SIAM J. OPTIM.
Vol. 30, No. 1, pp. 377–406

RISK-AVERSE MODELS IN BILEVEL STOCHASTIC LINEAR PROGRAMMING*

JOHANNA BURTSCHEIDT†, MATTHIAS CLAUS†, AND STEPHAN DEMPE‡

Abstract. We consider a two-stage stochastic bilevel linear program where the leader contemplates the follower's reaction at the second stage optimistically. In this setting, the leader's objective function value can be modeled by a random variable, which we evaluate based on some law-invariant (quasi-)convex risk measure. After establishing Lipschitzian properties and existence results, we derive sufficient conditions for differentiability when the choice function is a Lipschitzian transformation of the expectation. This allows us to formulate first-order necessary optimality conditions for models involving certainty equivalents or expected disutilities. Moreover, a qualitative stability result under perturbation of the underlying probability distribution is presented. Finally, for finite discrete distributions, we reformulate the bilevel stochastic problems as standard bilevel problems and propose a regularization scheme for solving a deterministic bilevel programming problem.

Key words. bilevel stochastic programming, risk measures, differentiability, stability, finite discrete models

AMS subject classifications. 90C15, 90C26, 90C31, 90C34, 91A65

DOI. 10.1137/19M1242240

1. Introduction.

Bilevel problems arise from the interplay of two decision makers at 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* reflecting his or her own goals and returns an optimal solution back to the leader. The leader's outcome depends on both his or her decision and the solution that is returned 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, models typically consider the case where the follower returns either the best (*optimistic model*) or the worst (*pessimistic model*) solution with respect to the leader's objective. The bilevel optimization problem is to find an optimal upper level decision which, even in a linear setting, results in a nonconvex, nondifferentiable, and NP-hard problem (cf. [15, Chapter 3]).

The present work is on bilevel stochastic linear problems, where the realization of some random vector whose distribution does not depend on the upper level decision enters the lower level problem as an additional parameter. It is assumed that the leader has to make his or her decision without knowing the realization of the randomness, while the follower decides under full information. This setting encapsulates two-stage stochastic programming with linear recourse as a special case, where the upper and lower level objective functions coincide.

In classical two-stage stochastic programming, the upper level objective function

*Received by the editors January 31, 2019; accepted for publication (in revised form) November 1, 2019; published electronically February 6, 2020.
https://doi.org/10.1137/19M1242240
Funding: The second author's work was supported by the Deutsche Forschungsgemeinschaft via the Collaborative Research Center TRR 154.
†Faculty of Mathematics, University of Duisburg-Essen, Campus Essen, D-45127 Essen, Germany (johanna.burtscheidt@uni-due.de, matthias.claus@uni-due.de).
‡Faculty of Mathematics and Computer Science, TU Bergakademie Freiberg, D-09599 Freiberg, Germany (dempe@math.tu-freiberg.de).

gives rise to a family of random variables defined by the optimal value function of the recourse problem. In contrast, the arising random variables in optimistic bilevel stochastic programming models depend on the optimal value of a problem where only optimal solutions of the lower level problem are feasible and the decision is made by a different actor. This is a crucial difference that entails a loss of convexity and poses additional challenges.

Nevertheless, bilevel stochastic problems are of great relevance for practical applications and have been discussed in the context of the pricing of electricity swing options [40], economics [10], supply chain planning [57], telecommunications [64], and general agency problems [26]. Other works focus on solution methods [8], bilevel stochastic problems with Knapsack constraints [39], and stochastic mathematical problems with equilibrium constraints (SMPECs) [44].

In [33], Ivanov examines bilevel stochastic linear problems with uncertainty in the right-hand side of the lower level problem and utilizes the value-at-risk to rank the arising random variables. The results include continuity of the objective function, the existence of a solution, and equivalence to a mixed-integer linear program if the underlying distribution is finite discrete. The latter result has been extended to the fully random case in [18].

In the present work, we rank the random variables arising from right-hand side uncertainty in the lower level by law-invariant convex risk measures and establish (local) Lipschitz continuity of the resulting objective function. This result allows us to formulate sufficient conditions for the existence of global minimizers.

It is well known that stochastic programming models may be smoother than their underlying deterministic counterparts. For instance, for a class of stochastic Stackelberg games employing the expectation, differentiability has been derived in [14]. Overcoming additional challenges arising from nondifferentiable integrands, we establish (continuous) differentiability for bilevel stochastic linear problems based on (Lipschitzian transformations of) the expectation, including certainty equivalents as well as expected disutility models. This leads to first-order necessary optimality conditions.

Incomplete information or the need for computational efficiency may lead to optimization models where an approximation of the true underlying distribution is employed. This motivates the analysis of the behavior of optimal values and (local) optimal solution sets under perturbations of the underlying distribution (see, e.g., [45], [49], and [50] for stability analysis of related models). For bilevel stochastic linear problems, we establish a qualitative stability result that holds for all law-invariant convex risk measures.

All of our results regarding finiteness, (Lipschitz) continuity, differentiability, and stability cover both the optimistic and the pessimistic approach of bilevel stochastic linear programming.

For finite discrete distributions and optimistic models, we show that the risk-averse bilevel stochastic linear problems using the expectation, the expected excess, or the mean upper semideviation are equivalent to standard bilevel linear problems. The resulting problems for the expectation and expected excess have at most one coupling constraint involving variables from different scenarios, which paves the way for decomposition approaches.

Finally, we show that a simplified version of the regularization scheme in [60] can be used to solve bilevel linear problems.

2. Model. We shall consider the optimistic formulation of a parametric bilevel linear problem

$$\min_x \left\{ c^\top x + \min_y \{ q^\top y \mid y \in \Psi(x,z) \} \mid x \in X \right\}, \tag{2.1}$$

where $z \in \mathbb{R}^s$ is a parameter and the data comprises a nonempty set $X \subseteq \mathbb{R}^n$, vectors $c \in \mathbb{R}^n$, $q \in \mathbb{R}^m$, and the lower level optimal solution set mapping $\Psi : \mathbb{R}^n \times \mathbb{R}^s \rightrightarrows \mathbb{R}^m$ defined by

$$\Psi(x,z) := \operatorname*{Argmin}_y \, \{ d^\top y \mid Ay \leq Tx + z \}$$

with $A \in \mathbb{R}^{s \times m}$, $T \in \mathbb{R}^{s \times n}$, and $d \in \mathbb{R}^m$. Let $f : \mathbb{R}^n \times \mathbb{R}^s \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ denote the extended real-valued mapping

$$f(x,z) := c^\top x + \min_y \{ q^\top y \mid y \in \Psi(x,z) \}.$$

LEMMA 2.1. *Assume that* dom $f \neq \emptyset$*, then f is real-valued and Lipschitz continuous on the polyhedron $F = \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^s \mid \exists y \in \mathbb{R}^m : Ay \leq Tx + z\}$.*

