:

Proposition 2.67. *For any gauge function* $\psi : \mathbb{R}^s \to [0,\infty)$ *and any* $\mathcal{M} \subseteq \mathcal{M}_s^\psi$, *the following statements are equivalent:*

(a) *$\mathcal{M}$ is locally uniformly $\psi-$integrating and relatively compact for the topology of weak convergence.*

(b) *$\mathcal{M}$ is relatively compact for the $\psi-$weak topology.*

Proof. ((a) $\Rightarrow$ (b)): By Lemma 2.66, the locally uniformly $\psi-$integrating set $\mathcal{M}$ is relatively compact for the topology of weak convergence iff it is relatively compact for the $\psi-$weak topology.

((b) $\Rightarrow$ (a)): Consider a sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{M}$ that converges weakly to $\mu \in \mathcal{P}(\mathbb{R}^s)$. By the relative compactness for the $\psi-$weak topology, there exists a subsequence of $\{\mu_l\}_{l\in\mathbb{N}}$ that converges $\psi-$weakly to some $\nu \in \mathcal{M}_s^\psi$. The equivalence of (a) and (b) in Lemma 2.61 implies $\nu = \mu$ and hence $\mu_l \stackrel{\psi}{\to} \mu$. Consequently, $\mathcal{M}$ is locally uniformly $\psi-$integrating by Theorem 2.59 and Lemma 2.66. As above, this implies that $\mathcal{M}$ is relatively compact for the topology of weak convergence. □

The classical Prokhorov theorem (see e.g. section 5 of chapter 1 in [49]) characterizes subsets of $\mathcal{P}(\mathbb{R}^s)$ that are weakly compact for the topology of weak convergence via tightness. In view of Proposition 2.67 and Lemma 2.66 the following generalization for $\psi-$weak topologies is of special interest:

Theorem 2.68 (**[142, Lemma 5.1]**).
For any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$ and any set $\mathcal{M} \subseteq \mathcal{M}_s^\psi$, the following statements are equivalent:

(a) *$\mathcal{M}$ is relatively compact for the $\psi-$weak topology.*

(b) *For any $\epsilon > 0$, there exists a compact set $K \subset \mathbb{R}^s$ such that*

$$\sup_{\mu\in\mathcal{M}} \int_{\mathbb{R}^s\setminus K} \psi(t)\ \mu(dt) \leq \epsilon.$$

(c) *There exists a measurable function $\kappa : \mathbb{R}^s \to [0, \infty)$ such that*

$$\sup_{\mu\in\mathcal{M}} \int_{\mathbb{R}^s} \kappa(t)\ \mu(dt) < \infty.$$

and the set $\{t \in \mathbb{R}^s \mid \kappa(t) \leq l\psi(t)\}$ is compact for any $l \in \mathbb{N}$.

The following result provides examples for sets that are relatively compact for the $\psi-$weak topology:

Lemma 2.69. *Let $\psi : \mathbb{R}^s \to [0, \infty)$ be a coercive gauge function, i.e. $\psi(t) \to \infty$ whenever $\|t\| \to \infty$. Then for any constants $K \geq 0$ and $q > 1$, the set*

$$U_s^{\psi}(K, q) := \{\mu \in \mathcal{P}(\mathbb{R}^s) \mid \int_{\mathbb{R}^s} \psi(t)^q \ \mu(dt) \leq K\}$$

is relatively compact for the $\psi-$weak topology.

Proof. Fix an arbitrary constant $\epsilon > 0$ and set $r := (\frac{K}{\epsilon})^{\frac{1}{q-1}}$. By the coercivity of ψ, there exists a finite constant $R \geq 0$ such that $\psi(t) \geq r$ whenever $\|t\| \geq R$. Let $\overline{B_R(0)}$ denote the closed $\| \cdot \|-$ball of radius R centered at 0. For any $\mu \in U_s^{\psi}(K, q)$, it holds that

$$\int_{\mathbb{R}^s \setminus \overline{B_R(0)}} \psi(t)^q \ \mu(dt) = \int_{\mathbb{R}^s \setminus \overline{B_R(0)}} \psi(t)^{q-1} \psi(t) \ \mu(dt) \geq r^{q-1} \int_{\mathbb{R}^s \setminus \overline{B_R(0)}} \psi(t) \ \mu(dt)$$

and hence

$$\int_{\mathbb{R}^s \setminus \overline{B_R(0)}} \psi(t) \ \mu(dt) \leq \frac{1}{r^{q-1}} \int_{\mathbb{R}^s \setminus \overline{B_R(0)}} \psi(t)^q \ \mu(dt) \leq \frac{K}{r^{q-1}} = \epsilon.$$

Since $\overline{B_R(0)}$ is compact, the equivalence of (a) and (b) in Theorem 2.68 yields that $U_s^{\psi}(K, q)$ is relatively compact for the $\psi-$weak topology. □

Remark 2.70. *Subsets of Borel probability measures having uniformly bounded moments of order $q' > 1$ are known to be useful in the context of stability analysis in stochastic programming (see e.g. [153], [196], [198], [200], [201]). Fix any constant $0 < q < q'$. By Lemma 2.69, those sets are relatively compact for the $\| \cdot \|^q-$weak topology and hence locally uniformly $\| \cdot \|^q-$integrating by Theorem 2.68.*

Remark 2.71. *In general, the statement of Lemma 2.69 does not hold if ψ is not coercive: Consider the constant gauge function $\psi_1 : \mathbb{R} \to \mathbb{R}$ given by $\psi_1 \equiv 1$ and fix any constants $K, q > 1$. Then the ψ_1-weak topology coincides with the topology of weak convergence and $U_1^{\psi_1}(K, q)$ is equal to $\mathcal{P}(\mathbb{R})$. Consequently, $U_1^{\psi_1}(K, q)$ is not relatively compact for the ψ_1-weak topology.*

Remark 2.72. *The statement of Lemma 2.69 does not hold if $q \leq 1$. In general, sets of the form $U_s^{\psi}(K, 1)$ are not locally uniformly $\psi-$integrating and hence not relatively compact for the $\psi-$weak topology, even if ψ is coercive: Consider the sequence $\{\mu_l\}_{l \in \mathbb{N}} \subseteq \mathcal{M}_1^1$ given by $\mu_l := (1 - \frac{1}{l})\delta_0 + \frac{1}{l}\delta_l$. Then $\{\mu_l\}_{l \in \mathbb{N}} \subseteq U_1^{|\cdot|}(1, 1)$ holds by the calculation in Example 2.64. Furthermore, the sequence is weakly convergent but not $| \cdot |-$weakly convergent. Consequently, $U_1^{|\cdot|}(1, 1)$ is not locally uniformly $| \cdot |-$integrating by Lemma 2.66.*

The fixed space version of the Skorohod representation theorem (Theorem 2.46) states that weak convergence of probability measures can be translated to almost sure convergence of random vectors on a fixed atomless probability space. Using Lemma 2.61, a similar result can be proven for $\psi-$weak converging sequences. While it is possible to formulate such a result whenever the gauge function is a finite Young function (see [141, Theorem 3.5] for details on the general result in a setting involving Orlicz spaces), only the case where $\psi = |\cdot|^p$ with $p \geq 1$ is needed in section 2.5:

Theorem 2.73 (A Skorohod representation for $|\cdot|^p-$weak convergence). *Fix an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a constant $p \geq 1$. For any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{M}_1^p$, $\mu_l \overset{|\cdot|^p}{\to} \mu_1$ holds iff there exists a sequence $\{Y_l\}_{l\in\mathbb{N}} \subseteq L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that for any $l \in \mathbb{N}$, μ_l is the law of Y_l and*

$$\lim_{l\to\infty} \|Y_l - Y_1\|_p = 0, \tag{2.14}$$

where $\|\cdot\|_p$ denotes the L^p-norm.

Proof. ($\Rightarrow$): By the equivalence of (a) and (b) in Lemma 2.61, $\mu_l \overset{|\cdot|^p}{\to} \mu_1$ implies $\mu_l \overset{w}{\to} \mu_1$. Hence, the fixed space version of the classical Skorohod representation theorem (Theorem 2.46) yields the existence of a sequence of random variables $\{Y_l\}_{l\in\mathbb{N}} \subseteq L^0(\Omega, \mathcal{F}, \mathbb{P})$ such that for any $l \in \mathbb{N}$, μ_l is the law of Y_l and $Y_l \to Y_1$ $\mathbb{P}-$almost surely. By the change-of-variable formula (Theorem A.1),

$$\|Y_l\|_p^p = \int_\Omega |Y_l(\omega)|^p \ \mathbb{P}(d\omega) = \int_\mathbb{R} |t|^p \ (\mathbb{P} \circ Y_l^{-1})(dt) = \int_\mathbb{R} |t|^p \ \mu_l(dt) \tag{2.15}$$

holds for any $l \in \mathbb{N}$. Since $\mu_l \in \mathcal{M}_1^p$, the last integral in (2.15) is finite. Consequently, $\{Y_l\}_{l\in\mathbb{N}} \subseteq L^p(\Omega, \mathcal{F}, \mathbb{P})$. Again by Lemma 2.61,

$$\lim_{l\to\infty} \|Y_l\|_p^p = \lim_{l\to\infty} \int_\mathbb{R} |t|^p \ \mu_l(dt) = \int_\mathbb{R} |t|^p \ \mu_1(dt) = \|Y_1\|_p^p$$

and hence $\lim_{l\to\infty} \|Y_l\|_p = \|Y_1\|_p$. Thus, (2.14) holds by the equivalence of (a) and (b) in Vitali's theorem (Theorem A.2), since $\mathbb{P}-$almost sure convergence implies convergence in probability.

($\Leftarrow$): By Vitali's theorem, (2.14) implies that the sequence $\{Y_l\}_{l\in\mathbb{N}}$ converges to Y_1 in probability. In particular, it convergences in distribution by [48, Theorem 25.2] and since μ_l is the law of Y_l for any $l \in \mathbb{N}$, the latter yields $\mu_l \overset{w}{\to} \mu_1$ by Lemma 2.43. Furthermore, from the equivalence of (a) and (b) in Vitali's theorem it follows that

$$\lim_{l\to\infty}\int_{\mathbb{R}} |t|^p\ \mu_l(dt) = \lim_{l\to\infty}\|Y_l\|_p^p = \|Y_1\|_p^p = \int_{\mathbb{R}} |t|^p\ \mu_1(dt).$$

Hence, $\mu_l \overset{|\cdot|^p}{\to} \mu_1$ holds by the equivalence (a) and (b) in Lemma 2.61. □

2.5. Proving weak continuity

Let $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ denote the product topology of the standard topology on $\mathbb{R}^n$ and the $\|\cdot\|^{\gamma p}$-weak topology on $\mathcal{M}_s^{\gamma p}$. The main result in this section is the continuity of Q with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ under assumptions $[\mathbb{A}_f]$, $[\mathbb{A}_\rho]$ and a condition originating from the continuous mapping theorem. The proof proceeds in two steps:

1. Prove that the mapping $\theta : \mathbb{R}^n \times \mathcal{M}_s^{\gamma p} \to \mathcal{M}_1^p$ given by

$$\theta(x,\mu) := (\delta_x \otimes \mu) \circ f^{-1} \tag{2.16}$$

   is continuous with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ and $\tau_{|\cdot|^p}$ (Lemma 2.74).

2. Prove that the mapping $\mathcal{R}_\rho : \mathcal{M}_1^p \to \mathbb{R}$ defined in (2.3) is continuous with respect to $\tau_{|\cdot|^p}$ (Lemma 2.81).

The continuity of Q with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ is then implied by the fact that $Q = \mathcal{R}_\rho \circ \theta$. Consequently, any restriction of Q to an appropriate subset is continuous with respect to $\tau_{\mathbb{R}^n} \otimes \tau_w^s$ by Lemma 2.66. The special case where the entering measure is absolutely continuous with respect to the Lebesgue measure is addressed in Corollary 2.86.

Lemma 2.74. *Assume $[\mathbb{A}_f]$ and fix a constant $p \geq 1$. Let $(x,\mu) \in \mathbb{R}^n \times \mathcal{M}_s^{\gamma p}$ be such that $(\delta_x \otimes \mu)[D_f] = 0$, where $D_f \subseteq \mathbb{R}^n \times \mathbb{R}^s$ denotes the set of discontinuities of f. Then $\theta : \mathbb{R}^n \times \mathcal{M}_s^{\gamma p} \to \mathcal{M}_1^p$ is continuous at (x,μ) with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ and $\tau_{|\cdot|^p}$.*

Proof. Consider any sequence $\{(x_l,\mu_l)\}_{l\in\mathbb{N}} \subseteq \mathbb{R}^n \times \mathcal{M}_s^{\gamma p}$ that converges to (x,μ) with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$.

Then $x_l \to x$ and hence $\delta_{x_l} \overset{w}{\to} \delta_x$. Furthermore, $\mu_l \overset{\|\cdot\|^{\gamma p}}{\to} \mu$ implies $\mu_l \overset{w}{\to} \mu$ by the equivalence of (a) and (b) in Lemma 2.61. Since all involved spaces are separable,

$$\delta_{x_l} \otimes \mu_l \overset{w}{\to} \delta_x \otimes \mu$$

holds by [49, Theorem 2.8]. f is Borel measurable by $[\mathbb{A}_f]$ and $(\delta_x \otimes \mu)[D_f] = 0$ allows to apply the continuous mapping theorem (Theorem 2.48):

$$(\delta_{x_l} \otimes \mu_l) \circ f^{-1} \overset{w}{\to} (\delta_x \otimes \mu) \circ f^{-1}. \tag{2.17}$$

Furthermore, the proof of Lemma 2.10 yields

$$|f(x_l, z)|^p \leq \eta(x_l)^p 2^p (\|z\|^{p\gamma} + 1)$$

for any $l \in \mathbb{N}$ and $z \in \mathbb{R}^s$. The constant $C := \sup_{l\in\mathbb{N}} \eta(x_l)$ is finite by the local boundedness of η. Without loss of generality, assume $C > 1$ and note that the indicator function $\chi_{(a,\infty)}(\cdot)$ is nondecreasing for any $a \in \mathbb{R}$. Thus,

$$\begin{aligned}
&\sup_{l\in\mathbb{N}} \int_{\mathbb{R}^n\times\mathbb{R}^s} |f(x,z)|^p \cdot \chi_{(a,\infty)}(|f(x,z)|^p)\ (\delta_{x_l} \otimes \mu_l)(d(x,z)) \\
\leq\ &\sup_{l\in\mathbb{N}} \int_{\mathbb{R}^n\times\mathbb{R}^s} \eta(x)^p 2^p (\|z\|^{p\gamma} + 1) \cdot \chi_{(a,\infty)}(\eta(x)^p 2^p (\|z\|^{p\gamma} + 1))\ (\delta_{x_l} \otimes \mu_l)(d(x,z)) \\
=\ &\sup_{l\in\mathbb{N}} \int_{\mathbb{R}^s} \eta(x_l)^p 2^p (\|z\|^{p\gamma} + 1) \cdot \chi_{(a,\infty)}(\eta(x_l)^p 2^p (\|z\|^{p\gamma} + 1))\ \mu_l(dz) \\
\leq\ &\sup_{l\in\mathbb{N}} \int_{\mathbb{R}^s} C^p 2^p (\|z\|^{p\gamma} + 1) \cdot \chi_{(a,\infty)}(C^p 2^p (\|z\|^{p\gamma} + 1))\ \mu_l(dz)
\end{aligned}$$

holds for any $a \in \mathbb{R}$ by the Fubini-Tonelli theorem (see e.g. [7, Theorem 11.27]). By Lemma 2.55, $\mu_l \overset{\|\cdot\|^{\gamma p}}{\to} \mu$ implies $\mu_l \overset{C^p 2^p (\|\cdot\|^{\gamma p}+1)}{\to} \mu$ and the equivalence of (a) and (d) in Lemma 2.61 yields

$$\lim_{a\to\infty} \sup_{l\in\mathbb{N}} \int_{\mathbb{R}^s} C^p 2^p (\|z\|^{p\gamma} + 1) \cdot \chi_{(a,\infty)}(C^p 2^p (\|z\|^{p\gamma} + 1))\ \mu_l(dz) = 0.$$

By the change-of-variable formula (Theorem A.1), it holds that

$$\begin{aligned}
&\lim_{a\to\infty} \sup_{l\in\mathbb{N}} \int_{\mathbb{R}} |t|^p \cdot \chi_{(a,\infty)}(|t|^p)\ ((\delta_{x_l} \otimes \mu_l) \circ f^{-1})(dt) \\
=\ &\lim_{a\to\infty} \sup_{l\in\mathbb{N}} \int_{\mathbb{R}^n\times\mathbb{R}^s} |f(x,z)|^p \cdot \chi_{(a,\infty)}(|f(x,z)|^p)\ (\delta_{x_l} \otimes \mu_l)(d(x,z)) = 0.
\end{aligned}$$

Combined with (2.17) the latter implies

$$\theta(x_l, \mu_l) = (\delta_{x_l} \otimes \mu_l) \circ f^{-1} \overset{|\cdot|^p}{\to} (\delta_x \otimes \mu) \circ f^{-1} = \theta(x, \mu),$$

i.e. the sequential continuity of θ with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ and $\tau_{|\cdot|^p}$. Since the involved topologies are metrizable by Theorem 2.59, that yields the desired continuity of θ with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ and $\tau_{|\cdot|^p}$. □

The continuity of R_ρ is proven using a result from the theory of Banach lattices (see chapter 9 in [7] for an introduction and [155] or [188] for more details). As Banach lattices are special Riesz spaces, it is convenient to introduce those spaces first:

Definition 2.75. *A **Riesz space** $(E, \leq_E)$ is a real vector space E endowed with a partial order $\leq_E$ such that the following statements hold true for any $x, y, z \in E$:*

(a) $x \leq_E y$ implies $x + z \leq_E y + z$ (translation invariance).

(b) $x \leq_E y$ implies $\alpha x \leq_E \alpha y$ for any nonnegative $\alpha \in \mathbb{R}$ (positive homogeneity).

(c) The set $\{x, y\}$ has a greatest lower bound $\inf_{\leq_E}\{x, y\} \in E$ and a least upper bound $\sup_{\leq_E}\{x, y\} \in E$ with respect to $\leq_E$.

Riesz spaces are named after Frigyes Riesz who first defined them in 1928 (see [171]). The theory of Riesz spaces (see chapter 8 in [7] for a basic introduction and refer to [152] and [217] for a detailed discussion) has many applications in economics (see [8]) and measure theory: For example, the Hahn decomposition theorem (see [48, Theorem 32.1]) and the Radon-Nikodym theorem (see [48, Theorem 32.2]) arise as special cases of general results on Riesz spaces.

A Banach lattice is a Riesz space endowed with a complete lattice norm:

Definition 2.76. *A **Banach lattice** $(E, \|\cdot\|_E, \leq_E)$ is a real Banach space $(E, \|\cdot\|_E)$ endowed with a partial order $\leq_E$ such that $(E, \leq_E)$ is a Riesz space and the following holds for any $x, y \in E$:*

$$|x|_{\leq_E} \leq_E |y|_{\leq_E} \Rightarrow \|x\|_E \leq \|y\|_E,$$

where $|z|_{\leq_E} := \sup_{\leq_E}\{\sup_{\leq_E}\{0, z\}, \sup_{\leq_E}\{0, -z\}\}$ denotes the absolute value of $z \in E$ with respect to $\leq_E$.

Remark 2.77. *The most familiar example of a Banach lattice is obtained by equipping $\mathbb{R}^s$ with the Euclidean norm and the order where $(x_1, \ldots, x_s) \geq (y_1, \ldots, y_s)$ whenever $x_i \geq y_i$ holds for all $i = 1, \ldots s$. Another example is given by the infinite dimensional space $C_b^0(\mathbb{R}^s)$ endowed with the supremum norm as defined in (2.8) and the order where $g \geq h$ iff $g(t) \geq h(t)$ holds for all $t \in \mathbb{R}^s$ (see section 9.1 in [7]).*

The proof of continuity of $\mathcal{R}_\rho$ relies heavily on the fact that the underlying L^p-space is a Banach lattice:

Theorem 2.78 (**Riesz-Fischer Theorem**, see e.g. Theorem 13.5 in [7])**.**
For any $1 \leq p \leq \infty$, $L^p(\Omega, \mathcal{F}, \mathbb{P})$ endowed with the L^p-norm and the $\mathbb{P}-$almost sure partial order is a Banach lattice.