Proof. By [19], $\emptyset \neq \operatorname{dom} f \subseteq \operatorname{dom} \Psi$ implies $\operatorname{dom} \Psi = F$. Consequently, the linear program in the definition of $f(x,z)$ is solvable for any $(x,z) \in F$ by parametric linear programming theory (see [2]). Consider any $(x,z), (x',z') \in F$. Without loss of generality, assume that $f(x,z) \geq f(x',z')$ and let $y' \in \Psi(x',z')$ be such that $f(x',z') = c^\top x' + q^\top y'$. Following [37] we obtain

$$\begin{aligned} |f(x,z) - f(x',z')| \; &= \; f(x,z) - c^\top x' - q^\top y' \; \leq \; c^\top x + q^\top y - c^\top x' - q^\top y' \\ &\leq \; \|c\| \|x - x'\| + \|q\| \|y - y'\| \end{aligned}$$

for any $y \in \Psi(x,z)$, where $\| \cdot \|$ denotes the Euclidean norm. Let $\mathbb{B}$ denote the Euclidean unit ball, then Theorem A.1 in the appendix yields

$$\Psi(x',z') \subseteq \Psi(x,z) + \Lambda \|(x,z) - (x',z')\| \mathbb{B}$$

and hence $|f(x,z) - f(x',z')| \; \leq \; (\|c\| + \Lambda \|q\|) \|(x,z) - (x',z')\|$. ∎

Remark 2.2. An alternate proof for Lemma 2.1 is given in [33, Theorem 1]. However, in view of Theorem A.1 in the appendix, Lemma 2.1 can be easily extended to the case of a convex quadratic lower level problem with linear constraints.

The next result follows directly from linear programming theory and provides verifiable conditions for dom $f \neq \emptyset$:

LEMMA 2.3. dom $f \neq \emptyset$ *holds if and only if there exists* $(x,z) \in \mathbb{R}^n \times \mathbb{R}^s$ *such that*

1. $\{y \mid Ay \leq Tx + z\}$ *is nonempty,*
2. *there is some* $u \in \mathbb{R}^s$ *satisfying* $A^\top u = d$ *and* $u \leq 0$*, and*
3. *the function* $y \mapsto q^\top y$ *is bounded from below on* $\Psi(x,z)$*.*

Under these conditions,

$$\min_y \{ q^\top y \mid y \in \Psi(x',z') \}$$

is attained for any $(x',z') \in F$*.*

A bilevel stochastic program arises from (2.1) if we assume the parameter $z = Z(\omega)$ to be the realization of a known random vector Z defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We impose an additional nonanticipativity constraint that creates the following pattern of decision and observation:

$$\text{Leader decides } x \quad \to \quad z = Z(\omega) \text{ is revealed} \quad \to \quad \text{Follower decides } y.$$

Throughout the analysis, we assume the stochasticity to be purely exogenous, i.e., the distribution of Z to be independent of x. It is well known that stochastic programs with decision dependent uncertainty pose additional conceptual challenges and are significantly harder to solve (see [29] for an outline of the underlying principles and [35] for an algorithm for a manageable subclass of stochastic programs with recourse).

We shall denote the Borel probability measure induced by the random vector Z by $\mu_Z := \mathbb{P} \circ Z^{-1} \in \mathcal{P}(\mathbb{R}^s)$. Moreover, we assume that the domain of f is nonempty and that the lower level problem is feasible for any leader's decision and any realization of the randomness, i.e.,

$$X \subseteq F_Z := \{x \in \mathbb{R}^n \mid (x, z) \in F \ \forall z \in \operatorname{supp} \mu_Z\}.$$

A similar assumption in two-stage stochastic programming is known as relatively complete recourse (cf. [61, sect. 2.1.3]). Let $L^0(\Omega, \mathcal{F}, \mathbb{P})$ denote the space of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, then we have $f(x, Z(\cdot)) \in L^0(\Omega, \mathcal{F}, \mathbb{P})$ for any leader's decision $x \in X$ by Lemma 2.1. Fixing any mapping $\mathcal{R} : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X}$ is a linear subspace of $L^0(\Omega, \mathcal{F}, \mathbb{P})$ that contains the constants and satisfies

$$\{f(x, Z(\cdot)) \mid x \in X\} \subseteq \mathcal{X},$$

we may thus consider the bilevel stochastic linear program

$$\min_x \{\mathcal{R}[f(x, Z(\cdot))] \mid x \in X\}. \tag{2.2}$$

Under suitable moment or boundedness conditions on Z, the classical L^p-spaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty]$ are natural choices for the domain $\mathcal{X}$ of $\mathcal{R}$. Let

$$\mathcal{M}_s^p := \left\{\mu \in \mathcal{P}(\mathbb{R}^s) \mid \int_{\mathbb{R}^s} \|z\|^p \ \mu(dz) < \infty\right\}$$

denote the set of Borel probability measures on $\mathbb{R}^s$ with finite moments of order $p \in [1, \infty)$ and set

$$\mathcal{M}_s^\infty := \{\mu \in \mathcal{P}(\mathbb{R}^s) \mid \operatorname{supp} \mu_Z \text{ is bounded}\}.$$

Lemma 2.4. *Assume* dom $f \neq \emptyset$ *and* $\mu_Z \in \mathcal{M}_s^p$ *for some* $p \in [1, \infty]$. *Then the mapping* $\mathbb{F} : F_Z \to L^0(\Omega, \mathcal{F}, \mathbb{P})$ *given by* $\mathbb{F}(x) := f(x, Z(\cdot))$ *takes values in* $L^p(\Omega, \mathcal{F}, \mathbb{P})$ *and is Lipschitz continuous with respect to (w.r.t.) the* L^p*-norm.*

Proof. We first consider the case that p is finite. By $(0, 0) \in F$ and Lemma 2.1, there exists a constant L_f such that

$$\begin{aligned}
\|\mathbb{F}(x)\|_{L^p}^p &\leq 2^p |f(0,0)|^p + 2^p \int_{\mathbb{R}^s} |f(x,z) - f(0,0)|^p \ \mu_Z(dz) \\
&\leq 2^p |f(0,0)|^p + 2^p L_f^p \|x\|^p + 2^p L_f^p \int_{\mathbb{R}^s} \|z\|^p \ \mu_Z(dz) < \infty
\end{aligned}$$

holds for any $x \in F_Z$. Furthermore, for any $x, x' \in F_Z$ we have

$$\|\mathbb{F}(x) - \mathbb{F}(x')\|_{L^p} = \left(\int_{\mathbb{R}^s} |f(x,z) - f(x',z)|^p \; \mu_Z(dz) \right)^{1/p} \leq L_f \|x - x'\|.$$

For $p = \infty$, Lemma 2.1 implies that for any fixed $x \in F_Z$, the mapping $f(x, \cdot)$ is continuous on supp μ_Z. Thus, $\mu_Z \in \mathcal{M}_s^\infty$ yields

$$\|\mathbb{F}(x)\|_{L^\infty} \leq \sup_{z \in \operatorname{supp} \mu_Z} |f(x,z)| < \infty.$$

Moreover,

$$\|\mathbb{F}(x) - \mathbb{F}(x')\|_{L^\infty} \leq \sup_{z \in \operatorname{supp} \mu_Z} |f(x,z) - f(x',z)| \leq L_f \|x - x'\|,$$

holds for any $x, x' \in F_Z$. ∎

The article [33] examines the case where $\mathcal{R}$ is given by the value-at-risk. We shall consider coherent (cf. [1], [24]) or convex (cf. [21], [23]) risk measures. A thorough discussion of their analytical traits is provided in [22].

Definition 2.5. *A mapping $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ defined on some linear subspace $\mathcal{X}$ of $L^0(\Omega, \mathcal{F}, \mathbb{P})$ containing the constants is called a* convex risk measure *if the following conditions are fulfilled:*

1. *(Convexity) For any $Y_1, Y_2 \in \mathcal{X}$ and $\lambda \in [0,1]$ we have*

$$\mathcal{R}[\lambda Y_1 + (1-\lambda) Y_2] \leq \lambda \mathcal{R}[Y_1] + (1-\lambda)\mathcal{R}[Y_2].$$

2. *(Monotonicity) $\mathcal{R}[Y_1] \leq \mathcal{R}[Y_2]$ for all $Y_1, Y_2 \in \mathcal{X}$ satisfying $Y_1 \leq Y_2$ with respect to the $\mathbb{P}$-almost sure partial order.*
3. *(Translation equivariance) $\mathcal{R}[Y + t] = \mathcal{R}[Y] + t$ for all $Y \in \mathcal{X}$ and $t \in \mathbb{R}$.*

   A convex risk measure $\mathcal{R}$ is coherent *if the following holds true.*
4. *(Positive homogeneity) $\mathcal{R}[tY] = t \cdot \mathcal{R}[Y]$ for all $Y \in \mathcal{X}$ and $t \in [0, \infty)$.*

Definition 2.6. *A mapping $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is called* law-invariant *if for all $Y_1, Y_2 \in \mathcal{X}$ with $\mathbb{P} \circ Y_1^{-1} = \mathbb{P} \circ Y_2^{-1}$ we have $\mathcal{R}[Y_1] = \mathcal{R}[Y_2]$.*

Example 2.7. Below we list some risk functionals that are commonly used in stochastic programming (cf. [52], [61, sect. 6.3.2]):

1. The *expectation* $\mathbb{E} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$,

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

   is a law-invariant and coherent risk measure.
2. The *expected excess of order* $p \in [1, \infty)$ over a predefined level $\eta \in \mathbb{R}$ is the mapping $\mathrm{EE}_\eta^p : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ given by

$$\mathrm{EE}_\eta^p[Y] := \Big(\mathbb{E}\big[\max\{Y - \eta, 0\}^p \big] \Big)^{1/p}.$$

   EE_η^p is law-invariant, convex, and nondecreasing, but neither positively homogeneous nor translation-equivariant (cf. [61, Example 6.22]).

3. The *mean upper semideviation of order* $p \in [1, \infty)$ is the mapping $\mathrm{SD}^p_\rho : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ defined by

$$\mathrm{SD}^p_\rho[Y] := \mathbb{E}[Y] + \rho \cdot \mathrm{EE}^p_{\mathbb{E}[Y]}[Y] = \mathbb{E}[Y] + \rho \cdot \Big(\mathbb{E}\big[\max\{Y - \mathbb{E}[Y], 0\}^p\big]\Big)^{1/p},$$

where $\rho \in (0, 1]$ is a parameter. SD^p_ρ is a law-invariant coherent risk measure (cf. [61, Example 6.20]).

4. The *conditional value-at-risk* $\mathrm{CVaR}_\alpha : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ at level $\alpha \in (0, 1)$ given by

$$\mathrm{CVaR}_\alpha[Y] := \inf\left\{\eta + \frac{1}{1-\alpha}\mathrm{EE}^1_\eta[Y] \mid \eta \in \mathbb{R}\right\}$$

is a law-invariant coherent risk measure (cf. [51, Proposition 2]). The variational representation above was established in [55, Theorem 10].

5. The *entropic risk measure* $\mathrm{Entr}_\alpha : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ defined by

$$\mathrm{Entr}_\alpha[Y] := \frac{1}{\alpha}\ln\Big(\mathbb{E}\big[\exp(\alpha Y)\big]\Big),$$

where $\alpha > 0$ is a parameter, is a law-invariant convex (but not coherent) risk measure (cf. [22, Example 4.13, Example 4.34]).

6. The *worst-case risk measure* $\mathcal{R}_{\max} : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ given by

$$\mathcal{R}_{\max}[Y] := \sup_{\omega \in \Omega} Y(\omega)$$

is law-invariant and coherent (cf. [22, Example 4.8]). This choice of $\mathcal{R}$ in (2.2) leads to a bilevel robust problem. $\mathcal{R}_{\max}$ only depends on the so-called *uncertainty set* $Z(\Omega) \subseteq \mathbb{R}^s$. Thus, the bilevel robust problem can be formulated without knowledge of the distribution of the uncertain parameter. In robust optimization, the uncertainty set is often assumed to be finite, polyhedral, or ellipsoidal (cf. [4]).