Every convex function $g : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous on all bounded subsets of $\mathbb{R}^k$ (Lemma 3.23) and hence in particular continuous on $\mathbb{R}^k$. This does no longer hold if $\mathbb{R}^k$ is replaced with an infinite dimensional Banach space E, since it is always possible to construct a linear, discontinuous functional using a Hamel basis of E (see e.g. [121, Example 4.2]). However, for Banach lattices, continuity results are available and will play an important role in the proof of Lemma 2.81. The following Theorem is a generalization of [193, Proposition 3.1]:

Theorem 2.79 ([**65, Theorem 4.1**])**.**
Let E be a Banach lattice and $\varrho : E \to (-\infty, \infty]$ a nondecreasing, convex function. Let

$$\text{dom } \varrho := \{e \in E \mid \varrho(e) < \infty\}$$

denote the domain of ϱ. Then the following statements hold for any $e \in \text{int}(\text{dom } \varrho)$:

(a) There exists a neighborhood of e on which ϱ is Lipschitz continuous with respect to the norm on E.

(b) ϱ is subdifferentiable at e.

(c) $\varrho(e) = \varrho^{**}(e)$, where ϱ^{**} denotes the biconjugate of ϱ (see section 11 A in [177]).

The corollary below is an immediate conclusion from Theorem 2.79.

Corollary 2.80. *Every finite, nondecreasing, convex functional on a Banach lattice is continuous.*

The above results can be applied to ρ and the Skorohod representation theorem for $|\cdot|^p-$weak convergence (Theorem 2.73) allows to conclude that $\mathcal{R}_\rho$ is continuous with respect to $\tau_{|\cdot|^p}$.

Lemma 2.81. *Under assumption* $[\mathbb{A}_\rho]$*, the mapping $\mathcal{R}_\rho$ defined in* (2.3) *is continuous with respect to the $|\cdot|^p-$weak topology.*

Proof. The space $L^p(\Omega, \mathcal{F}, \mathbb{P})$ from assumption $[\mathbb{A}_\rho]$ endowed with the L^p-norm and the $\mathbb{P}-$almost sure partial order is a Banach lattice by the Riesz-Fischer Theorem (Theorem 2.78). By $[\mathbb{A}_\rho]$, ρ is real-valued, convex and nondecreasing with respect to the $\mathbb{P}-$almost sure partial order. Consequently, Corollary 2.80 yields the continuity of ρ with respect to the L^p-norm.

Consider any sequence $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{M}_1^p$ satisfying $\mu_l \overset{|\cdot|^p}{\to} \mu_1$. By assumption $[\mathbb{A}_\rho]$, $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless and $p \geq 1$ holds. Hence, the Skorohod representation theorem for $|\cdot|^p-$weak convergence (Theorem 2.73) yields the existence of a sequence $\{Y_l\}_{l\in\mathbb{N}} \subseteq L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that for any $l \in \mathbb{N}$, μ_l is the law of Y_l and $\lim_{l\to\infty} \|Y_l - Y_1\|_p = 0$.

Since ρ is law-invariant by assumption $[\mathbb{A}_\rho]$, it holds that $\{\mathcal{R}(\mu_l)\}_{l\in\mathbb{N}} = \{\rho(Y_l)\}_{l\in\mathbb{N}}$ and the continuity of ρ with respect to the L^p-norm implies

$$\mathcal{R}_\rho(\mu_l) = \rho(Y_l) \to \rho(Y_1) = \mathcal{R}_\rho(\mu_1).$$

Hence, $\mathcal{R}_\rho$ is sequentially continuous with respect to $\tau_{|\cdot|^p}$. Since $\tau_{|\cdot|^p}$ is metrizable by Theorem 2.59, that entails the desired continuity of $\mathcal{R}_\rho$ with respect to the $|\cdot|^p-$weak topology. □

As stated at the beginning of the present section, the continuity of Q with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ is an immediate conclusion from Lemma 2.74 and Lemma 2.81.

Theorem 2.82. *Assume $[\mathbb{A}_f]$ and $[\mathbb{A}_\rho]$. Then Q is continuous with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ at any $(x, \mu) \in \mathbb{R}^n \times \mathcal{M}_s^{\gamma p}$ satisfying $(\delta_x \otimes \mu)[D_f] = 0$.*

Proof. Combine Lemma 2.74 and Lemma 2.81 with the fact that $Q = \mathcal{R}_\rho \circ \theta$. □

The inalienability of the assumption $(\delta_x \otimes \mu)[D_f] = 0$ is demonstrated by the following example:

Example 2.83. *Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given by $f(x, z) := \chi_{\{0\}}(z) \cdot (x^2 + \lambda)$, where $\lambda > 0$ is a fixed constant. Chose ρ to be the expectation, then Assumptions $[\mathbb{A}_f]$ and $[\mathbb{A}_\rho]$ are fulfilled for any exponent $\gamma > 0$ and $p = 1$. However, the mapping $Q : \mathbb{R} \times \mathcal{M}_1^1 \to \mathbb{R}$, $Q(x, \mu) = \int_{\mathbb{R}\times\mathbb{R}} f(x', z)\ (\delta_x \otimes \mu)(d(x', z))$ is not continuous with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{|\cdot|}$: Consider the sequence $\{\delta_{\frac{1}{l}}\}_{l\in\mathbb{N}} \subseteq \mathcal{M}_1^1$. For any continuous function $h : \mathbb{R} \to \mathbb{R}$, it holds that*

$$\lim_{l\to\infty} \int_{\mathbb{R}} h(t)\ \delta_{\frac{1}{l}}(dt) = \lim_{n\to\infty} h(\frac{1}{l}) = h(0) = \int_{\mathbb{R}} h(t)\ \delta_0(dt)$$

and hence $\delta_{\frac{1}{l}} \overset{|\cdot|}{\to} \delta_0$*. However, for any* $x \in \mathbb{R}$*, one obtains*

$$\lim_{l\to\infty} Q(x, \delta_{\frac{1}{l}}) = \lim_{l\to\infty} \int_{\mathbb{R}\times\mathbb{R}} f(x', z)\ (\delta_x \otimes \delta_{\frac{1}{l}})(d(x', z)) = \lim_{l\to\infty} \chi_{\{0\}}(\frac{1}{l}) \cdot (x^2 + \lambda) = 0,$$

while $Q(x, \delta_0) = x^2 + \lambda > 0$*. This is due to the fact that* $(\delta_x \otimes \delta_0)[D_f] = 1 \neq 0$.

The continuity of Q with respect to $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ directly translates to continuity with respect to $\tau_{\mathbb{R}^n} \otimes \tau_w^s$ whenever Q is restricted to an appropriate subset of $\mathbb{R}^n \times \mathcal{M}_s^{\gamma p}$:

Theorem 2.84. *Assume* $[\mathbb{A}_f]$, $[\mathbb{A}_\rho]$ *and let* $Q|_{\mathbb{R}^n\times\mathcal{M}}$ *denote the restriction of* Q *to the Cartesian product of* $\mathbb{R}^n$ *and some locally uniformly* $\|\cdot\|^{\gamma p}$*−integrating set* $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$*. Then* $Q|_{\mathbb{R}^n\times\mathcal{M}}$ *is continuous with respect to the product topology of the standard topology on* $\mathbb{R}^n$ *and the relative topology of weak convergence on* $\mathcal{M}$ *at any* $(x, \mu) \in \mathbb{R}^n \times \mathcal{M}$ *satisfying* $(\delta_x \otimes \mu)[D_f] = 0$.

Proof. Since $\mathcal{M}$ is locally uniformly $\|\cdot\|^{\gamma p}$−integrating, the relative topologies induced by $\tau_{\mathbb{R}^n} \otimes \tau_{\|\cdot\|^{\gamma p}}$ and $\tau_{\mathbb{R}^n} \otimes \tau_w^s$ on $\mathbb{R}^n \times \mathcal{M}$ coincide by Lemma 2.66. Hence, Theorem 2.84 follows directly from Theorem 2.82. □

The following example shows that the restriction to a locally uniformly $\|\cdot\|^{\gamma p}$−integrating set is essential:

Example 2.85. *Consider the mapping* $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *given by* $f(x, z) = x + z$ *and choose* ρ *to be the expectation. Then* $[\mathbb{A}_f]$ *holds with exponent* $\gamma = 1$ *and* $[\mathbb{A}_\rho]$ *is fulfilled with* $p = 1$*. Consequently,* $Q : \mathbb{R} \times \mathcal{M}_1^1 \to \mathbb{R}$, $Q(x, \mu) = \int_{\mathbb{R}\times\mathbb{R}} x' + z\ (\delta_x \otimes \mu)(d(x', z))$ *is continuous with respect to* $\tau_{\mathbb{R}} \otimes \tau_{|\cdot|}$ *by Theorem 2.84 and the fact that* $D_f = \emptyset$.

However, Q *is not continuous with respect to* $\tau_{\mathbb{R}} \otimes \tau_w^1$*: Consider sequence* $\{\mu_l\}_{l\in\mathbb{N}} \subseteq \mathcal{M}_1^1$ *given by* $\mu_l := (1 - \frac{1}{l})\delta_0 + \frac{1}{l}\delta_l$*. The calculation in Example 2.64 shows that* $\mu_l \overset{w}{\to} \delta_0$*, while*

$$\lim_{l\to\infty} Q(x, \mu_l) = x + 1 \neq x = Q(x, \delta_0)$$

holds for any $x \in \mathbb{R}$*. This is due to the fact that* $\{\mu_l \mid l \in \mathbb{N}\} \subset \mathcal{M}_1^1$ *is not locally uniformly* $|\cdot|$*−integrating by Example 2.64 and Lemma 2.66.*

If μ is absolutely continuous with respect to the Lebesgue measure, $(\delta_x \otimes \mu)[D_f] = 0$ holds whenever a certain projection of D_f has Lebesgue measure zero:

Corollary 2.86. *Assume* $[\mathbb{A}_f]$, $[\mathbb{A}_\rho]$ *and let* $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$ *be locally uniformly* $\|\cdot\|^{\gamma p}$*−integrating. Let* $(x, \mu) \in \mathbb{R}^n \times \mathcal{M}$ *be such that* μ *is absolutely continuous with respect to the Lebesgue measure* λ^s *and*

$$\lambda^s[\{z \in \mathbb{R}^s \mid (x, z) \in D_f\}] = 0. \tag{2.18}$$

Then $Q|_{\mathbb{R}^n \times \mathcal{M}}$ *is continuous with respect to the product topology of the standard topology on* $\mathbb{R}^n$ *and the relative topology of weak convergence on* $\mathcal{M}$ *at* (x, μ).

Proof. Since μ is absolutely continuous with respect to the Lebesgue measure, (2.18) implies

$$(\delta_x \otimes \mu)[D_f] = \mu[\{z \in \mathbb{R}^s \mid (x, z) \in D_f\}] = 0$$

and Theorem 2.84 is applicable. □

3. Stability of two-stage mean-risk models

In this chapter, the results of section 2.5 are applied to derive weak continuity of functionals arising from two-stage mean-risk models. Here, the function f is given by the optimal value function of the recourse problem and the verification of assumption $[\mathbb{A}_f]$ becomes a major issue. Furthermore, situations in which an explicit description of a suitable superset of the set of discontinuities of f is available are of special interest in view of Theorem 2.84. By a classical result from parametric optimization, continuity of the objective function allows for immediate conclusions about the stability of the mean-risk problem under perturbations of the underlying probability measure.

After introducing a general framework for two-stage mean-risk models in section 3.1, some preliminaries including the mentioned classical stability result by Claude Berge are provided in section 3.2. Sections 3.3 to 3.6 then examine classes of recourse problems for which the assumptions of Theorem 2.84 are fulfilled: Section 3.3 is devoted to linear recourse problems, while the mixed-integer linear case is considered in section 3.4. A class of mixed-integer quadratic recourse problems with linear constraints is examined in section 3.5. Finally, the results of section 3.6 apply to the comprehensive class of mixed-integer recourse models where the continuous relaxation admits a convex description.

Parts of this chapter have also been submitted for publication (see [69] for a preprint).

3.1. Two-stage mean-risk models

Two-stage stochastic programming problems arise from parametric optimization problems under stochastic data uncertainty and an information constraint. The latter dictates which decisions have to be taken without knowledge of the realization of the randomness. In the following analysis, parametric problems of the form

$$P(z) \qquad \min_{x,y}\{c(x,z) + q(x,y,z) \mid x \in X, y \in C(x,z)\}$$

shall be considered. Here, $X \subseteq \mathbb{R}^n$ is a fixed nonempty set, the objective is the sum of the functions $c : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ and $q : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s \to \mathbb{R}$ and both the objective and the feasible set described by the set-valued mapping $C : \mathbb{R}^n \times \mathbb{R}^s \to 2^{\mathbb{R}^m}$ depend on the parameter $z \in \mathbb{R}^s$.

Let $Z : \Omega' \to \mathbb{R}^s$ be a known random vector on some probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and consider the problem $P(Z(\omega))$. Depending on the which interplay between decision and observation is assumed, three settings are possible:

(a) $Z(\omega)$ can be observed before deciding on x and y. In this case, $P(Z(\omega))$ boils down to a deterministic problem.

(b) If both x and y have to be chosen without knowledge of $Z(\omega)$, the resulting problem is a one-stage stochastic optimization problem.

(c) $Z(\omega)$ can only be observed after making the decision on x. The variable y can then be chosen under complete information and the resulting problem is a two-stage stochastic optimization problem.

In (b) and (c), the stochasticity is usually assumed to be purely exogenous. The latter means that the distribution of Z does not depend on the choice of x and y. Note that (b) arises as a special case of (c), where the decision on y does not influence the outcome. Furthermore, the case where $Z(\omega)$ can only be observed after making the decision on y and x is chosen under complete information can be neglected due to the symmetry of $P(Z)$ in x and y.

In the following, setting (c) and purely exogenous stochasticity will be assumed. After deciding on x and observing $Z(\omega)$, the optimal decision y can be obtained by solving the so called recourse problem

$$\min_y \{q(x, y, Z(\omega)) \mid y \in C(x, Z(\omega))\}. \tag{3.1}$$

Since both x and $Z(\omega)$ are assumed to be known, (3.1) is a deterministic problem. This consideration allows to formulate the two-stage stochastic programming problem as

$$\min_x \{c(x, Z(\omega)) + \min_y \{q(x, y, Z(\omega)) \mid y \in C(x, Z(\omega))\} \mid x \in X\}. \tag{3.2}$$

Note that problem (3.2) is not well defined in view of setting (c). Possibilities of resolving this issue include robust approaches and models based on probabilistic constraints or stochastic dominance relations. However, the focus of the present analysis is on mean-risk models: Consider the mapping $f : \mathbb{R}^n \times \mathbb{R}^s \to \bar{\mathbb{R}} := [-\infty, \infty]$ defined by

$$f(x,z) := c(x,z) + \inf_y \{q(x,y,z) \mid y \in C(x,z)\}. \tag{3.3}$$

Under assumptions which guarantee finiteness and measurability, f induces a mapping $f_{(\cdot)} : \mathbb{R}^n \to L^0(\Omega', \mathcal{F}', \mathbb{P}')$ via $f_x(\omega) := f(x, Z(\omega))$. The decision on x can now be taken based on some ranking of the random variables in the family $f_X := \{f_x \mid x \in X\}$. Assuming that $f_X \subseteq L^1(\Omega', \mathcal{F}', \mathbb{P}')$, applying the expectation yields the well defined, risk neutral model

$$\min_x \{\int_{\Omega'} f(x, Z(\omega))\ \mathbb{P}'(d\omega) \mid x \in X\}. \tag{3.4}$$

Mean-risk models allow to take into account risk aversion by adding some quantification of risk to the objective. Note that these models depend on the underlying random vector Z. In practice, only an approximation of Z may be available, which has motivated the investigation of the behavior of mean-risk models under perturbations of Z. If the quantification of risk is law-invariant, one might equivalently work with the Borel probability measure $\mathbb{P}' \circ Z^{-1}$ induced by Z. For qualitative stability analysis, one typically considers perturbations of this measure with respect to the topology of weak convergence. The following example demonstrates that in such a setting, even problem (3.4) may be highly unstable:

Example 3.1. *Fix any parameter $\lambda > 0$ and consider the one-stage setting where $X = \mathbb{R}$, $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is given by $c(x,z) = \chi_{\{0\}}(z) \cdot (x^2 + \lambda)$, $q \equiv 0$ and $C \equiv \mathbb{R}$. Let the true random vector Z induce the measure $\mathbb{P}' \circ Z^{-1} = \delta_0$, so that problem (3.4) takes the form*

$$\min_{x \in \mathbb{R}} \int_{\mathbb{R}} f(x,z)\ \delta_0(dz) = \min_{x \in \mathbb{R}}\ \chi_{\{0\}}(0) \cdot (x^2 + \lambda) = \min_{x \in \mathbb{R}}\ x^2 + \lambda.$$

The unique optimal solution is given by $x = 0$ and yields the value λ. By Example 2.83, it holds that $\delta_{\frac{1}{l}} \xrightarrow{|\cdot|} \delta_0$ and hence $\delta_{\frac{1}{l}} \xrightarrow{w} \delta_0$ as $l \to \infty$. However, for any $l \in \mathbb{N}$, approximating δ_0 by $\delta_{\frac{1}{l}}$ results in the problem

$$\min_{x \in \mathbb{R}} \int_{\mathbb{R}} f(x,z)\ \delta_{\frac{1}{l}}(dz) = \min_{x \in \mathbb{R}}\ \chi_{\{0\}}(\frac{1}{l}) \cdot (x^2 + \lambda) = \min_{x \in \mathbb{R}}\ 0. \tag{3.5}$$

Every $x \in \mathbb{R}$ is an optimal solution of (3.5) and yields the value $0 < \lambda$. This instability results from the fact that $(\delta_x \otimes \delta_0)[D_f] = 1 \neq 0$ for any $x \in \mathbb{R}$.

Such instabilities may even occur if f is smooth:

Example 3.2. *Consider the one-stage setting where $X = [0,1]$, $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is given* *by* $c(x,z) = x + z$, $q \equiv 0$, $C \equiv \mathbb{R}$ *and* $\mathbb{P}' \circ Z^{-1} = \delta_0$. *Problem* (3.4) *takes the form*

$$\min_{x \in [0,1]} \int_{\mathbb{R}} f(x,z)\ \delta_0(dz) = \min_{x \in [0,1]}\ x$$

and yields the optimal value 0. *By Example 2.64,* $(1 - \frac{1}{l})\delta_0 + \frac{1}{l}\delta_l \xrightarrow{w} \delta_0$ *as* $l \to \infty$. *However, the calculation in Example 2.85 shows that for any* $l \in \mathbb{N}$, *approximating* δ_0 *by* $(1 - \frac{1}{l})\delta_0 + \frac{1}{l}\delta_l$ *results in the problem*

$$\min_{x \in [0,1]} \int_{\mathbb{R}} f(x,z)\ ((1 - \frac{1}{l})\delta_0 + \frac{1}{l}\delta_l)(dz) = \min_{x \in [0,1]}\ x + 1,$$

which has the optimal value 1. *Hence, the optimal value function of* (3.4) *can be discontinuous with respect to the topology of weak convergence even if* f *has continuous derivatives of all orders.*

In the following, the focus will be on situations where assumption $[\mathbb{A}_f]$ holds for the function f defined in (3.3) and the risk is quantified by a mapping ρ satisfying assumption $[\mathbb{A}_\rho]$. Note that the expectation fits into this framework by Example 2.25. Furthermore, Remark 2.19 allows to confine the analysis to a single mapping instead of a weighted sum. The models to be analyzed may be represented as

$$\min_x \{Q(x,\mu) \mid x \in X\}, \tag{3.6}$$

where $\mu = \mathbb{P}' \circ Z^{-1} \in \mathcal{P}(\mathbb{R}^s)$ and the objective function

$$Q(x,\mu) = \mathcal{R}_\rho((\delta_x \otimes \mu) \circ f^{-1})$$

is exactly as in (2.4). Hence, Theorem 2.84 can be applied to derive continuity of a restriction of Q with respect to the product topology of the standard topology on $\mathbb{R}^n$ and the relative topology of weak convergence on a suitable subset $\mathcal{M}$ of $\mathcal{P}(\mathbb{R}^s)$. The latter allows to draw conclusions about the optimal value function $\varphi : \mathcal{M} \to \bar{\mathbb{R}}$,

$$\varphi(\mu) := \inf_x \{Q(x,\mu) \mid x \in X\} \tag{3.7}$$

and the optimal solution set mapping $\Phi : \mathcal{M} \to 2^X$,

$$\Phi(\mu) := \{x \in X \mid Q(x,\mu) = \varphi(\mu)\} \tag{3.8}$$

of problem (3.6). The subsequent sections identify classes of parametric problems $P(z)$

for which the mapping f induced via (3.3) automatically fulfills assumption $[\mathbb{A}_f]$. This is done by imposing verifiable assumptions on the data, i.e. the mappings c, q and C.

3.2. Preliminaries

The following assumption allows to confine the analysis to the optimal value function of the recourse problem, which may be represented as $f - c$:

Assumption 3.3 ($[\mathbb{A}_c]$: **Assumptions on** c**).**
c is continuous and such that there exist a constant $\gamma_c \geq 0$ and a locally bounded mapping $\eta_c : \mathbb{R}^n \to \mathbb{R}$ satisfying

$$|c(x,z)| \leq \eta_c(x)(\|z\|^{\gamma_c} + 1)$$

for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$.

Remark 3.4. *If c is continuous and does not depend on z, assumption $[\mathbb{A}_c]$ is fulfilled with exponent 0.*

Lemma 3.5. *Assume $[\mathbb{A}_c]$ and let $[\mathbb{A}_f]$ be fulfilled for $f - c$ with exponent $\gamma_{f-c} \geq 0$. Then $[\mathbb{A}_f]$ is fulfilled for f with exponent $\gamma = \max\{\gamma_c, \gamma_{f-c}\}$. Furthermore, $D_f = D_{f-c}$.*

Proof. Under the given assumptions, f is the sum of two Borel measurable functions and hence Borel measurable. Let η_{f-c} denote the locally bounded mapping from assumption $[\mathbb{A}_f]$ for $f - c$. Then

$$\begin{aligned} |f(x,z)| &\leq |c(x,z)| + |(f-c)(x,z)| \\ &\leq \eta_c(x)(\|z\|^{\gamma_c} + 1) + \eta_{f-c}(x)(\|z\|^{\gamma_{f-c}} + 1) \\ &\leq (\eta_c(x) + \eta_{f-c}(x))(\|z\|^{\max\{\gamma_c, \gamma_{f-c}\}} + 1) \end{aligned}$$

holds for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$. Since the mapping $\eta(x) = \eta_c(x) + \eta_{f-c}(x)$ is locally bounded, the latter means that assumption $[\mathbb{A}_f]$ is fulfilled for f with exponent $\gamma = \max\{\gamma_c, \gamma_{f-c}\}$. Finally, $D_f = D_{f-c}$ is a direct conclusion from the continuity of c. □

Properties of set-valued mappings can be expressed using the notion of hemicontinuity:

Definition 3.6. *Let S and T be topological spaces. A mapping $\Upsilon : T \to 2^S$ is called **upper hemicontinuous** at some $t_0 \in T$ iff for any open set $\mathcal{O} \subseteq S$ satisfying $\Upsilon(t_0) \subseteq \mathcal{O}$, there exists a neighborhood $\mathcal{N}$ of t_0 such that $\Upsilon(t) \subseteq \mathcal{O}$ for any $t \in \mathcal{N}$. Υ is called **lower hemicontinuous** at t_0 iff for any open set $\mathcal{O} \subseteq S$ satisfying $\Upsilon(t_0) \cap \mathcal{O} \neq \emptyset$, there exists a neighborhood $\mathcal{N}$ of t_0 such that $\Upsilon(t) \cap \mathcal{O} \neq \emptyset$ for any $t \in \mathcal{N}$.*

The following classical result by Claude Berge will in particular be applied to derive qualitative stability of problem (3.6) from the continuity of Q:

Theorem 3.7 (**[41, Theorem 2]**).
Let S, T be metric spaces, $\Upsilon : T \to 2^S$ compact-valued and $\upsilon : S \times T \to \mathbb{R}$ continuous with respect to the first argument. Assume that for any $s \in S$, $\upsilon(s, \cdot)$ is continuous and that Υ is upper hemicontinuous at some $t_0 \in T$. Then the mapping $\upsilon^* : T \to \bar{\mathbb{R}}$ given by $\upsilon^*(t) = \inf_s\{\upsilon(s,t) \mid s \in \Upsilon(t)\}$ is lower semicontinuous at t_0. If, in addition, Υ is lower hemicontinuous at t_0, υ^* is continuous and the mapping $\Upsilon^* : T \to 2^S$ defined by $\Upsilon^*(t) = \{s \in \Upsilon(t) \mid \upsilon(s,t) = \upsilon^*(t)\}$ is upper hemicontinuous at t_0. Furthermore, $\Upsilon^*(t_0)$ is nonempty and compact.

Refer to section 4.2 in [20] or the recent work [154] for various generalizations of the above result. It is well known that the statement of the Theorem 3.7 does not hold in general if Υ is not compact valued (see e.g. [151] for a counterexample and [102] for an extension of Theorem 3.7 based on a so-called inf-compactness condition). However, the following lemma is still applicable if the feasible set is fixed (see e.g. section 4.1 in [55] for a proof):

Lemma 3.8. *Let S, T be metric spaces, $\upsilon : S \times T \to \mathbb{R}$ a function and $\Upsilon_0 \subseteq S$ a fixed set. Assume that for any $s \in S$, $\upsilon(s, \cdot)$ is upper semicontinuous at $t_0 \in T$. Then the mapping $\upsilon^* : T \to \bar{\mathbb{R}}$, $\upsilon^*(t) = \inf_s\{\upsilon(s,t) \mid s \in \Upsilon_0\}$ is upper semicontinuous at t_0.*

3.3. Linear recourse

This section examines the case where the recourse is given by a linear problem in y. Let (3.1) take the form

$$\min_y\{\bar{q}(x, Z(\omega))^\top y \mid Ay = h(x, Z(\omega)),\ y \geq 0\}, \tag{3.9}$$

where $A \in \mathbb{R}^{k \times m}$ is a matrix and $\bar{q} : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ and $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^k$ are mappings. The following classical result from parametric optimization is a conclusion from the basis decomposition theorem in [209] and will be instrumental in the analysis of (3.9):

Theorem 3.9. *Let $A \in \mathbb{R}^{k \times m}$ have full rank. Then the mapping $\varphi_{lin} : \mathbb{R}^m \times \mathbb{R}^k \to \bar{\mathbb{R}}$ defined by*

$$\varphi_{lin}(t_1, t_2) = \inf_y\{t_1^\top y \mid Ay = t_2,\ y \geq 0\}$$

is finite and continuous on the polyhedral cone $D(A) \times \mathrm{pos}(A)$, where

$$D(A) := \{t_1 \in \mathbb{R}^m \mid \{u \in \mathbb{R}^k \mid A^\top u \leq t_1\} \neq \emptyset\}$$

and $\text{pos}(A) := \{Ay \mid y \in \mathbb{R}^m,\ y \geq 0\}$*. Moreover, there exist matrices* $B_1, \ldots, B_N \in \mathbb{R}^{k\times m}$ *and polyhedral cones* $\mathcal{K}_1, \ldots, \mathcal{K}_N \subseteq \mathbb{R}^m \times \mathbb{R}^k$ *such that*

$$\bigcup_{j=1}^{N} \mathcal{K}_j = D(A) \times \text{pos}(A), \quad \text{int}\,\mathcal{K}_i \cap \text{int}\,\mathcal{K}_j = \emptyset \ \textit{ whenever } i \neq j$$

and

$$\varphi_{lin}(t_1, t_2) = t_2^\top B_j t_1 \quad \forall (t_1, t_2) \in \mathcal{K}_j.$$

Furthermore, for any $(t_1, t_2) \in D(A)\times\text{pos}(A)$*,* $\varphi_{lin}(t_1, \cdot)$ *is convex on* $\text{pos}(A)$ *and* $\varphi_{lin}(\cdot, t_2)$ *is concave on* $D(A)$*.*

Theorem 3.9 motivates the following assumption:

Assumption 3.10 ($[\mathbb{A}_{LP}]$: **Assumptions for linear recourse**).
A has full rank, $\bar{q}$ *and* h *are continuous and for any* $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$*, it holds that* $\bar{q}(x, z) \in D(A)$ *and* $h(x, z) \in \text{pos}(A)$*. Furthermore, there exist constants* $\gamma_h, \gamma_{\bar{q}} \geq 0$ *and locally bounded mappings* $\eta_h, \eta_{\bar{q}} : \mathbb{R}^n \to \mathbb{R}$ *such that*

$$\|h(x, z)\| \leq \eta_h(x)(\|z\|^{\gamma_h} + 1) \ \textit{ and } \ \|\bar{q}(x, z)\| \leq \eta_{\bar{q}}(x)(\|z\|^{\gamma_{\bar{q}}} + 1)$$

hold for any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$*.*

Remark 3.11. $\bar{q}(x, z) \in D(A)$ *holds in particular if* $\bar{q}(x, z) \geq 0$*.*

The first part of $[\mathbb{A}_{LP}]$ is a standard assumption in linear two-stage stochastic programming (see e.g. [181], [184]). The next theorem examines the stability of problem (3.6) with linear recourse:

Theorem 3.12. *Assume that* $[\mathbb{A}_c]$*,* $[\mathbb{A}_\rho]$ *and* $[\mathbb{A}_{LP}]$ *are fulfilled and let* $f : \mathbb{R}^n \times \mathbb{R}^s \to \bar{\mathbb{R}}$ *be given by*

$$f(x, z) = c(x, z) + \inf_y \{\bar{q}(x, z)^\top y \mid Ay = h(x, z),\ y \geq 0\}.$$

Set $\gamma := \max\{\gamma_c, \gamma_{\bar{q}} + \gamma_h\}$ *and let* $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$ *be locally uniformly* $\|\cdot\|^{\gamma p}$*−integrating. Then the optimal value function* φ *defined in* (3.7) *is upper semicontinuous with respect to the relative topology of weak convergence on* $\mathcal{M}$*. If* X *is compact,* φ *is continuous and the optimal solution set mapping* $\Phi : \mathcal{M} \to 2^X \setminus \{\emptyset\}$ *defined in* (3.8) *is compact-valued and upper hemicontinuous with respect to the relative topology of weak convergence on* $\mathcal{M}$*.*

Proof. By assumptions $[\mathbb{A}_c]$, $[\mathbb{A}_{LP}]$ and Theorem 3.9,

$$f(x, z) = c(x, z) + \varphi_{lin}(\bar{q}(x, z), h(x, z))$$

is a composition of real-valued, continuous functions. Consequently, f is real-valued and continuous on $\mathbb{R}^n \times \mathbb{R}^s$, i.e. $D_f = \emptyset$. In particular, f is Borel measurable.

Set $\kappa_B := \max_{j=1,\dots,N} \|B_j\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^k)} < \infty$, where $B_1, \dots, B_N$ are the matrices from Theorem 3.9 and $\|\cdot\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^k)}$ denotes the operator norm. Fix any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^k$. By $[\mathbb{A}_{LP}]$ there exists an index $j \in \{1, \dots, N\}$ such that $(\bar{q}(x,z), h(x,z)) \in \mathcal{K}_j \subseteq D(A) \times \operatorname{pos}(A)$. Consequently,

$$\begin{aligned} |(f-c)(x,z)| &= |h(x,z)^\top B_j \bar{q}(x,z)| \;\le\; \kappa_B \|h(x,z)\| \|\bar{q}(x,z)\| \\ &\le \kappa_B \eta_h(x) \eta_{\bar{q}}(x) (\|z\|^{\gamma_h} + 1)(\|z\|^{\gamma_{\bar{q}}} + 1) \\ &\le 3\kappa_B \eta_h(x) \eta_{\bar{q}}(x) (\|z\|^{\gamma_h + \gamma_{\bar{q}}} + 1). \end{aligned}$$

Since $\eta_{f-c}(x) := 3\kappa_B \eta_h(x) \eta_{\bar{q}}(x)$ is locally bounded, assumption $[\mathbb{A}_f]$ is fulfilled for f with exponent γ by Lemma 3.5. Hence, Theorem 2.84 is applicable and yields the continuity of $Q|_{\mathbb{R}^n \times \mathcal{M}}$ with respect to the relative topology induced by $\tau_{\mathbb{R}^n} \otimes \tau_w^s$.

Since the feasible set X is fixed, the continuity of the objective immediately entails the upper semicontinuity of φ by Lemma 3.8. The stronger statements for compact X follow directly from Theorem 3.7. □

3.4. Mixed-integer linear recourse

In this section, mean-risk models with mixed-integer linear recourse are considered, i.e. C and f take the form

$$C(x,z) = \{(y_1, y_2) \in \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2} \mid A_1 y_1 + A_2 y_2 = h(x,z),\ y_1, y_2 \ge 0\} \tag{3.10}$$

and

$$f(x,z) = c(x,z) + \inf_{y_1, y_2} \{q_1^\top y_1 + q_2^\top y_2 \mid (y_1, y_2) \in C(x,z)\}, \tag{3.11}$$

where $q_1 \in \mathbb{R}^{m_1}$, $q_2 \in \mathbb{R}^{m_2}$, $A_1 \in \mathbb{R}^{k \times m_1}$, $A_2 \in \mathbb{R}^{k \times m_2}$ and the mappings $c : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ and $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^k$ are fixed. Note that only the right-hand side of the constraint system depends on (x,z), while the objective function of the recourse problem is fixed. In the context of two-stage stochastic programming, similar recourse problems have been studied e.g. in [153], [195], [200] and [201]. As in the linear case, a classical result from parametric optimization (see [22], [53]) can be applied to analyze f:

Theorem 3.13. *Let A_1, A_2 have rational entries and assume that*

$$\mathrm{pos}(A_1) + \{A_2 y_2 \mid y_2 \in \mathbb{Z}^{m_2},\ y_2 \geq 0\} = \mathbb{R}^k \tag{3.12}$$

and

$$\{u \in \mathbb{R}^k \mid A_1^\top u \leq q_1, A_2^\top u \leq q_2\} \neq \emptyset. \tag{3.13}$$

Then the mapping $\varphi_{MILP} : \mathbb{R}^k \to \bar{\mathbb{R}}$ defined by

$$\varphi_{MILP}(t) := \inf_{y_1, y_2} \{q_1^\top y_1 + q_2^\top y_2 \mid A_1 y_1 + A_2 y_2 = t,\ y_1, y_2 \geq 0,\ y_2 \in \mathbb{Z}^{m_2}\}$$

has the following properties:

(a) φ_{MILP} is real-valued and lower semicontinuous on $\mathbb{R}^k$.

(b) There exists a countable partition $\mathbb{R}^k = \bigcup_{j=1}^\infty \mathcal{T}_j$ such that for any $j \in \mathbb{N}$, the restriction of φ_{MILP} to $\mathcal{T}_j$ is piecewise linear and Lipschitz continuous with a uniform constant not depending on j. The restriction $\varphi_{MILP}|_{\mathcal{T}_j}$ admits a representation as

$$\varphi_{MILP}|_{\mathcal{T}_j}(t) = \min_{y_2} \{q_2^\top y_2 + \max_{i=1,\ldots,M} u_i^\top (t - A_2 y_2) \mid y_2 \in Y_2(t)\},$$

where $Y_2(t) = \{y_2 \in \mathbb{Z}^{m_2} \mid t \in \mathrm{pos}(A_1) + A_2 y_2,\ y_2 \geq 0\}$ and $u_1, \ldots, u_M$ are the vertices of the polyhedron $\{u \in \mathbb{R}^k \mid A_1^\top u \leq q_1\}$. Moreover, for any $j \in \mathbb{N}$, there exist $t_{j,1}, \ldots, t_{j,N} \in \mathbb{R}^k$ satisfying

$$\mathcal{T}_j = (\{t_{j,1}\} + \mathrm{pos}(A_1)) \setminus \bigcup_{i=2}^N (\{t_{j,i}\} + \mathrm{pos}(A_1))$$

and N does not depend on j.

(c) There exist constants $\alpha, \beta > 0$ such that

$$|\varphi_{MILP}(t_1) - \varphi_{MILP}(t_2)| \leq \alpha \|t_1 - t_2\| + \beta$$

holds for any $t_1, t_2 \in \mathbb{R}^k$.

Remark 3.14. (3.12) *is referred to as* complete mixed-integer recourse*, while* (3.13) *is called* sufficiently expensive recourse*.*

Because of the above theorem, (3.12) and (3.13) are often assumed when dealing with mixed-integer linear recourse problems in two-stage stochastic programming (see e.g. [153],

[200], [201]). In view of (3.10) an additional assumption concerning h is needed:

Assumption 3.15 ($[\mathbb{A}_{MILP}]$: **Assumptions for mixed-integer linear recourse).** *A_1, A_2 have rational entries and are such that* (3.12) *and* (3.13) *are fulfilled. Furthermore, h is continuous and there exist a locally bounded mapping $\eta_h : \mathbb{R}^n \to \mathbb{R}$ and a constant $\gamma_h > 0$ such that*

$$\|h(x,z)\| \leq \eta_h(x)(\|z\|^{\gamma_h} + 1)$$

holds for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$.