The entropic risk measure is an example of a so-called certainty equivalent:

DEFINITION 2.8. *The* certainty equivalent *associated with a strictly increasing convex disutility function* $u : \mathbb{R} \to \mathbb{R}$ *is the mapping* $\mathrm{CE}_u : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ *defined by*

$$\mathrm{CE}_u[Y] = u^{-1}\big(\mathbb{E}[u(Y)]\big).$$

Certainty equivalents are law-invariant, monotone, quasi-convex, and normalized, i.e., $\mathrm{CE}_u[0] = 0$ (cf. [48, Theorem 2.1, Remark 2.3]), but not translation equivariant or convex in general. It is well known that the latter two properties hold if and only if u is linear or exponential (cf. [22, Proposition 2.46]).

3. Lipschitzian properties.

Aiming at sufficient conditions for the existence of optimal solutions for the bilevel stochastic linear program (2.2), we examine Lipschitzian properties of the objective function $\mathcal{Q}_\mathcal{R} : F_Z \to \mathbb{R}$,

$$\mathcal{Q}_\mathcal{R}(x) := \mathcal{R}[\mathbb{F}(x)].$$

THEOREM 3.1. *Assume* $\operatorname{dom} f \neq \emptyset$ *and* $\mu_Z \in \mathcal{M}^p_s$ *for some* $p \in [1, \infty]$. *Then the following statements hold true for any* $\mathcal{R} : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$:

1. *$\mathcal{Q}_{\mathcal{R}}$ is locally Lipschitz continuous if $\mathcal{R}$ is convex and continuous.*
2. *$\mathcal{Q}_{\mathcal{R}}$ is locally Lipschitz continuous if $\mathcal{R}$ is convex and nondecreasing.*
3. *$\mathcal{Q}_{\mathcal{R}}$ is locally Lipschitz continuous if $\mathcal{R}$ is a convex risk measure.*
4. *$\mathcal{Q}_{\mathcal{R}}$ is Lipschitz continuous if $\mathcal{R}$ is Lipschitz continuous.*
5. *$\mathcal{Q}_{\mathcal{R}}$ is Lipschitz continuous if $\mathcal{R}$ is a coherent risk measure.*

Proof. 1. It is well known that any real-valued convex and continuous mapping on a normed space is locally Lipschitz continuous (cf. [20]). The result is thus an immediate consequence of Lemma 2.4.

2. Any real-valued, convex, and nondecreasing functional on the Banach lattice $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is continuous (see, e.g., [11, Theorem 4.1]).

3. By definition, any convex risk measure is convex and nondecreasing.

4. This is a straightforward conclusion from Lemma 2.4.

5. Any coherent risk measure on $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is Lipschitz continuous by [32, Lemma 2.1]. ∎

Remark 3.2. Any coherent risk measure $\mathcal{R} : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is Lipschitz continuous with constant 1 by [22, Lemma 4.3]. Concrete Lipschitz constants for continuous coherent law-invariant risk measures $\mathcal{R} : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ with $p \in [1, \infty)$ may be obtained from representation results (see, e.g., [3]).

Corollary 3.3. *Assume* dom $f \neq \emptyset$, *$\mu_Z \in \mathcal{M}^p_s$ for some $p \in [1, \infty]$ and let $X \subseteq F_Z$ be nonempty and compact. Then* (2.2) *is solvable for any convex and nondecreasing mapping $\mathcal{R} : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$.*

While Theorem 3.1 only yields local Lipschitz continuity of $\mathcal{Q}_{\mathrm{EE}_\eta}$, it is easy to see that this mapping is actually Lipschitz continuous on F_Z.

Lemma 3.4. *Assume* dom $f \neq \emptyset$ *and $\mu_Z \in \mathcal{M}^1_s$, then $\mathcal{Q}_{\mathrm{EE}_\eta}$ is Lipschitz continuous for any $\eta \in \mathbb{R}$.*

Proof. By Lemma 2.1 there exists a constant L_f such that

$$\begin{aligned}|\mathcal{Q}_{\mathrm{EE}_\eta}(x) - \mathcal{Q}_{\mathrm{EE}_\eta}(x')| &\leq \int_{\mathbb{R}^s} |\max\{f(x,z) - \eta, 0\} - \max\{f(x',z) - \eta, 0\}| \; \mu_Z(dz) \\ &\leq \int_{\mathbb{R}^s} L_f \|(x,z) - (x',z)\| \; \mu_Z(dz) = L_f \|x - x'\|\end{aligned}$$

holds for any $x, x' \in F_Z$. ∎

Remark 3.5. The proof of Lemma 3.4 only requires the existence of some measurable mapping $L_f : \mathbb{R}^s \to \mathbb{R}$ such that

$$\|f(x,z) - f(x',z)\| \leq L_f(z)\|x - x'\| \;\; \forall x, x' \in F_Z \;\; \forall z \in \operatorname{supp} \mu_Z$$

and $\mathbb{P} \circ Z^{-1} \circ L_f^{-1} \in \mathcal{M}^1_1$. This may allow us to extend the result to more general models.

Due to the lack of convexity, Theorem 3.1 does not apply to $\mathcal{Q}_{\mathrm{CE}_u}$ in general. However, Lipschitz continuity is guaranteed under the strong assumption that u is bi-Lipschitz.

Lemma 3.6. *Let $u : \mathbb{R} \to \mathbb{R}$ be strictly increasing and bi-Lipschitz, i.e., there exists a finite constant $L \geq 1$ such that*

$$\frac{1}{L}\|t - t'\| \leq \|u(t) - u(t')\| \leq L\|t - t'\|$$

holds for all $t, t' \in \mathbb{R}$, then CE_u *is Lipschitz continuous.*

Proof. For any $Y, Y' \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ we have

$$\begin{aligned}\big|\mathrm{CE}_u(Y) - \mathrm{CE}_u(Y')\big| &= \big|u^{-1}\big(\mathbb{E}[u(Y)]\big) - u^{-1}\big(\mathbb{E}[u(Y')]\big)\big| \le L\big|\mathbb{E}[u(Y)] - \mathbb{E}[u(Y')]\big| \\ &\le L\|u(Y) - u(Y')\|_\infty = L^2\|Y - Y'\|_\infty,\end{aligned}$$

where the second inequality is justified by [22, Lemma 4.3] (see also Remark 3.2). ∎

4. Differentiability. We shall now focus on differentiability of mappings $\mathcal{Q}_{\mathcal{R}}$ that are defined via expectations (of Lipschitzian transformations) of f. In particular, we examine $\mathcal{Q}_{\mathbb{E}}$, $\mathcal{Q}_{\mathrm{EE}_\eta}$, and $\mathcal{Q}_{\mathrm{SD}_\rho}$ as well as some goal functions derived from certainty equivalents including $\mathcal{Q}_{\mathrm{Entr}_\alpha}$. While $f(\cdot, Z)$ is well known to be nondifferentiable in general (cf. [15, Figure 3.1]), we shall show that under mild assumptions, the aforementioned mappings are continuously differentiable at any x_0 for which a finite union of hyperplanes in $\mathbb{R}^s$ depending on x_0 is a null set w.r.t. the underlying probability measure μ_Z. Note that the latter condition holds automatically if μ_Z is absolutely continuous w.r.t. the Lebesgue measure.

While our primary motivation for the investigation of differentiability is the establishment of simple optimality conditions, the development of solution methods that exploit continuous differentiability of the objective function of the bilevel stochastic linear program remains a promising topic for future research.

It will be convenient to reformulate f as

$$f(x, z) = c^\top x + \min_{y_+, y_-, t}\big\{q^\top(y_+ - y_-) \mid (y_+, y_-, t) \in \Psi_=(x, z)\big\},$$

where

$$\Psi_=(x, z) = \operatorname*{Argmin}_{y_+, y_-, t}\{d^\top(y_+ - y_-) \mid A(y_+ - y_-) + t = Tx + z,\ y_+, y_-, t \ge 0\}.$$

Setting

$$\hat{q} := \begin{pmatrix} q \\ -q \\ 0_s \end{pmatrix}, \quad \hat{y} := \begin{pmatrix} y_+ \\ y_- \\ t \end{pmatrix}, \quad \hat{d} := \begin{pmatrix} d \\ -d \\ 0_s \end{pmatrix}, \quad \text{and } \hat{A} := (A, -A, I_s),$$

we obtain

$$f(x, z) = c^\top x + \min_{\hat{y}}\Big\{\hat{q}^\top \hat{y} \mid \hat{y} \in \operatorname*{Argmin}_{\hat{y}'}\{\hat{d}^\top \hat{y}' \mid \hat{A}\hat{y}' = Tx + z,\ \hat{y}' \ge 0\}\Big\}.$$

As the matrix $\hat{A}$ contains a unit matrix, its rows are linearly independent and we may consider the nonempty set

$$\mathcal{A} := \{\hat{A}_B \in \mathbb{R}^{s \times s} \mid \hat{A}_B \text{ is a regular submatrix of } \hat{A}\}$$

of lower level base matrices. A base matrix $\hat{A}_B \in \mathcal{A}$ is optimal for the lower level problem for a given (x, z) if it is feasible, i.e., $\hat{A}_B^{-1}(Tx + z) \ge 0$, and the associated reduced cost vector $\hat{d}_N^\top - \hat{d}_B^\top \hat{A}_B^{-1}\hat{A}_N$ is nonnegative. Furthermore, for any optimal base matrix $\hat{A}_{B'} \in \mathcal{A}$, there exists a feasible base matrix $\hat{A}_B \in \mathcal{A}$ satisfying

$$\hat{A}_{B'}^{-1}(Tx + z) = \hat{A}_B^{-1}(Tx + z) \quad \text{and} \quad \hat{d}_N^\top - \hat{d}_B^\top \hat{A}_B^{-1}\hat{A}_N \ge 0.$$

Set

$$\mathcal{A}^* := \{\hat{A}_B \in \mathcal{A} \mid \hat{d}_N^\top - \hat{d}_B^\top \hat{A}_B^{-1}\hat{A}_N \geq 0\}$$

and assume dom $f \neq \emptyset$, then

$$f(x,z) = c^\top x + \min_{\hat{A}_B} \left\{ \hat{q}_B^\top \hat{A}_B^{-1}(Tx+z) \mid \hat{A}_B^{-1}(Tx+z) \geq 0,\ \hat{A}_B \in \mathcal{A}^* \right\} \tag{4.1}$$

holds for any $(x,z) \in F$.

The set of parameters $(x,z) \in F$ on which a certain base matrix $\hat{A}_B \in \mathcal{A}^*$ is optimal for the lower level problem is called the region of stability associated with that particular base matrix (cf. [15, Definition 3.3]). For the present analysis, it will be convenient to also take into account the effect the base matrix has on the leader's objective function.

DEFINITION 4.1. *The* region of strong stability *associated with a base matrix $\hat{A}_B \in \mathcal{A}^*$ is the set of parameters $(x,z) \in F$ on which $\hat{A}_B$ is not only optimal for the lower level problem, but also most favorable among the optimal base matrices w.r.t. the leader's objective function, i.e.,*

$$\mathcal{S}(\hat{A}_B) := \{(x,z) \in F \mid \hat{A}_B^{-1}(Tx+z) \geq 0,\ c^\top x + \hat{q}_B^\top \hat{A}_B^{-1}(Tx+z) = f(x,z)\}.$$

In view of (4.1), we have

$$F = \bigcup_{\hat{A}_B \in \mathcal{A}^*} \mathcal{S}(\hat{A}_B).$$

Recall that the domain of f is empty or coincides with F by Lemma 2.1. On $\mathcal{S}(\hat{A}_B)$, f coincides with the affine linear mapping

$$f(x,z) = c^\top x + \hat{q}_B^\top \hat{A}_B^{-1}(Tx+z)$$

by definition. Consequently, $f(\cdot, z_0)$ is continuously differentiable at $x_0 \in \operatorname{int} F_Z$ whenever (x_0, z_0) belongs to the interior of some region of strong stability. The following result shows that this is the case for any $z_0 \in \operatorname{supp} \mu_Z \setminus \mathcal{N}_{x_0}$, where the set $\mathcal{N}_{x_0}$ is given as a union of three types of sets directly related to the three conditions defining a region of strong stability and contained in a finite union of hyperplanes.