Theorem 3.16. *Let f be given by* (3.11) *and assume that* $[\mathbb{A}_c]$, $[\mathbb{A}_\rho]$ *and* $[\mathbb{A}_{MILP}]$ *are fulfilled. Set $\gamma := \max\{\gamma_c, \gamma_h\}$, let $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$ be locally uniformly $\|\cdot\|^{\gamma p}$−integrating and $\mu_0 \in \mathcal{M}$ such that $(\delta_x \otimes \mu_0)[D_{f-c}] = 0$ holds for any $x \in X$. Then $\varphi : \mathcal{M} \to \bar{\mathbb{R}}$ defined in* (3.7) *is upper semicontinuous at μ_0 with respect to the relative topology of weak convergence on $\mathcal{M}$. If X is compact, φ is continuous at μ_0 and $\Phi : \mathcal{M} \to 2^X$ defined in* (3.8) *is upper hemicontinuous at μ_0 with respect to the relative topology of weak convergence on $\mathcal{M}$. In this case, $\Phi(\mu_0)$ is nonempty and compact.*

Proof. By assumption $[\mathbb{A}_{MILP}]$ and part (a) of Theorem 3.13,

$$(f-c)(x,z) = \varphi_{MILP}(h(x,z))$$

is real-valued and the composition of a continuous and an upper semicontinuous function. Consequently, f is real-valued and upper semicontinuous on $\mathbb{R}^n \times \mathbb{R}^s$ by assumption $[\mathbb{A}_c]$ and hence Borel measurable. Let α and β denote the constants from part (c) of Theorem 3.13 and set $\bar{\beta} := \alpha\|h(0,0)\| + \beta + |\varphi_{MILP}(h(0,0))|$. Then

$$\begin{aligned} |(f-c)(x,z)| &\leq |\varphi_{MILP}(h(x,z)) - \varphi_{MILP}(h(0,0))| + |\varphi_{MILP}(h(0,0))| \\ &\leq \alpha\|h(x,z) - h(0,0)\| + \beta + |\varphi_{MILP}(h(0,0))| \\ &\leq \alpha\|h(x,z)\| + \bar{\beta} \;\leq\; (\alpha\eta_h(x) + \bar{\beta})(\|z\|^{\gamma_h} + 1) \end{aligned}$$

holds for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$. Since $\eta_{f-c}(x) := \alpha\eta_h(x) + \bar{\beta}$ is locally bounded, assumption $[\mathbb{A}_f]$ is fulfilled for f with exponent γ by Lemma 3.5. Hence, Theorem 2.84 is applicable and yields the continuity of $Q|_{X\times\mathcal{M}}$ at any $(x,\mu) \in X \times \mathcal{M}$ satisfying $(\delta_x \otimes \mu)[D_{f-c}] = 0$ with respect to the relative topology induced by $\tau_{\mathbb{R}^n} \otimes \tau_w^s$ on $X \times \mathcal{M}$. Since the feasible set X is fixed, the stated stability is a direct conclusion from Lemma 3.8 and Theorem 3.7. □

The following result points out a special case in which $(\delta_x \otimes \mu)[D_{f-c}] = 0$ holds for all

$x \in X$:

Proposition 3.17. *Let f be given by (3.11) and assume $[\mathbb{A}_{MILP}]$. In addition, assume that μ is absolutely continuous with respect to the Lebesgue measure, $k = s$ and that for any $x \in X$, the mapping $h_x : \mathbb{R}^s \to \mathbb{R}^s$ defined by $h_x(z) = h(x,z)$ is a C^1-diffeomorphism. Then $(\delta_x \otimes \mu)[D_{f-c}] = 0$ holds for any $x \in X$.*

Proof. Part (b) of Theorem 3.13 implies that φ_{MILP} is continuous outside of

$$\bigcup_{y_2 \in \mathbb{Z}^{m_2},\, y_2 \geq 0} [\{A_2 y_2\} + \partial(\mathrm{pos}(A_1))]. \tag{3.14}$$

Since the set in (3.14) is contained in a countable union of hyperplanes, the Lebesgue measure of $D_{\varphi_{MILP}}$ is equal to zero. $[\mathbb{A}_{MILP}]$ implies $D_{f-c} \subseteq h^{-1}(D_{\varphi_{MILP}})$. Fix any $x \in X$, then

$$\begin{aligned}(\delta_x \otimes \mu_0)[D_{f-c}] &= \mu[\{z \in \mathbb{R}^s \mid (x,z) \in D_{f-c}\}] \\ &\leq \mu[\{z \in \mathbb{R}^s \mid (x,z) \in h^{-1}(D_{\varphi_{MILP}})\}] \\ &= \mu[h_x^{-1}(D_{\varphi_{MILP}})].\end{aligned}$$

h_x is a C^1-diffeomorphism and hence the Lebesgue measure of $h_x^{-1}(D_{\varphi_{MILP}})$ is equal to zero by [189, Lemma 7.25]. Consequently, $(\delta_x \otimes \mu_0)[D_{f-c}] = 0$, since μ is absolutely continuous with respect to the Lebesgue measure. □

3.5. Mixed-integer quadratic recourse

In this section, recourse problems with quadratic objective, linear constrains and mixed-integer variables are considered. Consequently, C and f take the form

$$C(x,z) = \{y \in \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2} \mid Ay \leq h(x,z)\}$$

and

$$f(x,z) = c(x,z) + \inf_y \{y^\top D y + d(x,z)^\top y \mid y \in C(x,z)\}, \tag{3.15}$$

where $m_1 + m_2 = m$, $D \in \mathbb{Q}^{m \times m}$ is a symmetric, positive definite matrix and $A \in \mathbb{Q}^{k \times m}$. The linear part of the objective and the right-hand side of the constraint system are given by mappings $d : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ and $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^k$, respectively.

Remark 3.18. *Stability of two-stage stochastic programs with mixed-integer quadratic* *recourse has been analyzed in [64]. While the authors of the mentioned paper also consider the situation where D is positive semidefinite, they confine their stability analysis to measures in*

$$\mathcal{P}_\Xi := \{\mu \in \mathcal{P}(\mathbb{R}^s) \mid \mu[\Xi] = 1\}, \tag{3.16}$$

where $\Xi \subset \mathbb{R}^s$ is a fixed compact set. Note that for any gauge function $\psi : \mathbb{R}^s \to [0, \infty)$ and any compact set $\Xi \subset \mathbb{R}^s$, $\mathcal{P}_\Xi \subseteq \mathcal{M}_s^\psi$ is relatively compact for the $\psi-$weak topology by Theorem 2.68 and hence locally uniformly $\psi-$integrating by Proposition 2.67.

Furthermore, [64] only examines an objective function that is based on the expectation (although their model may reflect some kind of risk-aversion, see page 465 in [64] for details) and assumes d and h to be of a special form (for any fixed z, d is linear in x, while h does not depend on x). Consequently, the present analysis allows to extend some results of [64] in various directions.

Several stability results are available for parametric programs with linear constraints (see e.g. [18], [45], [46], [95], [120], [136], [137], [146], [147], [205] for continuous and [118] for integer variables and refer to [20] (chapter 5), [21], [63] or [64] for the mixed-integer case). The following theorem is a combination of [63, Theorem 2.2] and [64, Lemma 2.7, Remark 2.8] and will be used to analyze f:

Theorem 3.19. *Assume that $C'(t) := \{y \in \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2} \mid Ay \leq t\} \neq \emptyset$ holds for any $t \in \mathbb{R}^k$. Then the mapping $\varphi_{MIQP} : \mathbb{R}^m \times \mathbb{R}^k \to \bar{\mathbb{R}}$ defined by*

$$\varphi_{MIQP}(u, t) = \inf_y \{y^\top Dy + u^\top y \mid y \in C'(t)\}$$

is real-valued and lower semicontinuous on $\mathbb{R}^m \times \mathbb{R}^k$. Furthermore, there exists a constant $\kappa_{MIPQ} > 0$ such that

$$|\varphi_{MIQP}(u,t) - \varphi_{MIQP}(u',t')| \leq \kappa_{MIQP}(\max\{\|(u,t)\|, \|(u',t')\|\}(\|u-u'\| + \|t-t'\| + 1) + 1)$$

holds for any $(u, t), (u', t') \in \mathbb{R}^m \times \mathbb{R}^k$.

The following assumption is motivated by Theorem 3.19:

Assumption 3.20 ($[\mathbb{A}_{MIQP}]$: **Assumptions for mixed-integer quadratic recourse).** *$C'(t) \neq \emptyset$ for any $t \in \mathbb{R}^k$, the mappings d and h are continuous and there exist constants $\gamma_d, \gamma_h \geq 0$ and locally bounded mappings $\eta_d, \eta_h : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\|d(x,z)\| \leq \eta_d(x)(\|z\|^{\gamma_d} + 1) \;\; \textit{and} \;\; \|h(x,z)\| \leq \eta_h(x)(\|z\|^{\gamma_h} + 1)$$

hold for any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$*.*

Assumption $[\mathbb{A}_{MIQP}]$ admits the following result:

Theorem 3.21. *Let f be given by* (3.15) *and assume that* $[\mathbb{A}_c]$, $[\mathbb{A}_\rho]$ *and* $[\mathbb{A}_{MIQP}]$ *are fulfilled. Set* $\gamma := \max\{\gamma_c, 2\gamma_h, 2\gamma_d\}$*, let* $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$ *be locally uniformly* $\|\cdot\|^{\gamma p}$*−integrating and* $\mu_0 \in \mathcal{M}$ *such that* $(\delta_x \otimes \mu_0)[D_{f-c}] = 0$ *holds for any* $x \in X$*. Then* $\varphi : \mathcal{M} \to \bar{\mathbb{R}}$ *defined by* (3.7) *is upper semicontinuous at* μ_0 *with respect to the relative topology of weak convergence on* $\mathcal{M}$*. If* X *is compact,* φ *is continuous at* μ_0 *and* $\Phi : \mathcal{M} \to 2^X$ *defined in* (3.8) *is upper hemicontinuous at* μ_0 *with respect to the relative topology of weak convergence on* $\mathcal{M}$*. In this case,* $\Phi(\mu_0)$ *is nonempty and compact.*

Proof. By assumption $[\mathbb{A}_{MIQP}]$ and Theorem 3.19,

$$(f - c)(x, z) = \varphi_{MIQP}(d(x, z), h(x, z))$$

is real-valued and the composition of two continuous and one lower semicontinuous function. Consequently, f is real-valued and lower semicontinuous on $\mathbb{R}^n \times \mathbb{R}^s$ by $[\mathbb{A}_c]$. In particular, f is Borel measurable. Set $\varphi_0 := |\varphi_{MIQP}(d(0,0), h(0,0))| + \kappa_{MIQP}$ and $n_0 := \|(d(0,0), h(0,0))\|$. For any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$, Theorem 3.19 yields

$$\begin{aligned}
|(f - c)(x, z)| &\le |\varphi_{MIQP}(d(x, z), h(x, z)) - \varphi_{MIQP}(d(0,0), h(0,0))| + \varphi_0 - \kappa_{MIQP} \\
&\le \kappa_{MIQP} \max\{\|(d(x, z), h(x, z))\|, n_0\}(\|d(x, z)\| + \|h(x, z)\| + 2n_0 + 1) + \varphi_0 \\
&\le \kappa_{MIQP}(\|d(x, z)\| + \|h(x, z)\| + 2n_0 + 1)^2 + \varphi_0 \\
&\le \kappa'(\|d(x, z)\|^2 + \|h(x, z)\|^2 + 1),
\end{aligned}$$

where $\kappa' := 9\kappa_{MIQP}(2n_0 + 1)^2 + \varphi_0$. By assumption $[\mathbb{A}_{MIQP}]$, the latter implies

$$\begin{aligned}
|(f - c)(x, z)| &\le \kappa' \eta_d(x)^2(\|z\|^{\gamma_d} + 1)^2 + \kappa' \eta_h(x)^2(\|z\|^{\gamma_h} + 1)^2 + \kappa' \\
&\le \eta_{f-c}(x)(\|z\|^{\max\{2\gamma_d, 2\gamma_h\}} + 1),
\end{aligned}$$

where $\eta_{f-c}(x) := 8\kappa'(\eta_d(x)^2 + \eta_h(x)^2 + 1)$ is locally bounded. Hence, $[\mathbb{A}_f]$ is fulfilled for f with exponent γ by Lemma 3.5 and assumption $[\mathbb{A}_c]$. Consequently, Theorem 2.84, Lemma 3.8 and Theorem 3.7 are applicable and yield the stated stability. □

The following remark points out a case in which the assumption $(\delta_x \otimes \mu_0)[D_{f-c}] = 0$ for all $x \in X$ of Theorem 3.21 is automatically fulfilled:

Remark 3.22. *Assume that* $[\mathbb{A}_c]$, $[\mathbb{A}_\rho]$ *and* $[\mathbb{A}_{MIQP}]$ *are fulfilled and that the projection of* $C(x,z)$ *to the integer components, i.e. the set*

$$\{y_2 \in \mathbb{Z}^{m_2} \mid \exists y_1 \in \mathbb{R}^{m_1} : (y_1, y_2) \in C(x,z)\},$$

does not depend on x *and* z*. Then* f *is continuous on* $\mathbb{R}^n \times \mathbb{R}^s$ *by [64, Lemma 2.9]. Hence,* $D_{f-c} = \emptyset$*.*

3.6. Mixed-integer recourse problems with convex continuous relaxation

In this section, a fairly general class of recourse problems is considered: Let C and f be given by

$$C(x,z) = \{y \in \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2} \mid g(y) \le h(x,z)\}$$

and

$$f(x,z) = c(x,z) + \inf_y \{v(y) \mid y \in C(x,z)\}, \tag{3.17}$$

where $m_1 + m_2 = m$, the right-hand side of the constraint system is given by the mapping $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^k$, $v : \mathbb{R}^m \to \mathbb{R}$ is convex and $g = (g_1, \dots, g_k)^\top : \mathbb{R}^m \to \mathbb{R}^k$ is such that for any $i \in \{1, \dots, k\}$, g_i is convex and has a closed epigraph. The following well known result about convex functions will be useful in the analysis of f:

Lemma 3.23. *Let* $\bar{v} : \mathbb{R}^m \to \mathbb{R}$ *be convex. Then for every* $r > 0$, $\bar{v}$ *is Lipschitz continuous on* $B_r(0)$ *with constant*

$$L_{\bar{v}}(r) := \frac{2}{r}\left(\max_{y \in \{2r, -2r\}^m} |\bar{v}(y)| + 2|\bar{v}(0)|\right). \tag{3.18}$$

In particular, $\bar{v}$ *is continuous on* $\mathbb{R}^m$*.*

Proof. Combine Lemma A and Theorem A in [173]. □

Denote by $C_{rel} : \mathbb{R}^k \to 2^{\mathbb{R}^m}$ the set-valued mapping

$$C_{rel}(t) := \{y \in \mathbb{R}^m \mid g(y) \le t\}.$$

The following is known about C_{rel} :

Theorem 3.24 (**[19, Corollary 5]**).
Assume that $C_{rel}(t) \neq \emptyset$ *holds for any* $t \in \mathbb{R}^k$ *and that* $C_{rel}(0)$ *is compact. Let* $t_1, \ldots, t_k$ *denote the components of* $t \in \mathbb{R}^k$*, then for any* $r > 0$,

$$K(r) := \sup_{t \in B_r(0),\, y \notin C_{rel}(t)} \frac{\inf\{\|y - y'\| \mid y' \in C_{rel}(t)\}}{\max\{g_j(y) - t_j \mid j = 1, \ldots, k\}}$$

is finite and such that

$$d_\infty(C_{rel}(t), C_{rel}(t')) \leq K(r)\|t - t'\| \quad \forall t, t' \in B_r(0),$$

where d_∞ *denotes the Hausdorff distance.*

In view of $[\mathbb{A}_f]$, an additional assumption allowing to bound the growth of $K(r)$ is needed:

Assumption 3.25 ($[\mathbb{A}_{Conv}]$: **Assumptions for mixed-integer convex recourse**).
$C_{rel}(0)$ *is compact,* $C_{rel}(t) \cap (\mathbb{R}^{m_1} \times \mathbb{Z}^{m_2}) \neq \emptyset$ *for any* $t \in \mathbb{R}^k$*,* h *is continuous and there exist a locally bounded mapping* $\eta_h : \mathbb{R}^n \to \mathbb{R}$ *and constants* $\gamma_h, \kappa_v, \gamma_v, \kappa_K, \gamma_K > 0$ *such that*

$$\|h(x,z)\| \leq \eta_h(x)(\|z\|^{\gamma_h} + 1), \quad |v(y)| \leq \kappa_v(\|y\|^{\gamma_v} + 1) \ \text{ and } \ K(r) \leq \kappa_K(r^{\gamma_K} + 1)$$

hold for any $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s$ *and* $r > 0$.

The above assumption admits the following result:

Theorem 3.26. *Let* f *be given by* (3.17) *and assume that* $[\mathbb{A}_c]$, $[\mathbb{A}_\rho]$ *and* $[\mathbb{A}_{Conv}]$ *are fulfilled. Set*

$$\gamma := \max\{\gamma_c, \gamma_h(\gamma_K + 1)(\gamma_v + 1)\},$$

let $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$ *be locally uniformly* $\|\cdot\|^{\gamma p}$*−integrating and assume that* $\mu_0 \in \mathcal{M}$ *is such that* $(\delta_x \otimes \mu_0)[D_{f-c}] = 0$ *holds for any* $x \in X$*. Then* $\varphi : \mathcal{M} \to \bar{\mathbb{R}}$ *defined by* (3.7) *is upper semicontinuous at* μ_0 *with respect to the relative topology of weak convergence on* $\mathcal{M}$*. If* X *is compact,* φ *is continuous at* μ_0 *and* $\Phi : \mathcal{M} \to 2^X$ *defined in* (3.8) *is upper hemicontinuous at* μ_0 *with respect to the relative topology of weak convergence on* $\mathcal{M}$*. In this case,* $\Phi(\mu_0)$ *is nonempty and compact.*

Proof. Fix any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$. Since g is continuous by Lemma 3.23, $C_{rel}(h(x,z))$ is closed and the boundedness of $C_{rel}(0)$ and Theorem 3.24 yield that $C_{rel}(h(x,z))$ is bounded. Consequently,

$$C(x,z) = C_{rel}(h(x,z)) \cap (\mathbb{R}^{m_1} \times \mathbb{Z}^{m_2})$$

is the intersection of a compact and closed set and thus compact. Furthermore, v continuous by Lemma 3.23, which implies that

$$\inf_y \{v(y) \mid y \in C(x,z)\}$$

is finite and the infimum is attained. Hence, f is real-valued and admits the representation $f(x,z) = c(x,z) + \min_y\{v(y) \mid y \in C(x,z)\}$.

C is upper hemicontinuous: Otherwise, there would exist a point $(x_0, z_0) \in \mathbb{R}^n \times \mathbb{R}^s$, an open set $\mathcal{O} \subset \mathbb{R}^m$ and sequences $\{(x_l, z_l)\}_{l\in\mathbb{N}} \subseteq \mathbb{R}^n \times \mathbb{R}^s$ and $\{y_l\}_{l\in\mathbb{N}} \subseteq \mathbb{R}^m$ such that

$$C(x_0,z_0) \subset \mathcal{O},\ (x_l,z_l) \in B_{\frac{1}{l}}(x_0,z_0),\ h(x_l,z_l) \in B_{\frac{1}{l}}(h(x_0,z_0)),\ y_l \in C(x_l,z_l) \text{ and } y_l \notin \mathcal{O}$$

hold for any $l \in \mathbb{N}$. By Theorem 3.24,

$$\sup_{l\in\mathbb{N}} d_\infty(C_{rel}(h(x_0,z_0)), C_{rel}(h(x_l,z_l))) \le K(\|h(x_0,z_0)\| + 1),$$

which implies that

$$\bigcup_{l=1}^{\infty}\{y_l\} \subseteq \bigcup_{l=1}^{\infty} C(x_l,z_l) \subseteq \bigcup_{l=1}^{\infty} C_{rel}(h(x_l,z_l)) \subseteq \{C_{rel}(h(x_0,z_0))\} + B_{K(\|h(x_0,z_0)\|+1)}(0)$$

is bounded. Consequently, $y_l \to \bar{y}$ for some $\bar{y} \in \mathbb{R}^m$ can be assumed without loss of generality. $y_l \in C(x_l,z_l) \subseteq \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2}$ holds for any $l \in \mathbb{N}$ and implies $\bar{y} \in \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2}$. Furthermore, by $h(x_l,z_l) \to h(x_0,z_0)$, Theorem 3.24 yields that $\bar{y} \in C_{rel}(h(x_0,z_0))$. Thus,

$$\bar{y} \in C_{rel}(h(x_0,z_0)) \cap (\mathbb{R}^{m_1} \times \mathbb{Z}^{m_2}) = C(x_0,z_0) \subset \mathcal{O}.$$

On the other hand, $y_l \notin \mathcal{O}$ for any $l \in \mathbb{N}$ and $\mathcal{O}$ is open, which yields the contradiction $\bar{y} \notin \mathcal{O}$. Hence, C is upper hemicontinuous.

Since h and v are continuous, the upper hemicontinuity of C allows to apply Theorem 3.7 to conclude that $f - c$ is lower semicontinuous. By $[\mathbb{A}_c]$, the latter implies that f is lower semicontinuous and hence Borel measurable.