LEMMA 4.2. *Assume* dom $f \neq \emptyset$ *and let x_0 be an interior point of F_Z. Then $f(\cdot, z_0)$ is continuously differentiable at x_0 for any $z_0 \in \operatorname{supp} \mu_Z \setminus \mathcal{N}_{x_0}$, where*

$$\mathcal{N}_{x_0} = \mathcal{F}_{x_0} \ \cup \bigcup_{\hat{A}_B \in \mathcal{A}^*} \mathcal{Z}_{x_0}(\hat{A}_B) \setminus \operatorname{int} \mathcal{Z}_{x_0}(\hat{A}_B) \ \cup \bigcup_{\substack{\hat{A}_B, \hat{A}_{B'} \in \mathcal{A}^*: \\ \hat{q}_B^\top \hat{A}_B^{-1} \neq \hat{q}_{B'}^\top \hat{A}_{B'}^{-1}}} \mathcal{V}_{x_0}(\hat{A}_B, \hat{A}_{B'})$$

with

$$\mathcal{F}_{x_0} := \{z \in \mathbb{R}^s \mid (x_0, z) \in F \setminus \operatorname{int} F\},$$
$$\mathcal{Z}_{x_0}(\hat{A}_B) := \{z \in \mathbb{R}^s \mid \hat{A}_B^{-1}(Tx_0 + z) \geq 0\}, \text{ and}$$
$$\mathcal{V}_{x_0}(\hat{A}_B, \hat{A}_{B'}) := \{z \in \mathbb{R}^s \mid (\hat{q}_B^\top \hat{A}_B^{-1} - \hat{q}_{B'}^\top \hat{A}_{B'}^{-1})(z + Tx_0) = 0\}.$$

Furthermore, $\mathcal{N}_{x_0}$ is contained in a finite union of affine hyperplanes in $\mathbb{R}^s$ and we have

$$\nabla_x f(x_0, z_0) \in \{c^\top + \hat{q}_B^\top \hat{A}_B^{-1} T \mid \hat{A}_B \in \mathcal{A}^*\}.$$

Proof. $x_0 \in \operatorname{int} F_Z$ and $z_0 \in \operatorname{supp} \mu \setminus \mathcal{N}_{x_0} \subseteq \operatorname{supp} \mu \setminus \mathcal{F}_{x_0}$ imply $(x_0, z_0) \in \operatorname{int} F$ by definition. If $(x_0, z_0) \in \operatorname{int} \mathcal{S}(\hat{A}_B)$ holds for some $\hat{A}_B \in \mathcal{A}^*$, there is a neighborhood U of x_0 such that $f(x, z_0) = c^\top x + \hat{q}_B^\top \hat{A}_B^{-1}(Tx + z_0)$ holds for all $x \in U$. In particular, $f(\cdot, z_0)$ is continuously differentiable at x_0 and $\nabla_x f(x_0, z_0) = c^\top + \hat{q}_B^\top \hat{A}_B^{-1} T$.

Suppose that $(x_0, z_0) \notin \operatorname{int} \mathcal{S}(\hat{A}_B)$ for all $\hat{A}_B \in \mathcal{A}^*$. The continuity of f implies that there are $k \geq 2$ pairwise different base matrices $\hat{A}_{B^1}, \dots, \hat{A}_{B^k} \in \mathcal{A}^*$ such that

$$(x_0, z_0) \in \bigcap_{i=1,\dots,k} \mathcal{S}(\hat{A}_{B^i}) \ \cap \ \operatorname{int} \left(\bigcup_{i=1,\dots,k} \mathcal{S}(\hat{A}_{B^i}) \right).$$

In particular, we have

$$\hat{q}_{B^1}^\top \hat{A}_{B^1}^{-1}(Tx_0 + z_0) = \cdots = \hat{q}_{B^k}^\top \hat{A}_{B^k}^{-1}(Tx_0 + z_0),$$

i.e., $z_0 \in \mathcal{V}_{x_0}(\hat{A}_{B^i}, \hat{A}_{B^j})$ for all $i, j \in \{1, \dots, k\}$. Thus, $z_0 \in \operatorname{supp} \mu \setminus \mathcal{N}_{x_0}$ implies

$$\hat{q}_{B^1}^\top \hat{A}_{B^1}^{-1} = \cdots = \hat{q}_{B^k}^\top \hat{A}_{B^k}^{-1}. \tag{4.2}$$

For any $i \in \{1, \dots, k\}$ we shall consider the sets

$$\begin{aligned} &\mathcal{Z}(\hat{A}_{B^i}) := \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s \mid \hat{A}_{B^i}^{-1}(Tx + z) \geq 0\} \ \text{ and} \\ &\mathcal{O}(\hat{A}_{B^i}) := \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s \mid c^\top x + \hat{q}_{B^i}^\top \hat{A}_{B^i}^{-1}(Tx + z) = f(x, z)\}. \end{aligned}$$

By (4.2) we have

$$\mathcal{S}(\hat{A}_{B^i}) = \mathcal{O}(\hat{A}_{B^i}) \cap \mathcal{Z}(\hat{A}_{B^i}) = \mathcal{O}(\hat{A}_{B^1}) \cap \mathcal{Z}(\hat{A}_{B^i})$$

for all $i \in \{1, \dots, k\}$. Thus,

$$\begin{aligned} (x_0, z_0) \in \operatorname{int} \bigcup_{i=1,\dots,k} \mathcal{S}(\hat{A}_{B^i}) \ &= \ \operatorname{int} \left(\mathcal{O}(\hat{A}_{B^1}) \ \cap \bigcup_{i=1,\dots,k} \mathcal{Z}(\hat{A}_{B^i}) \right) \\ &= \operatorname{int} \mathcal{O}(\hat{A}_{B^1}) \ \cap \ \operatorname{int} \bigcup_{i=1,\dots,k} \mathcal{Z}(\hat{A}_{B^i}). \end{aligned}$$

We have $(x_0, z_0) \in \mathcal{Z}(\hat{A}_{B^1})$, i.e., $z_0 \in \mathcal{Z}_{x_0}(\hat{A}_{B^1})$. Thus, $z_0 \in \operatorname{supp} \mu \setminus \mathcal{N}_{x_0}$ implies $z_0 \in \operatorname{int} \mathcal{Z}_{x_0}(\hat{A}_{B^1})$. Consequently, there is a neighborhood W of z_0 such that $\hat{A}_{B^1}^{-1}(Tx_0 + z) \geq 0$ for all $z \in W$. This implies $(x_0, z_0) \in \operatorname{int} \mathcal{Z}(\hat{A}_{B^1})$ for continuity reasons. Hence, $(x_0, z_0) \in \operatorname{int} \mathcal{O}(\hat{A}_{B^1}) \cap \operatorname{int} \mathcal{Z}(\hat{A}_{B^1}) = \operatorname{int} \mathcal{S}(\hat{A}_{B^1})$, which contradicts $(x_0, z_0) \notin \operatorname{int} \mathcal{S}(\hat{A}_B)$ for all $\hat{A}_B \in \mathcal{A}^*$.

It remains to show that $\mathcal{N}_{x_0}$ is contained in a finite union of affine hyperplanes. Suppose that z is such that $Ay < Tx_0 + z$ holds for some $y \in \mathbb{R}^m$, then $(x_0, z) \in \operatorname{int} F$. Consequently,

$$\mathcal{F}_{x_0} \subseteq \bigcup_{i=1,\dots,s} \{z \in \mathbb{R}^s \mid (x, z_0) \in F, \ e_i^\top Ay = e_i^\top (Tx_0 + z) \ \forall y : \ Ay \leq Tx_0 + z\}$$

is contained in a finite union of affine hyperplanes. Similarly, we have

$$\mathcal{Z}_{x_0}(\hat{A}_B) \setminus \operatorname{int} \mathcal{Z}_{x_0}(\hat{A}_B) \subseteq \bigcup_{i=1,\dots,s} \{z \in \mathbb{R}^s \mid \hat{A}_B^{-1}(Tx_0 + z) \leq 0, \ e_i^\top (\hat{A}_B^{-1}(Tx_0 + z)) = 0\}$$

$\hat{A}_B \in \mathcal{A}^*$, where $e_i^\top \hat{A}_B^{-1} \neq 0$ due to the regularity of $\hat{A}_B$. Finally, $\mathcal{V}_{x_0}(\hat{A}_B, \hat{A}_{B'})$ is an affine hyperplane for any $\hat{A}_B, \hat{A}_{B'} \in \mathcal{A}^*$ satisfying $\hat{q}_B^\top \hat{A}_B^{-1} \neq \hat{q}_{B'}^\top \hat{A}_{B'}^{-1}$. ∎

By the above considerations, the set

$$\{z \in \text{supp } \mu_Z \mid f(\cdot, z) \text{ is not differentiable at } x_0\}$$

is contained in $\mathcal{N}_{x_0}$. Assuming $\mu_Z[\mathcal{N}_{x_0}] = 0$ and invoking Lipschitz continuity of f, differentiability of $\mathcal{Q}_{\mathbb{E}}(x) = \int_{\mathbb{R}^s} f(x,z)\ \mu_Z(dz)$ at x_0 follows from Lebesgue's dominated convergence theorem, which justifies differentiation under the integral sign.

THEOREM 4.3. *Assume* dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}_s^1$, *and let* $x_0 \in \text{int } F_Z$ *be such that* $\mu_Z[\mathcal{N}_{x_0}] = 0$. *Then* $\mathcal{Q}_{\mathbb{E}}$ *is continuously differentiable at* x_0 *and*

$$\nabla \mathcal{Q}_{\mathbb{E}}(x_0) = \int_{\text{supp } \mu_Z \setminus \mathcal{N}_{x_0}} \nabla_x f(x_0, z)\ \mu_Z(dz).$$

Proof. We shall prove that Lemma A.2 in the appendix is applicable. First, note that condition (a) is satisfied by $\mu_Z[\mathcal{N}_{x_0}] = 0$ and Lemma 4.2. Furthermore, by $x_0 \in \text{int } F_Z$ there is neighborhood U of x_0 that is contained in F_Z. In particular, $\mathcal{Q}_{\mathbb{E}}$ is well defined and finite by Lemma 2.4, i.e., the first part of condition (b) of Lemma A.2 is satisfied. To see that the second part holds as well, let L denote the Lipschitz constant from Lemma 2.1. Fix any $x \in U \setminus \{x_0\}$ and $z_0 \in \text{supp } \mu_Z \setminus \mathcal{N}_{x_0}$, then

$$\|x - x_0\|^{-1} \big| f(x, z_0) - f(x_0, z_0) - \nabla_x f(x_0, z_0)(x - x_0) \big| \ \leq\ L + \max_{\hat{A}_B \in \mathcal{A}^*} \|\hat{q}_B^\top \hat{A}_B^{-1} B\|$$

follows immediately from the characterization of the derivative in Lemma 4.2. Thus, Lemma A.2 yields the differentiability of $\mathcal{Q}_{\mathbb{E}}$.

We shall now prove that the derivative is indeed continuous. By construction, there exists a neighborhood $U \subseteq \text{int } F_Z$ of x_0 such that $\mathcal{N}_x \subseteq \mathcal{N}_{x_0}$ holds for any $x \in U$. Consequently, by $\mu_Z[\mathcal{N}_x] = 0$ and the previous arguments, $\mathcal{Q}_{\mathbb{E}}$ is differentiable at any $x \in U$ and we have

$$\nabla \mathcal{Q}_{\mathbb{E}}(x) = \int_{\text{supp } \mu_Z \setminus \mathcal{N}_x} \nabla_x f(x, z)\ \mu_Z(dz) = c^\top + \sum_{\Delta \in D} \mu_Z[\mathcal{W}(x, \Delta)]\Delta, \tag{4.3}$$

where $D := \{\hat{q}_B^\top \hat{A}_B^{-1} T \mid \hat{A}_B \in \mathcal{A}^*\}$ and

$$\mathcal{W}(x, \Delta) := \{z \in \text{supp } \mu_Z \setminus \mathcal{N}_x \mid \exists \hat{A}_B \in \mathcal{A}^* : (x, z) \in \text{int } \mathcal{S}(\hat{A}_B),\ \Delta = \hat{q}_B^\top \hat{A}_B^{-1} T\}.$$

By Lemma A.3 in the appendix, the set-valued mapping $\overline{\mathcal{W}} : \mathbb{R}^n \times D \rightrightarrows \mathbb{R}^s$,

$$\begin{aligned}\overline{\mathcal{W}}(x, \Delta) &= \{z \in \text{supp } \mu_Z \mid \exists \hat{A}_B \in \mathcal{A}^* : (x, z) \in \text{cl int } \mathcal{S}(\hat{A}_B),\ \Delta = \hat{q}_B^\top \hat{A}_B^{-1} T\} \\ &= \Bigg\{z \in \text{supp } \mu_Z \mid (x, z) \in \bigcup_{\hat{A}_B \in \mathcal{A}^* :\, \hat{q}_B^\top \hat{A}_B^{-1} T = \Delta} \text{cl int } \mathcal{S}(\hat{A}_B)\Bigg\}\end{aligned}$$

is outer semicontinuous. Furthermore, by the arguments used in the proof of Lemma 4.2 we obtain

$$\overline{\mathcal{W}}(x, \Delta) \setminus \mathcal{W}(x, \Delta) \subseteq \mathcal{N}_x$$

and thus $\mu_Z[\mathcal{N}_x] = 0$ implies $\mu_Z[\mathcal{W}(x, \Delta)] = \mu_Z[\overline{\mathcal{W}}(x, \Delta)]$.