Let $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$ be fixed. Based on the considerations above, there exist $y^* \in C(x,z)$ and $y_0^* \in C(0,0)$ satisfying $f(x,z) = c(x,z) + v(y^*)$ and $f(0,0) = c(0,0) + v(y_0^*)$. Since $C_{rel}(h(0,0))$ is compact,

$$d_0 := \max\{\|y - y'\| \mid y, y' \in C_{rel}(h(0,0))\}$$

is finite. Theorem 3.24 and assumption $[\mathbb{A}_{Conv}]$ yield

$$\begin{aligned}\|y^* - y_0^*\| &\le d_\infty(C_{rel}(h(x,z)), C_{rel}(h(0,0))) + d_0\\ &\le K(\|h(x,z)\| + \|h(0,0)\| + 1)(\|h(x,z)\| + \|h(0,0)\|) + d_0\\ &\le 2\kappa_K(\|h(x,z)\| + \|h(0,0)\| + 1)^{\gamma_K+1} + d_0\\ &\le \kappa^*(\|h(x,z)\|^{\gamma_K+1} + 1),\end{aligned}$$

where $\kappa^* := 2^{\gamma_K+2}\kappa_K(\|h(0,0)\| + 1)^{\gamma_K+1} + d_0$. Let $L_v(\|y^* - y_0^*\| + \|y_0^*\| + 1)$ be given by (3.18). $[\mathbb{A}_{Conv}]$ implies

$$\begin{aligned}L_v(\|y^* - y_0^*\| + \|y_0^*\| + 1) &\le \max_{y \in \{\pm 2(\|y^* - y_0^*\| + \|y_0^*\| + 1)\}^m} 2|v(y)| + 4|v(0)|\\ &\le 2\kappa_v(2\sqrt{m})^{\gamma_v}(\|y^* - y_0^*\| + \|y_0^*\| + 1)^{\gamma_v} + 2\kappa_v + 4|v(0)|\\ &\le \kappa_L(\|y^* - y_0^*\|^{\gamma_v} + 1),\end{aligned}$$

where $\kappa_L := 2\kappa_v(4\sqrt{m})^{\gamma_v}(\|y_0^*\| + 1)^{\gamma_v} + 2\kappa_v + 4|v(0)|$. Since

$$\|y^*\| < \|y^* - y_0^*\| + \|y_0^*\| + 1,$$

Lemma 3.23 and the above inequalities yield

$$\begin{aligned}|(f-c)(x,z)| &\le |v(y^*) - v(y_0^*)| + |v(y_0^*)|\\ &\le L_v(\|y^* - y_0^*\| + \|y_0^*\| + 1)\|y^* - y_0^*\| + |v(y_0^*)|\\ &\le 2\kappa_L\|y^* - y_0^*\|^{\gamma_v+1} + 2\kappa_L + |v(y_0^*)|\\ &\le 2\kappa_L(\kappa^*)^{\gamma_v+1}(\|h(x,z)\|^{\gamma_K+1} + 1)^{\gamma_v+1} + 2\kappa_L + |v(y_0^*)|\\ &\le \bar{\kappa}(\|h(x,z)\|^{(\gamma_K+1)(\gamma_v+1)} + 1),\end{aligned}$$

where $\bar{\kappa} := 2^{\gamma_v+2}\kappa_L(\kappa^*)^{\gamma_v+1} + 2\kappa_L + |v(y_0^*)|$. Finally, $[\mathbb{A}_{Conv}]$ implies

$$\begin{aligned}|(f-c)(x,z)| &\le \bar{\kappa}\eta_h(x)^{(\gamma_K+1)(\gamma_v+1)}(\|z\|^{\gamma_h} + 1)^{(\gamma_K+1)(\gamma_v+1)} + \bar{\kappa}\\ &\le \eta_{f-c}(x)(\|z\|^{\gamma_h(\gamma_K+1)(\gamma_v+1)} + 1),\end{aligned}$$

where $\eta_{f-c}(x) := \bar{\kappa}(2\eta_h(x))^{(\gamma_K+1)(\gamma_v+1)} + \bar{\kappa}$ is locally bounded. Consequently, $[\mathbb{A}_f]$ is fulfilled for f with exponent γ by Lemma 3.5 and Theorem 2.84 is applicable. The stated stability is a direct conclusion from Lemma 3.8 and Theorem 3.7. □

If all variables in the recourse problem are continuous, the assumption $(\delta_x \otimes \mu_0)[D_{f-c}] = 0$

of Theorem 3.26 holds automatically:

Remark 3.27. *Assume that* $[\mathbb{A}_c]$*,* $[\mathbb{A}_\rho]$ *and* $[\mathbb{A}_{Conv}]$ *are fulfilled and that there are no integrality constraints in the description of* C*, i.e. that* $C(x,z) = C_{rel}(h(x,z))$ *holds for any* $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$*. Then* C *is both upper and lower hemicontinuous by Theorem 3.24. Thus, the mapping* f *given by* (3.17) *is continuous by Theorem 3.7. Consequently,* $D_{f-c} = \emptyset$*.*

4. Stability in stochastic bilevel programming

This chapter applies the results of section 2.5 to derive weak continuity of functionals arising from stochastic bilevel problems with mean-risk objective functions. In this setting, the function f is essentially given by the optimal value function of a parametric problem, where only optimal solutions to the lower level problem are feasible. This results in weaker analytical properties and poses additional difficulties in view of the verification of assumption $[\mathbb{A}_f]$.

The framework of mean-risk stochastic bilevel programming is introduced in section 4.1, while section 4.2 is devoted to the case where the lower level problem is quadratic and uniquely solvable. Finally, section 4.3 examines the situation where the lower level problem is quadratic and allowed to have more than a single solution. However, in section 4.3 it is assumed that the randomness does not affect the lower level objective function. In the vein of chapter 3, sufficient conditions for stability with respect to perturbations of the underlying measure are discussed.

Parts of this chapter have also been submitted for publication (see [68] for a preprint).

4.1. Mean-risk stochastic bilevel problems

Consider the parametric bilevel optimization problem

$$\text{"}\min_x\text{"}\{c(x,z) + q(x,y,z) \mid x \in X, y \in C(x,z)\}, \tag{4.1}$$

where the leader variable x is to be chosen from a fixed nonempty set $X \subseteq \mathbb{R}^n$ and the upper level objective function is given as the sum of the mappings $c : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ and $q : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s \to \mathbb{R}$. In (4.1), $z \in \mathbb{R}^s$ is a parameter, while y reflects the follower's decision and is an optimal solution to the lower level problem given by the multifunction

$C : \mathbb{R}^n \times \mathbb{R}^s \to 2^{\mathbb{R}^m}$,

$$C(x, z) = \operatorname{argmin}_y \{y^\top Dy + d(x, z)^\top y \mid Ay \leq h(x, z)\},$$

which involves matrices $A \in \mathbb{R}^{k \times m}$ and $D \in \mathbb{R}^{m \times m}$ and mappings $d : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ and $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^k$. Without loss of generality, D is assumed to be symmetric. Note that the minimum in (4.1) is only taken with respect to x, which reflects the assumption that the decision on y is made by a different actor (the follower) who is able to observe x beforehand. Since the quadratic program defining C may have more than one optimal solution, problem (4.1) is not well defined in general. In bilevel optimization, this issue is typically resolved by considering either the best (optimistic approach) or the worst (pessimistic approach) y with respect to the upper level objective function. While the optimistic approach yields

$$P_{opt}(z) \qquad \min_x \{\underbrace{c(x, z) + \min_y \{q(x, y, z) \mid y \in C(x, z)\}}_{=: f_{opt}(x,z)} \mid x \in X\},$$

the pessimistic one results in the problem

$$\begin{aligned} P_{pes}(z) \qquad & \min_x \{c(x, z) + \max_y \{q(x, y, z) \mid y \in C(x, z)\} \mid x \in X\} \\ = \; & \min_x \{\underbrace{c(x, z) - \min_y \{-q(x, y, z) \mid y \in C(x, z)\}}_{=: f_{pes}(x,z)} \mid x \in X\}. \end{aligned}$$

Both $P_{opt}(z)$ and $P_{opt}(z)$ are well defined and still depend on the parameter z. Under stochastic data uncertainty and an information constraint, these programs give rise to stochastic bilevel problems: For the following analysis, z is assumed to be the realization of a known random vector $Z(\omega) : \Omega' \to \mathbb{R}^s$ defined on some probability space $(\Omega', \mathcal{F}', \mathbb{P}')$. Depending on which interplay between decision and observation is assumed, three settings are possible:

(a) $Z(\omega)$ can be observed before deciding on x. In this case, $P_{opt}(Z(\omega))$ and $P_{pes}(Z(\omega))$ boil down to deterministic bilevel problems.

(b) If both the leader and the follower have to make their decisions without knowledge of $Z(\omega)$, the lower level problem turns into a one-stage stochastic program. The follower has various possibilities of handling the stochastic uncertainty, e.g. via mean-risk models, chance constraints or models based on stochastic dominance. For any of the resulting well defined problems, the set $\bar{C}(x)$ of optimal solutions only depends on the leader's decision x. Assuming that the leader is aware of the follower's model, $P_{opt}(Z(\omega))$ and $P_{pes}(Z(\omega))$ turn into one-stage stochastic bilevel problems of the form

$$\min_x \{c(x, Z(\omega)) \pm \min_y \{\pm q(x, y, Z(\omega)) \mid y \in \bar{C}(x)\} \mid x \in X\},$$

where the objective function is subject to stochastic uncertainty.

(c) The leader has to decide on x without knowledge of $Z(\omega)$, while the follower solves the lower level problem knowing both x and the realization of the parameter $Z(\omega)$. The resulting problems $P_{opt}(Z(\omega))$ and $P_{opt}(Z(\omega))$ bear close structural similarities to the problems considered in chapter 3 and take the form

$$\min_x \{f(x, Z(\omega)) \mid x \in X\}, \tag{4.2}$$

where $f : \mathbb{R}^n \times \mathbb{R}^s \to \bar{\mathbb{R}}$ is given by $f = f_{opt}$ or $f = f_{pes}$, depending on which approach is considered.

Throughout this chapter, setting (c) and purely exogenous stochasticity shall be assumed. Note that this interplay between decision and observation is also considered in [9] and [131].

Remark 4.1. *Classical two-stage stochastic problems can be seen as special stochastic bilevel programs, where the optimistic approach is taken and every feasible point of the lower level is optimal. On the other hand,* (4.2) *can be understood as a two-stage problem, where the recourse is given by an optimization problem over the set of optimal solutions to the lower level problem.*

Since the decision on x has to be made without knowledge of $Z(\omega)$, the problem in (4.2) is not well defined. However, under assumptions $[\mathbb{A}_f]$ and $[\mathbb{A}_\rho]$, it gives rise to the well defined mean-risk problem

$$\min_x \{Q(x, \mu) \mid x \in X\}, \tag{4.3}$$

where $\mu = \mathbb{P}' \circ Z^{-1} \in \mathcal{P}(\mathbb{R}^s)$ and the objective function

$$Q(x, \mu) = \mathcal{R}_\rho((\delta_x \otimes \mu) \circ f^{-1})$$

is exactly as in (2.4) (see section 3.1 for details). The following analysis examines the behavior of (4.3) under perturbations of μ with respect to the topology of weak convergence. Let φ be given by (3.7) and denote the optimal value function of problem (4.3) with respect to the parameter μ. In the same vein, let the optimal solution set mapping Φ be given by (3.8).

The subsequent sections focus on identifying sufficient conditions on the data of problem (4.1), under which assumption $[\mathbb{A}_f]$ automatically holds for f_{opt} or f_{pes}. In such situations, Theorem 2.84 yields continuity of a restriction of Q with respect to the product topology of the standard topology on $\mathbb{R}^n$ and the relative topology of weak convergence on a suitable subset $\mathcal{M}$ of $\mathcal{P}(\mathbb{R}^s)$. Theorem 3.7 and Lemma 3.8 then allow for immediate conclusions about φ and Φ.

As before, Lemma 3.5 allows to work with $f - c$ instead of f whenever assumption $[\mathbb{A}_c]$ is fulfilled. In addition, the following assumption will be imposed on q:

Assumption 4.2 ($[\mathbb{A}_q]$: **Assumptions on** q)**.**
q is continuous and there exist a locally bounded mapping $\eta_q : \mathbb{R}^n \to \mathbb{R}$ and constants $\gamma_{q,y}, \gamma_{q,z} \geq 0$ such that

$$|q(x, y, z)| \leq \eta_q(x)(\|y\|^{\gamma_{q,y}} + 1)(\|z\|^{\gamma_{q,z}} + 1)$$

holds for any $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s$.

4.2. Quadratic lower level problems with unique solutions

This section examines the case where the lower level problem is always uniquely solvable. The following result will be helpful:

Theorem 4.3 (**[136, Corollary 5.1]**)**.**
Let

$$\text{dom } C^* := \{(t_1, t_2) \in \mathbb{R}^m \times \mathbb{R}^k \mid C^*(t_1, t_2) \neq \emptyset\}$$

denote the domain of the set-valued mapping $C^* : \mathbb{R}^m \times \mathbb{R}^k \to 2^{\mathbb{R}^m}$ given by

$$C^*(t_1, t_2) := \text{argmin}_y\{y^\top Dy + t_1^\top y \mid Ay \leq t_2\}.$$

Assume that $\mathcal{C}$ is a convex subset of dom C^* *on which C^* is single-valued. Then C^* is Lipschitzian on $\mathcal{C}$, i.e. there exists a constant $L^* > 0$ such that for any $(t_1, t_2), (t_1', t_2') \in \mathcal{C}$, it holds that*

$$\|y - y'\| \leq L^*\|(t_1, t_2) - (t_1', t_2')\|,$$

where $\{y\} = C^*(t_1, t_2)$ and $\{y'\} = C^*(t_1', t_2')$.

Theorem 4.3 motivates the following assumption:

Assumption 4.4 ($[\mathbb{A}_{ULLS}]$: **Assumptions for unique lower level solution).**
C^* is single-valued on the convex hull of

$$\{(d(x,z), h(x,z)) \mid (x,z) \in \mathbb{R}^n \times \mathbb{R}^s\}.$$

Furthermore, d and h are continuous and there exist a constant $\gamma_{d,h} \geq 0$ and a locally bounded mapping $\eta_{d,h} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\|(d(x,z), h(x,z))\| \leq \eta_{d,h}(x)(\|z\|^{\gamma_{d,h}} + 1)$$

holds for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$.

Remark 4.5. *Since assumption $[\mathbb{A}_{ULLS}]$ implies $f_{opt} = f_{pes}$, it is not necessary to distinguish between the optimistic and the pessimistic approach.*

Remark 4.6. *If D is positive definite, the mapping $y \mapsto y^\top Dy + t_1^\top y$ is strictly convex by [177, Theorem 2.14] and $\inf_{y \in \mathbb{R}^m} y^\top Dy + t_1^\top y > -\infty$ for any $t_1 \in \mathbb{R}^m$. Consequently, C^* is single-valued on the convex set*

$$\mathbb{R}^m \times \{t_2 \in \mathbb{R}^k \mid \{y \in \mathbb{R}^m \mid Ay \leq t_2\} \neq \emptyset\}$$

(see e.g. the theorem in [54]). In this case, the first part of assumption $[\mathbb{A}_{ULLS}]$ can weakened to

$$\{y \in \mathbb{R}^m \mid Ay \leq h(x,z)\} \neq \emptyset$$

for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$.

Assumption $[\mathbb{A}_{ULLS}]$ admits the following result concerning stability of the stochastic bilevel problem (4.3):

Theorem 4.7. *Assume that $[\mathbb{A}_c]$, $[\mathbb{A}_q]$, $[\mathbb{A}_\rho]$ and $[\mathbb{A}_{ULLS}]$ are fulfilled and let f be given by $f = f_{opt} = f_{pes}$. Set $\gamma := \max\{\gamma_c, \gamma_{d,h}\gamma_{q,y} + \gamma_{q,z}\}$ and let $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$ be locally uniformly $\|\cdot\|^{\gamma p}$-integrating. Then the optimal value function φ defined in* (3.7) *is upper semicontinuous with respect to the relative topology of weak convergence on $\mathcal{M}$. If X is compact, φ is continuous and the optimal solution set mapping $\Phi : \mathcal{M} \to 2^X \setminus \{\emptyset\}$ defined in* (3.8) *is compact-valued and upper hemicontinuous with respect to the relative topology of weak convergence on $\mathcal{M}$.*

Proof. By assumption $[\mathbb{A}_{ULLS}]$ and Theorem 4.3, there is a Lipschitz continuous mapping $y^* : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m$ satisfying $C(x,z) = \{y^*(d(x,z), h(x,z))\}$ for any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$. Consequently, by assumptions $[\mathbb{A}_c]$ and $[\mathbb{A}_q]$,

$$f(x,z) = c(x,z) + q(x, y^*(d(x,z), h(x,z)), z)$$

is continuous on $\mathbb{R}^n \times \mathbb{R}^s$ and hence Borel measurable. Fix any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$ and let L^* denote the Lipschitz constant from Theorem 4.3. Assumption $[\mathbb{A}_{ULLS}]$ implies

$$\begin{aligned}
\|y^*(d(x,z), h(x,z))\| &\le L^* \|(d(x,z), h(x,z))\| + L^* \|(d(0,0), h(0,0))\| + \|y^*(d(0,0), h(0,0))\| \\
&\le L^* \eta_{d,h}(x)(\|z\|^{\gamma_{d,h}} + 1) + L^* \|(d(0,0), h(0,0))\| + \|y^*(d(0,0), h(0,0))\| \\
&\le \eta^*(x)(\|z\|^{\gamma_{d,h}} + 1),
\end{aligned}$$

where $\eta^*(x) := L^* \eta_{d,h} + L^* \|(d(0,0), h(0,0))\| + \|y^*(d(0,0), h(0,0))\|$ is locally bounded. In combination with assumption $[\mathbb{A}_q]$, the latter yields

$$\begin{aligned}
|(f-c)(x,z)| &\le \eta_q(x)(\|y^*(d(x,z), h(x,z))\|^{\gamma_{q,y}} + 1)(\|z\|^{\gamma_{q,z}} + 1) \\
&\le \eta_q(x)((2\eta^*(x))^{\gamma_{q,y}} + 1)(\|z\|^{\gamma_{d,h}\gamma_{q,y}} + 1)(\|z\|^{\gamma_{q,z}} + 1) \\
&\le \eta_{f-c}(x)(\|z\|^{\gamma_{d,h}\gamma_{q,y} + \gamma_{q,z}} + 1),
\end{aligned}$$

where $\eta_{f-c}(x) := 3\eta_q(x)((2\eta^*(x))^{\gamma_{q,y}} + 1)$ is locally bounded. Consequently, $[\mathbb{A}_f]$ is fulfilled for f with exponent γ by assumption $[\mathbb{A}_c]$ and Lemma 3.5. Since $D_f = \emptyset$, Theorem 2.84 is applicable and the stated stability is a direct conclusion from Lemma 3.8 and Theorem 3.7. $\square$

4.3. Quadratic lower level problems with random right-hand side

In this section, the case where the lower level problem may have more than one solution is considered. However, it is assumed that only the right-hand side of the inequalities describing the feasible set of the lower level depends on x and z, while the mapping d in the lower level objective function is constant. Consequently, f takes the form

$$f(x,z) = c(x,z) \pm \min_y \{\pm q(x,y,z) \mid y \in \operatorname{argmin}_{y'} \{y'^\top D y' + d_0^\top y' \mid Ay' \le h(x,z)\}\}, \quad (4.4)$$

where $d_0 \in \mathbb{R}^m$ is fixed. The following result will be used to analyze f:

Theorem 4.8 ([**138, Theorem 4.2**]).
If D positive semidefinite, the set-valued mapping $\hat{C} : \mathbb{R}^k \to 2^{\mathbb{R}^m}$ given by

$$\hat{C}(t) := \operatorname{argmin}_y \{y^\top Dy + d_0^\top y \mid Ay \leq t\}$$

is Lipschitzian on $\operatorname{dom} \hat{C} := \{t \in \mathbb{R}^k \mid \hat{C}(t) \neq \emptyset\}$*, i.e. there exists a constant $\hat{L} > 0$ such that*

$$d_\infty(\hat{C}(t), \hat{C}(t')) \leq \hat{L}\|t - t'\|$$

holds for any $t, t' \in \operatorname{dom} \hat{C}$.

Theorem 4.8 motivates the following assumptions. Note that the assumption imposed on q is more restrictive than $[\mathbb{A}_q]$:

Assumption 4.9 ($[\mathbb{A}_{RRHS}]$: **Assumptions for random right-hand side**).
There exist constants $\gamma^*_{q,y}, \gamma^*_{q,z} > 0$ and a continuous mapping $\eta^*_q : \mathbb{R}^n \to \mathbb{R}$ such that

$$|q(x, y, z) - q(x', y', z')| \leq \eta^*_q(x - x')(\|y - y'\|^{\gamma^*_{q,y}} + \|z - z'\|^{\gamma^*_{q,z}})$$

holds for any $(x, y, z), (x', y', z') \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s$. Furthermore, D is positive semidefinite, $\{y \in \mathbb{R}^m \mid Ay \leq h(x, z)\} \neq \emptyset$ holds for any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$ and there exists a vector $t_0 \in \mathbb{R}^k$ such that

$$|\inf_y \{y^\top Dy + d_0^\top y \mid Ay \leq t_0\}| < \infty. \tag{4.5}$$

In addition, h is continuous and there exist a constant $\gamma_h \geq 0$ and a locally bounded mapping $\eta_h : \mathbb{R}^n \to \mathbb{R}$ such that

$$\|h(x, z)\| \leq \eta_h(x)(\|z\|^{\gamma_h} + 1) \quad \forall (x, z) \in \mathbb{R}^n \times \mathbb{R}^s.$$

For the optimistic approach $f = f_{opt}$, further assume that $\min_y\{q(x, y, z) \mid y \in C(x, z)\}$ *is solvable for any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$ and* $\sup_y\{q(x_0, y, z_0) \mid y \in C(x_0, z_0)\}$ *is finite for some $(x_0, z_0) \in \mathbb{R}^n \times \mathbb{R}^s$.*

For the pessimistic approach $f = f_{pes}$, assume that $\min_y\{-q(x, y, z) \mid y \in C(x, z)\}$ *is solvable for any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$ and* $\sup_y\{-q(x_0, y, z_0) \mid y \in C(x_0, z_0)\}$ *is finite for some $(x_0, z_0) \in \mathbb{R}^n \times \mathbb{R}^s$.*

Remark 4.10. *By the existence theorem of quadratic programming (see e.g. [54]),* (4.5) *yields $\hat{C}(t_0) \neq \emptyset$. The latter implies*

$$\operatorname{dom} \hat{C} = \{t \in \mathbb{R}^k \mid \{y \in \mathbb{R}^m \mid Ay \leq t\} \neq \emptyset\},$$

i.e. the solvability of the lower level problem whenever its feasible set is nonempty. Furthermore, there exists a finite set $E \subset \mathbb{R}^m$ such that for any $t \in \operatorname{dom} \hat{C}$, $\hat{C}(t)$ is a polyhedron having exactly the elements of E as its extreme directions (see [95]).

Remark 4.11. *Consider the case where $q(x, y, z) = q_0^\top y$ for some fixed vector $q_0 \in \mathbb{R}^m$. If there exists some $t_0 \in \mathbb{R}^k$ such that $\max_y\{|q_0^\top y| \mid y \in \hat{C}(t_0)\}$ is solvable, Remark 4.10 implies $q_0^\top e = 0$ for any $e \in E$ and hence the solvability of $\min_y\{\pm q_0^\top y \mid y \in \hat{C}(t)\}$ for any $t \in \mathbb{R}^k$.*

Remark 4.12. *Assumption $[\mathbb{A}_{RRHS}]$ does not exclude the linear case $D = 0 \in \mathbb{R}^{m\times m}$, since only positive semidefiniteness is assumed.*

Assumption $[\mathbb{A}_{RRHS}]$ admits the following stability result for the stochastic bilevel problem (4.3):

Theorem 4.13. *Let f be given by (4.4) and assume that $[\mathbb{A}_c]$, $[\mathbb{A}_\rho]$ and $[\mathbb{A}_{RRHS}]$ are fulfilled. Set $\gamma := \max\{\gamma_c, \gamma_h\gamma^*_{q,y} + \gamma^*_{q,z}\}$ and let $\mathcal{M} \subseteq \mathcal{M}^{\gamma p}_s$ be a locally uniformly $\|\cdot\|^{\gamma p}-$integrating set. Then the optimal value function φ defined in (3.7) is upper semicontinuous with respect to the relative topology of weak convergence on $\mathcal{M}$. If X is compact, φ is continuous and the optimal solution set mapping $\Phi : \mathcal{M} \to 2^X \setminus \{\emptyset\}$ defined in (3.8) is compact-valued and upper hemicontinuous with respect to the relative topology of weak convergence on $\mathcal{M}$.*

Proof. $f - c$ is continuous: Consider any converging sequence $\{(x_l, z_l)\}_{l\in\mathbb{N}} \subset \mathbb{R}^n \times \mathbb{R}^s$ and denote its limit by $(\bar{x}, \bar{z}) := \lim_{l\to\infty}(x_l, z_l)$. By assumption $[\mathbb{A}_{RRHS}]$ and Remark 4.10, there exists a vector $\bar{y} \in C(\bar{x}, \bar{z})$ satisfying $(f - c)(\bar{x}, \bar{z}) = q(\bar{x}, \bar{y}, \bar{z})$. Furthermore, since

$$\{h(x, z) \mid (x, z) \in \mathbb{R}^n \times \mathbb{R}^s\} \subseteq \operatorname{dom} \hat{C},$$

Theorem 4.8 yields $d_\infty(C(\bar{x}, \bar{z}), C(x_l, z_l)) \leq \hat{L}\|h(\bar{x}, \bar{z}) - h(x_l, z_l)\|$, which, by the continuity of h, implies the existence of a sequence $\{y_l\}_{l\in\mathbb{N}}$ satisfying $\lim_{l\to\infty} y_l = \bar{y}$ and $y_l \in C(x_l, z_l)$ for any $l \in \mathbb{N}$. In the optimistic setting, the latter yields $(f - c)(x_l, z_l) \leq q(x_l, y_l, z_l)$ for any $l \in \mathbb{N}$ and hence

$$\limsup_{l\to\infty}(f - c)(x_l, z_l) \leq \limsup_{l\to\infty} q(x_l, y_l, z_l) = q(\bar{x}, \bar{y}, \bar{z}) = (f - c)(\bar{x}, \bar{z}),$$

i.e. the upper semicontinuity of $f - c$. In the pessimistic setting, a similar argument shows that $f - c$ is lower semicontinuous.

Again by assumption $[\mathbb{A}_{RRHS}]$ and Theorem 4.8, there exist sequences $\{y_l^*\}_{l\in\mathbb{N}}$ and $\{\bar{y}_l\}_{l\in\mathbb{N}}$ such that $y_l^* \in C(x_l, z_l)$, $\bar{y}_l \in C(\bar{x}, \bar{z})$, $(f-c)(x_l, z_l) = q(x_l, y_l^*, z_l)$ and

$$\|y_l^* - \bar{y}_l\| \leq \hat{L}\|h(\bar{x}, \bar{z}) - h(x_l, z_l)\|$$

hold for any $l \in \mathbb{N}$. Consequently, in the optimistic setting,

$$(f-c)(\bar{x}, \bar{z})-(f-c)(x_l, z_l) \leq q(\bar{x}, \bar{y}_l, \bar{z})-q(x_l, y_l^*, z_l) \leq \eta_q^*(\bar{x}-x_l)(\|\bar{y}_l-y_l^*\|^{\gamma_{q,y}^*}+\|\bar{z}-z_l\|^{\gamma_{q,z}^*})$$

holds for any $l \in \mathbb{N}$, which implies $(f-c)(\bar{x}, \bar{z}) - \liminf_{l\to\infty}(f-c)(x_l, z_l) \leq 0$ and hence the lower semicontinuity of $f-c$. In the pessimistic setting, a similar argument shows that $f-c$ is upper semicontinuous.

By $[\mathbb{A}_c]$ and the considerations above, f is continuous and hence Borel measurable. Fix an arbitrary vector $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$ and let $y \in C(x, z)$ be such that $(f-c)(x, z) = q(x, y, z)$. By Theorem 4.8, there exists a vector $y_0 \in C(x_0, y_0)$ such that

$$\begin{aligned} \|y - y_0\| &\leq \hat{L}\|h(x, z) - h(x_0, z_0)\| \\ &\leq \hat{\eta}(x)(\|z\|^{\gamma_h} + 1), \end{aligned}$$

where $\hat{\eta}(x) := \hat{L}(\eta_h(x) + \|h(x_0, z_0)\|)$ is locally bounded. By assumption $[\mathbb{A}_{RRHS}]$, the constant $\check{q} := \max_{y'}\{|q(x_0, y', z_0)| \mid y' \in C(x_0, z_0)\}$ is finite. Hence,

$$\begin{aligned} |(f-c)(x, z)| &\leq |q(x, y, z) - q(x_0, y_0, z_0)| + \check{q} \\ &\leq \eta_q^*(x - x_0)(\|y - y_0\|^{\gamma_{q,y}^*} + \|z - z_0\|^{\gamma_{q,z}^*}) + \check{q} \\ &\leq \eta_{f-c}(x)(\|z\|^{\gamma_h\gamma_{q,y}^*+\gamma_{q,z}^*} + 1), \end{aligned}$$

where $\eta_{f-c}(x) := 2\eta_q^*(x - x_0)((2\hat{\eta}(x))^{\gamma_{q,y}^*} + 2^{\gamma_{q,z}^*}(\|z_0\|^{\gamma_{q,y}^*} + 1)) + \check{q}$ is locally bounded. Consequently, $[\mathbb{A}_f]$ is fulfilled for f with exponent γ by assumption $[\mathbb{A}_c]$ and Lemma 3.5. Since $D_f = \emptyset$, Theorem 2.84 is applicable and the stated stability is a direct conclusion from Lemma 3.8 and Theorem 3.7. □

If the set of optimal solutions to the lower level problem is compact, a weaker assumption on q is sufficient:

Assumption 4.14 ($[\mathbb{A}_{Alt}]$: **Alternative assumptions for random right-hand side**). *D is positive semidefinite, $\{y \in \mathbb{R}^m \mid Ay \leq h(x, z)\} \neq \emptyset$ holds for any $(x, z) \in \mathbb{R}^n \times \mathbb{R}^s$*

and there exists a vector $t_0 \in \mathbb{R}^k$ *such that* $\hat{C}(t_0)$ *is bounded and*

$$|\inf_y\{y^\top Dy + d_0^\top y \mid Ay \leq t_0\}| < \infty.$$

Furthermore, h *is continuous and there exist a constant* $\gamma_h \geq 0$ *and a locally bounded mapping* $\eta_h : \mathbb{R}^n \to \mathbb{R}$ *such that*

$$\|h(x,z)\| \leq \eta_h(x)(\|z\|^{\gamma_h} + 1) \quad \forall (x,z) \in \mathbb{R}^n \times \mathbb{R}^s.$$

Under assumption $[\mathbb{A}_{Alt}]$, the following holds for problem (4.3):

Theorem 4.15. *Let* f *be given by* (4.4) *and assume that* $[\mathbb{A}_c]$, $[\mathbb{A}_q]$, $[\mathbb{A}_\rho]$ *and* $[\mathbb{A}_{Alt}]$ *are fulfilled. Set* $\gamma := \max\{\gamma_c, \gamma_h\gamma_{q,y} + \gamma_{q,z}\}$ *and let* $\mathcal{M} \subseteq \mathcal{M}_s^{\gamma p}$ *be locally uniformly* $\|\cdot\|^{\gamma p}$*−integrating. Then the optimal value function* φ *defined in* (3.7) *is upper semicontinuous with respect to the relative topology of weak convergence on* $\mathcal{M}$*. If* X *is compact,* φ *is continuous and the optimal solution set mapping* $\Phi : \mathcal{M} \to 2^X \setminus \{\emptyset\}$ *defined in* (3.8) *is compact-valued and upper hemicontinuous with respect to the relative topology of weak convergence on* $\mathcal{M}$*.*

Proof. Under assumption $[\mathbb{A}_{Alt}]$, the mapping C is compact-valued and both upper and lower hemicontinuous on $\mathbb{R}^n \times \mathbb{R}^s$ by Theorem 4.8 and Remark 4.10. Furthermore, c and q are continuous by assumptions $[\mathbb{A}_c]$ and $[\mathbb{A}_q]$. Consequently,

$$f(x,z) = c(x,z) \pm \min_y\{\pm q(x,y,z) \mid y \in C(x,z)\}$$

is continuous by Theorem 3.7 and hence Borel measurable. Set

$$\kappa_0 := \max\{\|y_0'\| \mid y_0' \in C(0,0)\} < \infty$$

and fix any $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$ and let $y \in C(x,z)$ be such that $f(x,z) = c(x,z) + q(x,y,z)$. By Theorem 4.8, there exists a vector $y_0 \in C(0,0)$ such that

$$\|y - y_0\| \ \leq \ \hat{\eta}(x)(\|z\|^{\gamma_h} + 1),$$

where $\hat{\eta}(x) := \hat{L}(\eta_h(x) + \|h(0,0)\|)$ is locally bounded. Thus, by assumption $[\mathbb{A}_q]$,

$$\begin{aligned}
|(f-c)(x,z)| &= |q(x,y,z)| \;\leq\; \eta_q(x)(\|y-y_0\|^{\gamma_{q,y}} + \kappa_0^{\gamma_{q,y}} + 1)(\|z\|^{\gamma_{q,z}} + 1) \\
&\leq \eta_q(x)(\kappa_0^{\gamma_{q,y}} + 1)(\hat{\eta}(x)^{\gamma_{q,y}}(\|z\|^{\gamma_h} + 1)^{\gamma_{q,y}} + 1)(\|z\|^{\gamma_{q,z}} + 1) \\
&\leq 2^{\gamma_{q,y}}\eta_q(x)(\kappa_0^{\gamma_{q,y}} + 1)(2^{\gamma_{q,y}}\hat{\eta}(x)^{\gamma_{q,y}} + 1)(\|z\|^{\gamma_h\gamma_{q,y}} + 1)(\|z\|^{\gamma_{q,z}} + 1) \\
&\leq \eta_{f-c}(x)(\|z\|^{\gamma_h\gamma_{q,y}+\gamma_{q,z}} + 1)
\end{aligned}$$

$\eta_{f-c}(x) := 3\eta_q(x)2^{\gamma_{q,y}}(\kappa_0^{\gamma_{q,y}}+1)(2^{\gamma_{q,y}}\hat{\eta}(x)^{\gamma_{q,y}}+1)$ is locally bounded. Consequently, $[\mathbb{A}_f]$ is fulfilled for f with exponent γ by assumption $[\mathbb{A}_c]$ and Lemma 3.5. Since $D_f = \emptyset$, Theorem 2.84, Lemma 3.8 and Theorem 3.7 are applicable and yield the stated stability. □

A. Appendix

In the following, selected results of relevance for the proofs in the present thesis are recalled for the convenience of the reader.

Theorem A.1 (**Change of variable**, see e.g. Theorem 16.13 in [48])**.**
Let $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ be measurable spaces and fix a measure ν on $(\Omega_1, \mathcal{F}_1)$. For any $(\mathcal{F}_1, \mathcal{F}_2)-$measurable mapping $T : \Omega_1 \to \Omega_2$, let $\nu \circ T^{-1}$ denote the image measure of ν under T, i.e.

$$(\nu \circ T^{-1})[B] = \nu[T^{-1}(B)] \quad \forall B \in \mathcal{F}_2.$$

Then

$$\int_{\Omega_2} g(r)\ (\nu \circ T^{-1})(dr) = \int_{\Omega_1} (g \circ T)(t)\ \nu(dt)$$

holds for any nonnegative, $(\mathcal{F}_2, \mathcal{B}(\mathbb{R}))-$measurable function $g : \Omega_2 \to \mathbb{R}$.

Theorem A.2 (**Vitali's theorem**, see e.g. Proposition 3.12 in [134])**.**
Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as well as a constant $p \geq 0$ and consider a sequence $\{Y_n\}_{n\in\mathbb{N}} \subseteq L^p(\Omega, \mathcal{F}, \mathbb{P})$. If $\{Y_n\}_{n\in\mathbb{N}}$ converges in probability to Y_1, the following statements are equivalent:

(a) $\lim_{n\to\infty} \|Y_n - Y_1\|_p = 0$.

(b) $\lim_{n\to\infty} \|Y_n\|_p = \|Y_1\|_p$.

(c) The random variables $|Y_n|^p$, $n \in \mathbb{N}$ are uniformly integrable.

Conversely, $\{Y_n\}_{n\in\mathbb{N}}$ converges in probability to Y_1 whenever (a) holds.

B. List of symbols

Functions and measures

δ_x	the Dirac measure at x, see e.g. [7, Definition 12.17]
λ^k	the Lebesgue measure on $\mathbb{R}^k$
λ^k_A	the restriction of λ^k to $A \subseteq \mathbb{R}^k$
$\mu \otimes \nu$	the product probability measure of probability measures μ and ν
F_σ	the distribution function induced by $\sigma \in \mathcal{P}(\mathbb{R})$ via $F_\sigma(t) = \sigma[(-\infty, t]]$
F_σ^{-1}	the quantile function associated with $\sigma \in \mathcal{P}(\mathbb{R})$, see (2.2)
$\mathcal{R}_\rho$	see (2.3)
Q	see (2.4)
$\mathbb{E}$	the expectation, see Example 2.25
Var	the variance, see Example 2.26
Cov	the covariance, see Example 2.26
ρ^{EE}_α	the expected excess, see Example 2.27
$\rho^{EE,q}_\alpha$	the expected excess of order q, see Example 2.27
ρ^{SD}	the semideviation, see Example 2.28
ρ^P_α	the excess probability, see Example 2.29
VaR_α	the value-at-risk, see Example 2.30
CVaR_α	the conditional value-at-risk, see Example 2.31
$\lVert \cdot \rVert$	the Euclidean norm
$\lVert \cdot \rVert_\infty$	the supremum norm, see (2.8)
$\lVert \cdot \rVert^*_\infty$	see (2.9)
$\lVert \cdot \rVert_p$	the L^p-norm
$\lVert \cdot \rVert_{\mathcal{L}(\mathbb{R}^m, \mathbb{R}^k)}$	the operator norm
g_h	see (2.7)
π	the Prokhorov metric, see Definition 2.35
Λ	see 2.38
μ^{Y_l}	see (2.10)
Ψ	see Remark 2.58
θ	see (2.16)

χ_A	the indicator function of the set A
$\sup_{\leq_E}$	the supremum with respect to $\leq_E$, see Definition 2.75
$\inf_{\leq_E}$	the infimum with respect to $\leq_E$, see Definition 2.75
$\vert\cdot\vert_{\leq_E}$	the absolute value with respect to $\leq_E$, see Definition 2.76
ϱ^{**}	the biconjugate of ϱ, see section 11 A in [177]
d_ψ	see Corollary 2.62
$d_{W,s,q}$	the Wasserstein metric of order q, see Proposition 2.63
$d_{FM,q}$	the Fortet-Mourier metric of order q, see Proposition 2.63
d_∞	the Hausdorff distance
φ	see (3.7)
Φ	see (3.8)

Sets and spaces

$\bar{\mathbb{R}}$	$= [-\infty, \infty]$
$C_b^0(\mathbb{R}^s)$	the space of all bounded and continuous functions $h : \mathbb{R}^s \to \mathbb{R}$
E^*	the dual space of the normed space E
$\mathcal{B}(\mathbb{R}^s)$	the Borel $\sigma-$algebra of $\mathbb{R}^s$, see Definition 2.1
$\mathcal{P}(\mathbb{R}^s)$	the space of Borel probability measures on $\mathbb{R}^s$, see Definition 2.2
$\mathcal{M}_s^\psi$	see Definition 2.3
$\mathcal{M}_s^p$	$= \mathcal{M}_s^{\Vert\cdot\Vert^p}$, the space of Borel probability measures on $\mathbb{R}^s$ having finite moments of order p
$\mathbb{M}(\mu, \nu)$	see Proposition 2.63
$\mathbb{L}(\mu, \nu)$	see Proposition 2.63
$L^p(\Omega, \mathcal{F}, \mathbb{P})$	the standard L^p- space on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $0 < p \leq \infty$
$L^0(\Omega, \mathcal{F}, \mathbb{P})$	the space of finite-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$
$B_\epsilon(t)$	the open $\Vert\cdot\Vert-$ball of radius ϵ centered at t
$\overline{B_\epsilon(t)}$	the closed $\Vert\cdot\Vert-$ball of radius ϵ centered at t
int A	the interior of the set A
∂A	the topological boundary of the set A
2^A	the power set of the set A
$\mathcal{D}^s$	see (2.13)
C_s^ψ	see Definition 2.53
dom ϱ	the domain of the mapping ϱ, see (2.79)
D_f	the set of discontinuities of the function f
$D(A)$	see Theorem 3.9
pos(A)	see Theorem 3.9

Topologies and notions of convergence

τ_w^s	the topology of weak convergence, see Definition 2.32
$\overset{w}{\to}$	weak convergence in the sense of Definition 2.32
$\overset{w^*}{\to}$	weak* convergence, see Corollary 2.39
$\overset{d}{\to}$	convergence in distribution, see Definition 2.42
τ_ψ	see Definition 2.53
$\overset{\psi}{\to}$	convergence with respect to τ_ψ, see Definition 2.53
$\tau_{\mathbb{R}^n}$	the standard topology on $\mathbb{R}^n$
$\tau_1 \otimes \tau_2$	the product topology of τ_1 and τ_2

Bibliography

[1] C. Acerbi, *Spectral measures of risk: A coherent representation of subjective risk aversion*, Journal of Banking and Finance, 26, pp. 1505-1518 (2002).

[2] C. Acerbi, D. Tasche, *Expected Shortfall: A natural coherent alternative to Value at Risk*, Economic Notes, 31, pp. 379-388 (2002).

[3] C. Acerbi, D. Tasche, *On the coherence of Expected Shortfall*, Journal of Banking and Finance, 26, pp. 1487-1503 (2002).

[4] S. Ahmed, *Convexity and decomposition of mean-risk stochastic programs*, Mathematical Programming, 106, pp. 433-446 (2006).

[5] E. Aiyoshi, Y. Ishizuka, *Double penalty method for bilevel optimization problems*, Annals of Operations Research, 34(1), pp. 73-88 (1992).

[6] E. Aiyoshi, K. Shimizu, *A solution method for the static constrained Stackelberg problem via penalty method*, IEEE Transactions on Automatic Control, 29(12), pp. 1111-1114 (1984).

[7] C. D. Aliprantis, K. C. Border, *Infinite Dimensional Analysis - A Hitchhiker's Guide*, Springer, Berlin, Heidelberg, 3rd Edition (2006).

[8] C. D. Aliprantis, O. Burkinshaw, *Locally solid Riesz spaces with applications to economics*, Mathematical Surveys and Monographs, 105, American Mathematical Society, Providence (2003).

[9] S. M. Alizadeh, P. Marcotte, G. Savard, *Two-stage stochastic bilevel programming over a transportation network*, Transportation Research Part B: Methodological, 58, pp. 58-105 (2013).

[10] F. Andersson, H. Mausser, D. Rosen, S. Uryasev, *Credit risk optimization with Conditional Value-at-Risk criterion*, Mathematical Programming, Series B, 89, pp. 273-291 (2001).

[11] G. Angelsberg, F. Delbaen, I. Kaelin, M. Kupper, J. Näf, *On a class of law invariant convex risk measures*, Finance and Stochastics, 15(2), pp. 343-363 (2011).

[12] J. M. Arroyo, M. Carrion, A. J. Conejo, *A bilevel stochastic programming approach for retailer futures market trading*, Power Systems, IEEE Transactions on, 24(3), pp. 1446-1456 (2009).

[13] Z. Artstein, *Sensitivity with respect to the underlying information in stochastic programs*, Journal of Computational and Applied Mathematics, 56, pp. 1487-1503 (1994).

[14] Z. Artstein, R. J.-B. Wets, *Stability results for stochastic programs and sensors, allowing for discontinuous objective functions*, SIAM Journal on Optimization, 4, pp. 537-550 (1994).

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

[16] M.-B. Aryanezhad, E. Roghanian, S. J. Sadjadi, *A probabilistic bi-level linear multi-objective programming problem to supply chain planning*, Applied Mathematics and Computation, 188(1), pp. 786-800 (2007).

[17] T. Asamov, A. Ruszczyński, *Time-consistent approximations of risk-averse multistage stochastic optimization problems*, Mathematical Programming, Series A, 153, pp. 459-493 (2015).

[18] A. A. Auslender, P. Coutat, *Sensitivity Analysis for Generalized Linear-Quadratic Problems*, Journal of Optimization Theory and Applications, Vol. 88(3), pp. 541-559 (1996).

[19] A. A. Auslender, J.-P. Crouzeix, *Global Regularity Theorems*, Mathematics of Operations Research, 13(2), pp. 243-253 (1988).

[20] B. Bank, J. Guddat, D. Klatte, B. Kummer, K. Tammer, *Non-Linear Parametric Optimization*, Akademie-Verlag, Berlin (1982).

[21] B. Bank, R. Hansel, *Stability of mixed-integer quadratic programming problems*, Math Programming Study, 21, pp. 1-17 (1984).

[22] B. Bank, R. Mandel, *Parametric Integer Optimization*, Akedemie-Verlag, Berlin (1988).

[23] J. F. Bard, *Optimality Conditions for the Bilevel Programming Problem*, Naval research logistics quarterly, 31(1), pp. 13-26 (1984).

[24] J. F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publishers, Dordrecht (1998).

[25] J. F. Bard, S. Dempe, *Bundle trust-region algorithm for bilinear bilevel programming*, Journal of Optimization Theory and Applications, 110(2), pp. 265-288 (2001).

[26] J. F. Bard, J. E. Falk, *An explicit solution to the multi-level programming problem*, Computers and Operations Research, 9(1), pp. 77-100 (1982).

[27] L. A. Barroso, D. Candal, M. Fampa, L. Simonetti, *Bilevel optimization applied to strategic pricing in competitive electricity markets.* Computational Optimization and Applications, 39(2), pp. 121-142 (2008).

[28] H. Bauer, *Measure and Integration Theory*, de Gruyter, Berlin (2001).

[29] E. M. L. Beale, *On minimizing a convex function subject to linear inequalities*, Journal of the Royal Statistical Society, Series B, 17, pp. 173-184 (1955).

[30] E. M. L. Beale, J. J. H. Forrest, C. J. Taylor, *Multi-time-period stochastic programming*, in: M. A. H. Dempster (ed.), Stochastic Programming, Conference Series, Academic Press, London, pp. 387-402 (1980).

[31] D. Belomestny, V. Krätschmer, *Central limit theorems for law-invariant risk measures*, J. Appl. Prob. 49, pp. 1-21 (2012).

[32] A. Ben-Tal, O. Boni, A. Nemirovski, *Robust solutions to conic quadratic problems and their applications*, Optimization and Engineering, 9, pp. 1-18 (2008).

[33] A. Ben-Tal, A. Goryashko, E. Guslitzer, A. Nemirovski, *Adjustable robust solutions of uncertain linear programs*, Mathematical Programming, 99(2), pp. 351-376 (2004).

[34] A. Ben-Tal, A. Nemirovski, *Robust Convex Optimization*, Mathematics of Operations Research, 23(4), pp. 769-805 (1998).

[35] A. Ben-Tal, A. Nemirovski, *Robust optimization - methodology and applications*, Mathematical Programming, Series B, 92, pp. 453-480 (2002).

[36] A. Ben-Tal, A. Nemirovski, *Robust solutions of linear programming problems contaminated with uncertain data*, Mathematical Programming, Series A, 88(3), pp. 411-424 (2000).

[37] A. Ben-Tal, A. Nemirovski, *Robust solutions to uncertain linear programs*, Operations Research Letters, 25, pp. 1-13 (1999).

[38] A. Ben-Tal, A. Nemirovski, *Selected topics in robust convex optimization*, Mathematical Programming, Series B, 112, pp. 125-158 (2008).

[39] A. Ben-Tal, A. Nemirovski, C. Roos, *Robust solutions of uncertain quadratic and conic-quadratic problems*, SIAM Journal on Optimization, Vol. 13(2), pp. 535-560 (2002).

[40] B. Bereanu, *Minimum risk criterion in stochastic optimization*, Economic Computation and Economic Cybernetics Studies and Research, 2, pp. 31-39 (1981).

[41] C. Berge, *Topological Spaces*, Macmillian, New York (1963).

[42] P. Berti, L. Pratelli, P. Rigo, *Skorohod representation on a given probability space*, Probability Theory and Related Fields, 137, pp. 277-288 (2007).

[43] D. J. Bertsimas, *A Vehicle Routing Problem with Stochastic Demands*, Operations Research, Vol. 40, pp. 574-585 (1992).

[44] D. Bertsimas, M. Sim, *The Price of Robustness*, Operations Research, 52, pp. 35-53 (2004).

[45] M. J. Best, N. Chakravarti, *Stability of Linearly Constrained Convex Quadratic Programs*, Journal of Optimization Theory and Applications, Vol. 64(1), pp. 43-53 (1990).

[46] M. J. Best, B. Ding, *On the Continuity of the Minimum in Parametric Quadratic Programs*, Journal of Optimization Theory and Applications, Vol. 86(1), pp. 245-250 (1995).

[47] W. Bialas, M. Karwan, *Two-level linear programming*, Management Science, 30, pp. 1004-1020 (1984).

[48] P. Billingsley, *Probability and Measure*, 3. ed., Wiley, New York (1995).

[49] P. Billingsley, *Convergence of Probability Measures*, 2. ed., Wiley, New York (1999).

[50] S. I. Birbil, G. Gürkan, O. Listes, *Simulation-based solution of stochastic mathematical problems with complementarity constraints: Sample-path analysis*, Technical report, ERIM Report Series Research in Management (2014).

[51] J. R. Birge, *Decomposition and partitioning methods for multistage stochastic linear programs*, Operations Research, 33, pp. 989-1007 (1985).

[52] J.R. Birge, F. Louveaux, *Introduction to Stochastic Programming*, 2nd Edition, Springer, New York (2011).

[53] C. E. Blair, R. G. Jeroslow, *The calue function of a mixed integer program: I*, Discrete Mathematics, 19, pp. 121-138 (1977).

[54] E. Blum, W. Oettli, *Direct Proof for the Existence Theorem for Quadratic Programming*, Operations Research, Vol. 20(1), pp. 165-167 (1972).

[55] J. F. Bonnans, A. Shapiro, *Perturbation analysis of optimization problems*, Springer Series in Operations Research, Springer, New York (2000).

[56] M. Brandtner, *Risikomessung mit kohärenten, spektralen und konvexen Risikomaßen - Konzeption, entscheidungstheoretische Implikationen und finanzwirtschaftliche Anwendungen*, Schriften zur Quantitativen Betriebswirtschaftslehre, 27, Gabler Verlag, Wiesbaden (2012).

[57] J, Burtscheidt, M. Claus, *A Note on Stability for Risk Averse Stochastic Complementarity Problems*, `https://www.uni-due.de/imperia/md/content/mathematik/2016_preprint_799.pdf` (2016).

[58] W. Candler, R. Norton, *Multilevel programming*, Technical report, World Bank Developement Research Center, Washington D.C., (1977).

[59] W. Candler, R. Townsley, *A linear two-level programming problem*, Computers and Operations Research, 9(1), pp. 59-76 (1982).

[60] M. Carrión, U. Gotzes, R. Schultz, *Risk aversion for an electricity retailer with second-order stochastic dominance constraints*, Computational Management Science 6, pp. 233-250 (2009).

[61] C. Carøe, R. Schultz, *Dual decomposition in stochastic integer programming*, Operations Research Letters, Vol. 24, pp. 37-45 (1999).

[62] Y. Chen, M. Florian, *The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions*, Optimization, 32(3), pp. 193-209 (1995).

[63] Z. Chen, Y. Han, *Continuity of the optimal value function and optimal solutions of parametric mixed-integer quadratic programs*, Appl. Math. J. Chinese Univ., 25(4), pp. 391-399 (2010).

[64] Z. Chen, Z. Xu, *Continuity and Stability of a Quadratic Mixed-Integer Stochastic Program*, Numerical Functional Analysis and Optimization, 30(5-6), pp. 462-477 (2009).

[65] P. Cheridito, T. Li, *Risk measures on orlicz hearts*, Mathematical Finance, Vol. 18(2), pp. 189-214 (2009).

[66] S. Christiansen, M. Patriksson, L. Wynter, *Stochastic Bilevel Programming in Structural Optimization*, Struct. Multidisciplinary Optimization, 21(5), pp. 361-371 (2001).

[67] C. Cicerone, G. Di Stefano, M. Schachtebeck, A. Schöbel, *Multi-Stage Recovery Robustness for Optimization Problems: A new Concept for Planning under Disturbances*, Information Sciences, 190, pp. 107-126 (2012).

[68] M. Claus, *On stability in risk averse stochastic bilevel programming*, `https://www.uni-due.de/imperia/md/content/mathematik/2016_preprint_798.pdf` (2016).

[69] M. Claus, V. Krätschmer, R. Schultz, *Weak continuity of risk functionals with applications to stochastic programming*, `https://www.uni-due.de/imperia/md/content/mathematik/2015_preprint_790.pdf` (2015).

[70] M. Claus, R. Schultz, *Lipschitzian properties and stability of a class of first-order stochastic dominance constraints*, SIAM Journal on Optimization, 25(1), pp. 396-415 (2015).

[71] D. L. Cohn, *Measure Theory: Second Edition*, Springer, New York (2013).

[72] R. A. Collado, D. Papp, A. Ruszczyński, *Scenario decomposition of risk-averse multistage stochastic programming problems*, Annals of Operations Research, 200, pp. 147-170 (2012).

[73] S. Conti, H. Held, M. Pach, M. Rumpf, R. Schultz, *Risk averse shape optimization*, SIAM Journal on Control and Optimization, 49, pp. 927-947 (2011).

[74] B. Colson, P. Marcotte, G. Savard, *A trust-region method for nonlinear bilevel programming: Algorithm and computational experience*, Computational Optimization and Applications, 30(3), pp. 211-227 (2005).

[75] B. Colson, P. Marcotte, G. Savard, *An overview of bilevel optimization*, Annals of Operations Research, 153, pp. 235-256 (2007).

[76] G. B. Dantzig, *Linear Programming Under Uncertainty*, Management Science, Vol. 1, pp. 197-206 (1955).

[77] F. Delbaen, *Coherent risk measures on general probability spaces*, in K. Sandmann, P. J. Schönbucher (editors), Advances in Finance and Stochastics - Essays in Honour of Dieter Sondermann, Springer, Berlin (2002).

[78] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht (2002).

[79] S. Dempe, J. Dutta, B. S. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization, 56(5-6), pp. 577-604 (2007).

[80] S. Dempe, V. V. Kalashnikov, N. Kalashnykova, *Optimality conditions for bilevel programming problems*, in: S. Dempe, P. Raskolnikova (eds.), Optimization with Multivalued Mappings, Springer, pp. 3-28 (2006).

[81] S. Dempe, V. V. Kalashnikov, N. Kalashnykova, G. A. Pérez-Valdés, *Natural gas bilevel cash-out problem: Convergence of a penalty function method*, European Journal of Operational Research, 215(3), pp. 532-538 (2011).

[82] S. Dempe, A. B. Zemkoho, *The Generalized Mangasarian-Fromowitz Constraint Qualification and Optimality Conditions for Bilevel Programs*, Journal of Optimization Theory and Applications, 148(1), pp. 46-68 (2011).

[83] S. Dempe, A. B. Zemkoho, *The bilevel programming problem: reformulations, constraint qualifications and optimality conditions*, Mathematical Programming, 138(1), pp. 447-473 (2013).

[84] D. Dentcheva, R. Henrion, A. Ruszczyński, *Stability and sensitivity of optimization problems with first-order stochastic dominance constraints*, SIAM Journal on Optimization, 18, pp. 322-337 (2007).

[85] D. Dentcheva, G. Martinez, *Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse*, European Journal of Operational Research, 219, pp. 1-8 (2012).

[86] D. Dentcheva, W. Römisch, *Stability and sensitivity of stochastic dominance constrained optimization problems*, SIAM Journal on Optimization, 23, pp. 1672-1688 (2013).

[87] D. Dentcheva, A. Ruszczyński, *Inverse cutting plane methods for optimization problems with second-order stochastic dominance constraints*, Optimization, 59, pp. 323-338 (2010).

[88] D. Dentcheva, A. Ruszczyński, *Optimality and duality theory for stochastic optimization with nonlinear dominance constraints*, Mathematical Programming, 99, pp. 329-350 (2004).

[89] D. Dentcheva, A. Ruszczyński, *Optimization with stochastic dominance constraints*, SIAM Journal on Optimization, 14, pp. 548-566 (2003).

[90] D. Dentcheva, A. Ruszczyński, *Portfolio optimization with stochastic dominance constraints*, Journal on Banking and Finance, 30, pp. 433-451 (2006).

[91] D. Dentcheva, A. Ruszczyński, *Robust stochastic dominance and its application to risk-averse optimization*, Mathematical Programming, 123, pp. 85-100 (2010).

[92] D. Dentcheva, A. P. Ruszczyński, A. Shapiro, *Lectures on Stochastic Programming: Modeling and Theory*, MPS SIAM Series on Optimization, 9, SIAM, Philadelphia, 2nd Edition (2014).

[93] D. Drapkin, R. Schultz, *An algorithm for stochastic programs with first-order dominance constraints induced by linear recourse*, Discrete Applied Mathematics, 158, pp. 291-297 (2010).

[94] J. Dupacŏvá, Multistage stochastic programs: The state-of-the-art and selected bibliography, Kybernetika, 31, pp. 151-174 (1995).

[95] B. C. Eaves, *On Quadratic Programming*, Management Science, Vol. 17(11), pp. 698-711 (1971).

[96] A. Eichhorn, W. Römisch, *Polyhedral risk measures in stochastic programming*, SIAM Journal on Optimization, Vol. 16(1), pp. 69-95 (2005).

[97] L. El Ghaoui, H. Lebret, *Robust Solutions to Least-Squares Problems with Uncertain Data*, SIAM Journal on Matrix Analysis and Applications, Vol. 18(4), pp. 1035-1064 (1997).

[98] L. El Ghaoui, H. Lebret, F. Oustry, *Robust Solutions to Uncertain Semidefinite Programs*, SIAM Journal on Optimization, Vol. 9(1), pp. 33-52 (1998).

[99] S. Engell, A. Märkert, G. Sand, R. Schultz, *Aggregated scheduling of a multiproduct batch plant by two-stage stochastic integer programming*, Optimization and Engineering, 5, pp. 335-359 (2004).

[100] T. Ermolieva, Y. Ermoliev, *Catastrophic risk management: Flood and seismic risks case studies*, in: S. W. Wallace and W. T. Ziemba (eds.), Applications of Stochastic Programming, MPS-SIAM Series on Optimization, SIAM (2005).

[101] C.S. Fabian, G. Mitra, D. Roman, *Processing second-order stochastic dominance models using cutting-plane representations*, Mathematical Programming, 130, pp. 33-57 (2011).

[102] E. A. Feinberg, P. O. Kasyanov, N. V. Zadoianchuk, *Berge's theorem for noncompact image sets*, Journal of Mathematical Analysis and Applications, 379, pp. 255-259 (2013).

[103] O. Fiedler, W. Römisch, *Stability in multistage stochastic programming*, Annals of Operations Research, 56, pp. 79-93 (1995).

[104] D. Filipović, G. Svindland, *The canonical model space for law-invariant convex risk measures is L^1*, Mathematical Finance, Vol. 22(3), pp. 585-589 (2012).

[105] H. Föllmer, T. Knispel, *Convex Risk Measures: Basic Facts, Law-invariance and beyond, Asymptotics for Large Portfolios*, in L. C. MacLean and W. T. Ziemba (editors), Handbook of the Fundamentals of Financial Decision Making, Part II, pp. 507-554, World Scientific (2013).

[106] H. Föllmer, A. Schied, *Convex measures of risk and trading constraints*, Finance and Stochastics, 6(4), pp. 429-447 (2002).

[107] H. Föllmer, A. Schied, *Stochastic Finance*, de Gruyter, Berlin, 3rd Edition (2011).

[108] M. Frittelli, E. R. Gianin, *Law Invariant Convex Risk Measures*, Advances in Mathematical Economics, 7, pp. 42-53 (2005).

[109] M. Frittelli, E. R. Gianin, *Putting order in risk measures*, Journal of Banking and Finance, 26(7), pp. 1473-1486 (2002).

[110] M. Fukushima, G.-H. Lin, *Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: a survey.* Pacific Journal of Optimization, 6(3), pp. 455-482 (2010).

[111] A. A. Gaivoronski, A. Werner, *Stochastic Programming Perspective on the Agency Problems under Uncertainty*, In Managing Safety of Heterogeneous Systems, Springer, pp. 137-167 (2012).

[112] M. Georigk, A. Schöbel, *Recovery-to-optimality: A new two-stage approach to robustness with an application to aperiodic timetabling*, Computers and Operations Research, 52, pp. 1-15 (2014).

[113] V. Goel, I. E. Grossmann, *A class of stochastic programs with decision dependent uncertainty*, Mathematical Programming, 108(2-3), pp. 355-394 (2006).

[114] D. Goldfarb, G. Iyengar, *Robust Portfolio Selection Problems*, Mathematics of Operations Research, 28(1), pp. 1-38 (2003).

[115] R. Gollmer, U. Gotzes, F. Neise, R. Schultz, *Risk modeling via stochastic dominance in power systems with dispersed generation*, in "Stochastic Optimization Methods in Finance and Energy", M. Bertocchi, G. Consigli, and M.A.H. Dempster, eds., Springer, New York, pp. 253-271 (2011).

[116] R. Gollmer, U. Gotzes, R. Schultz, *A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse*, Mathematical Programming, 127, pp. 179-190 (2011).

[117] R. Gollmer, F. Neise, R. Schultz, *Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse*, SIAM Journal on Optimization, 19, pp. 552-571 (2008).

[118] F. Granot, J. Skorin-Kapov, *Some proximity and sensitivity results in quadratic integer programming*, Math. Programming, 47, pp. 259-268 (1990).

[119] P. Gross, G. C. Pflug, *Behavioral pricing of energy swing options by stochastic bilevel optimization*, Energy Systems, pp. 1-26 (2016).

[120] A. G. HADIGHEH, O. ROMANKO, T. TERLAKY, *Sensitivity analysis in convex quadratic optimization: Simultaneous perturbation of the objective and right-hand-side vectors*, Algorithmic Operations Research, 2, pp. 95-111 (2007).

[121] C. HEIL, *A Basis Theory Primer: Expanded Edition*, Applied and Numerical Harmonic Analysis, Springer, Boston (2011).

[122] T. HEINZE, R. SCHULTZ, *A branch-and-bound method for multistage stochastic integer programs with risk objectives*, Optimization, 57, pp. 277-293 (2008).

[123] H. HEITSCH, W. RÖMISCH, *Scenario tree modeling for multistage stochastic programs*, Mathematical Programming, Series A, 118, pp. 371-406 (2009).

[124] H. HEITSCH, W. RÖMISCH, C. STRUGAREK, *Stability of multistage stochastic programs*, SIAM Journal on Optimization, 17(2), pp. 511-525 (2006).

[125] R. HENRION, *Qualitative stability of convex programs with probabilistic constraints*, V.H. Nguyen, J.-J. Strodiot and P. Tossings (eds.): Optimization (Lect. Notes in Economics and Mathematical Systems, Vol. 481, Springer, Berlin), pp. 164-180 (2000).

[126] R. HENRION, W. RÖMISCH, *Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints*, Mathematical Programming, 100, pp. 589-611 (2004).

[127] R. HENRION, W. RÖMISCH, *Metric regularity and quantitative stability in stochastic programs with probabilistic constraints*, Mathematical Programming, 84, pp. 55-88 (1999).

[128] R. HENRION, W. RÖMISCH, *Stability of solutions to chance constrained stochastic programs*, in: J. Guddat, R. Hirabayashi, H. T. Jongen, F. Twilt (eds.), Parametric Optimization and Related Topics V, Peter Lang Verlag, Frankfurt a. M., pp. 95-114 (2000).

[129] R. HENRION, T. SUROWIEC, *On calmness conditions in convex bilevel programming*, Applicable Analysis, 90(6), pp. 951-970 (2011).

[130] T. HOMEM-DE MELLO, S. MEHROTRA, *A cutting-surface method for uncertain linear programs with polyhedral stochastic dominance constraints*. SIAM Journal on Optimization, 20(3), pp. 1250-1273 (2010).

[131] S. V. IVANOV, *Bilevel stochastic linear programming problems with quantile criterion*, Automation and Remote Control, 75(1) pp. 107-118 (2014).

[132] J. Júdice, G. Savard, L. Vincente, *Descent approaches for quadratic bilevel programming*, Journal of Optimization Theory and Applications, 81(2), pp. 379-399 (1994).

[133] P. Kall, *On approximations and stability in stochastic programming*, in: J. Guddat, H. Th. Jongen, B. Kummer, F. Nožička (eds.), Parametric Optimization and Related Topics, Akademie Verlag, Berlin, pp. 387-407 (1987).

[134] O. Kallenberg, *Foundations of Modern Probability*, Springer, New York (1997).

[135] V. Kankova, *Stability in the stochastic programming*, Kybernetika 14, pp. 339-349 (1978).

[136] D. Klatte, *Lipschitz continuity of infima and optimal solutions in parametric optimization: The polyhedral case*, in J. Guddat, H. Th. Jongen, B. Kummer, and F. Nozicka, editors, Parametric Optimization and Related Topics, Akademie-Verlag, Berlin, pp. 229-248 (1987).

[137] D. Klatte, *On the Lipschitz Behavior of Optimal Solutions in Parametric Problems of Quadratic Optimization and Linear Complementarity*, Optimization, 16, pp. 819-831 (1985).

[138] D. Klatte, G. Thiere, *Error bounds for solutions of linear equations and inequalities*, ZOR - Mathematical Methods of Operations Research 41, pp. 191-214 (1995).

[139] S. Kosuch, P. Le Bodic, J. Leung, A. Lisser, *On a stochastic bilevel programming problem*, Networks, 59(1), pp. 107-116 (2012).

[140] R. M. Kovacevic, G. C. Pflug, *Electricity swing option pricing by stochastic bilevel optimization: a survey and new approaches*, European Journal of Operational Research, Vol. 237(2), pp. 389-403 (2014).

[141] V. Krätschmer, A. Schied, H. Zähle, *Comparative and qualitative robustness for law-invariant risk measures*, Finance and Stochastics, 18, pp. 271-295 (2014).

[142] V. Krätschmer, A. Schied, H. Zähle, *Domains of weak continuity of statistical functionals with a view on robust statistics*, submitted, see `http://arxiv.org/pdf/1511.08677v1.pdf` for a preprint (2015).

[143] V. Krätschmer, A. Schied, H. Zähle, *Qualitative and infinitesimal robustness of tail-dependent statistical functionals*, Journal of Multivariate Analysis, 103, pp. 35-47 (2012).

[144] V. Krätschmer, A. Schied, H. Zähle, *Quasi-Hadamard differentiability of general risk functionals and its application*, Statistics and Risk Modeling, 32, pp. 25-47 (2015).

[145] G. Laporte, F. V. Louveaux, The integer L-shaped method for stochastic integer programs with complete recourse, Operations Research Letters, Vol. 13, pp. 133-142 (1993).

[146] G. M. Lee, N. N. Tam, N. D. Yen, *On the Optimal Value Function of a Linearly Perturbed Quadratic Program*, Journal of Global Optimization, 32, pp. 119-134 (2005).

[147] G. M. Lee, N. N. Tam, N. D. Yen, *Continuity of the Solution Map in Quadratic Programs under Linear Perturbations*, Journal of Optimization Theory and Applications, Vol. 129(3), pp. 415-423 (2006).

[148] G.-H. Lin, Y. Liu, H. Xu, *Stability Analysis of Two-Stage Stochastic Mathematical Programs with Complementarity Constraints via NLP Regularization*, SIAM Journal on Optimization, 21(3), pp. 669-705 (2011).

[149] Y. Liu, H. Xu, *Stability analysis of stochastic programs with second order dominance constraints*, Mathematical Programming, 142, pp. 435-460 (2013).

[150] J. Luedtke, *New formulations for optimization under stochastic dominance constraints*, SIAM Journal on Optimization, 19, pp. 1433-1450 (2008).

[151] F. Luque-Vásquez, O. Hernández-Lerma, *A counterexample on the semicontinuity of minima*, Proceedings of the American Mathematical Society, Vol. 123(10), pp. 3175-3176 (1995).

[152] W. A. J. Luxemburg, A. C. Zaanen, *Riesz Spaces I*, North Holland, Amsterdam (1971).

[153] A. Märkert, R. Schultz, *On deviation measures in stochastic integer programming*, Operations Research Letters, 33, pp. 441-449 (2005).

[154] A. Matani, Y. Terazono, *Continuity of optimal solution functions and their conditions on objective functions*, SIAM Journal on Optimization, 25(4), pp. 2050-2060 (2015).

[155] P. Meyer-Nieberg, *Banach lattices*, Springer, Berlin (1991).

[156] N. Miller, A. Ruszczyński, *Risk-Averse Two-Stage Stochastic Linear Programming - Modeling and Decomposition*, Operations Research, Vol. 59(1), pp. 125-132 (2011).

[157] N. Noyan, G. Rudolf, A. Ruszczyński, *Relaxations of linear programming problems with first order stochastic dominance constraints*, Operations Research Letters 34, pp. 653-659 (2006).

[158] W. Ogryczak, A. Ruszczyński, *Dual stochastic dominance and related mean-risk models*, SIAM Journal on Optimization, 13, pp. 60-78 (2002).

[159] W. Ogryczak, A. Ruszczyński, *From stochastic dominance to mean-risk models: Semideviations as risk measures*, European Journal of Operations Research, 116, pp. 33-50 (1999).

[160] W. Ogryczak, A. Ruszczyński, *On consistency of stochastic dominance and mean-semideviation models*, Mathematical Programming, 89, pp. 217-232 (2001).

[161] M. Patriksson, *On the applicability and solution of bilevel optimization models in transportation science: A study on the existence, stability and computation of optimal solutions to stochastic mathematical programs with equilibrium constraints*, Transportation Research Part B: Methodological 42(10), pp. 843-860 (2008).

[162] M. Patriksson, L. Wynter, *Stochastic Nonlinear Bilevel Programming*, Technical report, PRISM, Universitée de Versailles-Saint Quentin en Yvelines, Versailles, France (1997).

[163] M. Patriksson, L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Operations Research Letters 25, pp. 159-167 (1999).

[164] G. C. Pflug, *Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk*, in S. P. Uryasev (editor), Probabilistic Constrained Opimization - Methodology and Applications, pp. 272-281, Kluwer Academic Publishers, Dordrecht (2000).

[165] G. C. Pflug, A. Pichler, *Multistage Stochastic Optimization*, Springer Series in Operations Research and Financial Engineering, Springer (2014).

[166] G. C. Pflug, W. Römisch, *Modeling, Measuring and Managing Risk*, World Scientific, Singapore (2007).

[167] A. B. Philpott, *Stochastic optimization and yacht racing*, in S. W. Wallace, W. T. Ziemba (eds.), Applications of Stochastic Programming, MPS-SIAM Series on Optimization, SIAM (2005).

[168] A. Prékopa, *Stochastic Programming Problem with Random RHS Obeying a Discrete Probability Distribution*, ZOR - Methods and Models of Operations Research, 34, pp. 441-461 (1990).

[169] S. T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley, Chichester et al., (1991).

[170] S.T. Rachev, W. Römisch, *Quantitative stability in stochastic programming: The method of probability metrics*, Mathematics of Operations Research, 27, pp. 792-818 (2002).

[171] F. Riesz, *Sur la décomposition des opérations fonctionnelles linéaires*, Atti congress. internaz. mathematici (Bologna, 1928), 3, Zanichelli, pp. 143-148 (1930).

[172] M. Riis, R. Schultz, *Applying the Minimum Risk Criterion in Stochastic Recourse Programs*, Computational Optimization and Applications, 24, pp. 267-287 (2003).

[173] A. W. Roberts, D. E. Varberg, *Another Proof that Convex Functions are Locally Lipschitz*, Am. Math. Mon., 81, pp. 1014-1016 (1974).

[174] S. M. Robinson, R. J.-B. Wets, *Stability in two-stage stochastic programming*, SIAM Journal on Control and Optimization, 25, pp. 1409-1416 (1987).

[175] R. T. Rockafellar, S. Uryasev, *Conditional Value-at-Risk for General Loss Distributions*, Journal of Banking and Finance, 26, pp. 1443-1471 (2002).

[176] R. T. Rockafellar, S. Uryasev, *Optimization of Conditional Value-at-Risk*, Journal of Risk, 2, pp. 21-41 (2000).

[177] R. T. Rockafellar, R. J.-B. Wets, *Variational Analysis*, Springer, Berlin (2009).

[178] W. Römisch, *Stability of Stochastic Programming Problems*, in [192], pp. 483-554 (2003).

[179] W. Römisch, R. Schultz, *Distribution sensitivity for certain classes of chance-constrained models with application to power dispatch*, Journal of Optimization Theory and Applications, 71, pp. 569-588 (1991).

[180] W. Römisch, R. Schultz, *Distribution sensitivity in stochastic programming*, Mathematical Programming, 50, pp. 197-226 (1991).

[181] W. Römisch, R. Schultz, *Lipschitz stability for stochastic programs with complete recourse*, SIAM Journal on Optimization, 6, pp. 531-547 (2006).

[182] W. Römisch, R. Schultz, *Multistage stochastic integer programs: an introduction*, in: M. Grötschel, S. O. Krumke, J. Rambau (eds.), Online optimization of large scale systems, pp. 581-622 (2001).

[183] W. Römisch, R. Schultz, *Stability analysis for stochastic programs*, Annals of Operations Research, 30, pp. 241-266 (1991).

[184] W. Römisch, R. Schultz, *Stability of solutions for stochastic programs with complete recourse*, Mathematics of Operations Research, 18, pp. 590-609 (1993).

[185] W. Römisch, S. Vigerske, *Quantitative stability of fully random mixed-integer two-stage stochastic programs*, Optimization Letters, 2, pp. 377-388 (2008).

[186] W. Römisch, A. Wakolbinger, *Obtaining convergence rates for approximations in stochastic programming*, in: J. Guddat, H. Th. Jongen, B. Kummer, F. Nožička (eds.), Parametric Optimization and Related Topics, Akademie Verlag, Berlin, pp. 327-343 (1987).

[187] G. Rudolf, A. Ruszczyński, *Optimization problems with second order stochastic dominance constraints: duality, compact formulations and cut generation methods*, SIAM Journal on Optimization, 19, pp. 1326-1343 (2008).

[188] W. Rudin, *Banach lattices and positive operators*, Springer, Berlin (1974).

[189] W. Rudin, *Real and Complex Analysis*, 3. ed., McGraw-Hill, New York (1987).

[190] L. Rüschendorf, *Mathematical risk analysis*, Springer, Berlin, Heidelberg (2013).

[191] A. Ruszczyński, *Decomposition methods in stochastic programming*, Mathematical Programming, 79, pp. 333-353 (1997).

[192] A. Ruszczyński, A. Shapiro (eds.), *Handbooks in Operations Research and Management Science, 10: Stochastic Programming*, Elsevier, Amsterdam (2003).

[193] A. Ruszczyński, A. Shapiro, *Optimization of Convex Risk Functions*, Mathematics of Operations Research, 31(3), pp. 433-452 (2006).

[194] A. Schöbel, *Generalized light robustness and the trade-off between robustness and nominal quality*, Mathematical Methods of Operations Research, 80, pp. 161-191 (2014).

[195] R. Schultz, *Mixed-integer value functions in stochastic programming*, in: M. Jünger, G. Reinelt, G. Rinaldi (eds.), "Combinatorial Optimization - Eureka, You Shrink! Papers Dedicated to Jack Edmonds", Springer, Lecture Notes in Computer Science, 2570, pp. 171-184 (2003).

[196] R. Schultz, *On structure and stability in stochastic programs with random technology matrix and complete integer recourse*, Mathematical Programming, 70, pp. 73-89 (1995).

[197] R. Schultz, *Rates of convergence in stochastic programs with complete integer recourse*, SIAM Journal on Optimization, 6, pp. 1138-1152 (1996).

[198] R. Schultz, *Some Aspects of Stability in Stochastic Programming*, Annals of Operations Research, Vol. 100, Issue 1, pp. 55-84 (2000).

[199] R. Schultz, *Stochastic programming with integer variables*, Mathematical Programming, 97, pp. 285-309 (2003).

[200] R. Schultz, S. Tiedemann, *Conditional value-at-risk in stochastic programs with mixed-integer recourse*, Mathematical Programming 105, pp. 365-386 (2006).

[201] R. Schultz, S. Tiedemann, *Risk Aversion via Excess Probabilities in Stochastic Programs with Mixed-Integer Recourse*, SIAM Journal on Optimization, 14, pp. 115-138 (2003).

[202] A. Shapiro, *Asymptotic analysis of stochastic programs*, Annals of Operations Research, 30, pp. 169-186 (1991).

[203] A. Shapiro, *Asymptotic behavior of optimal solutions in stochastic programming*, Mathematics of Operations Research, 18, pp. 829-845 (1993).

[204] A. Shapiro, *Quantitative stability in stochastic programming*, Mathematical Programming, 67, pp. 99-108 (1994).

[205] N. N. Tam, *Continuity of the optimal value function in indefinite quadratic programming*, Journal of Global Optimization, 23, pp. 43-61 (2002).

[206] A. W. van der Vaart, *Asymptotic statistics*, Cambridge University Press, Cambridge (1998).

[207] R. M. Van Slyke, R. Wets, *L-shaped linear programs with applications to optimal control and stochastic programming*, SIAM Journal on Applied Mathematics, 17, pp. 638-663 (1969).

[208] H. von Stackelberg, *Marktform und Gleichgewicht*, Springer (1934).

[209] D. W. Walkup, R. J.-B. Wets, *Lifting projections of convex polyhedra*, Pacific Journal of Mathematics, 28, pp. 465-475 (1969).

[210] J. Wang, *Continuity of feasible solution sets of probabilistic constrained programs*, Journal of Optimization Theory and Applications 63, pp. 79-89 (1989).

[211] J. Wang, *Stability of multistage stochastic programming*, Annals of Operations Research, 56, pp. 313-322 (1995).

[212] A. Werner, *Bilevel stochastic programming problems: Analysis and application to telecommunications*, PhD thesis, Norwegian University of Science and Technology (2005).

[213] J. J. Ye, D. L. Zhu *A Note on Optimality Conditions for Bilevel Programming Problems*, Optimization, 39(4), pp. 361-366 (1997).

[214] J. J. Ye, D. L. Zhu, *New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches*, SIAM Journal on Optimization, 20(4), pp. 1885-1905 (2010).

[215] J. J. Ye, D. L. Zhu, *Optimality Conditions for Bilevel Programming Problems*, Optimization, 33(1), pp. 9-27 (1995).

[216] J. J. Ye, D. L. Zhu, Q. J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM Journal on Optimization, 7(2), pp. 481-507 (1997).

[217] A. C. Zaanen, *Riesz Spaces II*, North Holland, Amsterdam (1983).

[218] H. Zähle, *A definition of qualitative robustness for general point estimators, and examples*, Journal of Multivariate Analysis, 143, pp. 12-31 (2016).