We shall use the above representation to prove that for any $\Delta \in D$, the mapping $M_\Delta : \mathbb{R}^n \to \mathbb{R}$, $M_\Delta(x) := \mu_Z[\mathcal{W}(x, \Delta)]$ is continuous at x_0. Consider any sequence

$\{x_l\}_{l\in\mathbb{N}} \subset \mathbb{R}^n$ that converges to x_0. Without loss of generality we may assume that $x_l \in U$ holds for all $l \in \mathbb{N}$. We have

$$\begin{aligned}\limsup_{l\to\infty} M_\Delta(x_l) &= \limsup_{l\to\infty} \mu_Z[\overline{\mathcal{W}}(x_l,\Delta)] = \limsup_{l\to\infty} \int_{\overline{\mathcal{W}}(x_l,\Delta)} 1\ \mu_Z(dz)\\ &= \limsup_{l\to\infty} \int_{\mathrm{supp}\ \mu_Z} 1_{\overline{\mathcal{W}}(x_l,\Delta)}(z)\ \mu_Z(dz)\\ &\le \int_{\mathrm{supp}\ \mu_Z} \limsup_{l\to\infty}\ 1_{\overline{\mathcal{W}}(x_l,\Delta)}(z)\ \mu_Z(dz),\end{aligned}$$

where

$$1_{\overline{\mathcal{W}}(x_l,\Delta)} := \begin{cases} 1 & \text{if } z \in \overline{\mathcal{W}}(x_l,\Delta),\\ 0 & \text{else}\end{cases}$$

denotes the indicator function associated with the set $\overline{\mathcal{W}}(x_l,\Delta)$ and the final inequality is obtained by using Fatou's lemma. We shall show that

$$\limsup_{l\to\infty} 1_{\overline{\mathcal{W}}(x_l,\Delta)}(z) \le 1_{\limsup_{l\to\infty}\overline{\mathcal{W}}(x_l,\Delta)}(z) \tag{4.4}$$

holds for any $z \in \mathrm{supp}\ \mu_Z$. If the left-hand side in (4.4) equals zero, the above inequality holds because the right-hand side is nonnegative. On the other hand, $\limsup_{l\to\infty} 1_{\overline{\mathcal{W}}(x_l,\Delta)}(z) = 1$ implies that there is a subsequence $\{x'_l\}_{l\in\mathbb{N}}$ of $\{x_l\}_{l\in\mathbb{N}}$ such that $z \in \overline{\mathcal{W}}(x'_l,\Delta)$ holds for all $l \in \mathbb{N}$. Thus, $z \in \limsup_{l\to\infty}\overline{\mathcal{W}}(x_l,\Delta)$ by definition and (4.4) is satisfied.

Invoking (4.4) and the previous estimates we obtain

$$\begin{aligned}\limsup_{l\to\infty} M_\Delta(x_l) &\le \int_{\mathrm{supp}\ \mu_Z} 1_{\limsup_{l\to\infty}\overline{\mathcal{W}}(x_l,\Delta)}(z)\ \mu_Z(dz)\\ &\le \int_{\mathrm{supp}\ \mu_Z} 1_{\overline{\mathcal{W}}(x_0,\Delta)}(z)\ \mu_Z(dz)\\ &= \mu_Z[\overline{\mathcal{W}}(x_0,\Delta)] = M_\Delta(x_0),\end{aligned}$$

where the second inequality holds due to the outer semicontinuity of $\overline{W}$ and the monotonicity of the indicator function. Consequently, M_Δ is upper semicontinuous at x_0 for any $\Delta \in D$.

By $U \subseteq \mathrm{int}\ F_Z$ and the arguments used in the proof of Lemma (4.2),

$$\mathrm{supp}\ \mu_Z \subseteq \bigcup_{\Delta\in D} \mathcal{W}(x,\Delta) \tag{4.5}$$

holds for any $x \in U$. By $\mathcal{W}(x,\Delta_1) \cap \mathcal{W}(x,\Delta_2) = \emptyset$ for any $\Delta_1, \Delta_2 \in D$ satisfying $\Delta_1 \neq \Delta_2$, (4.5) implies

$$\sum_{\Delta\in D} M_\Delta(x) = \sum_{\Delta\in D} \mu_Z[\mathcal{W}(x,\Delta)] = 1$$

for any $x \in U$. Consequently, as M_Δ is upper semicontinuous at x_0 for any $\Delta \in D$, we obtain that

$$M_\Delta(x) = 1 - \sum_{\Delta'\in D\setminus\{\Delta\}} M_{\Delta'}(x)$$

is representable as a sum of functions that are lower semicontinuous at x_0. Thus, M_Δ is continuous at x_0 for any $\Delta \in D$, which implies the continuity of

$$\nabla \mathcal{Q}_{\mathbb{E}}(x) = c^\top + \sum_{\Delta \in D} M_\Delta(x)\Delta$$

at x_0. ∎

When working with the expected excess, the inner maximum may cause additional points of nondifferentiability.

THEOREM 4.4. *Assume* dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}_s^1$, *and let* $x_0 \in \operatorname{int} F_Z$ *and* $\eta \in \mathbb{R}$ *be such that* $\mu_Z[\mathcal{N}_{x_0} \cup \mathcal{L}(x_0, \eta)] = 0$, *where*

$$\mathcal{L}(x_0, \eta) := \bigcup_{\hat{A}_B \in \mathcal{A}^* :\, \hat{q}_B^\top \hat{A}_B^{-1} \neq 0} \left\{ z \in \mathbb{R}^s \mid c^\top x_0 + \hat{q}_B^\top \hat{A}_B^{-1}(Tx_0 + z) = \eta \right\}.$$

Then $\mathcal{Q}_{\mathrm{EE}_\eta}$ *is continuously differentiable at any* x_0 *satisfying* $c^\top x_0 \neq \eta$.

Proof. Consider the mapping $g_\eta : \mathbb{R}^n \times \mathbb{R}^s \to \overline{\mathbb{R}}$ given by

$$g_\eta(x, z) := \max\{f(x, z) - \eta, 0\},$$

which is finite and Lipschitz continuous on F by Lemma 2.1. Consider any fixed $z_0 \in \operatorname{supp} \mu_Z \setminus \big(\mathcal{N}_{x_0} \cup \mathcal{L}(x_0, \eta)\big)$. If $f(x_0, z_0) \neq \eta$, there is a neighborhood U of x_0 such that either $g_\eta(x, z_0) = f(x, z_0) - \eta$ for all $x \in U$ or $g_\eta(x, z_0) = 0$ for all $x \in U$. In both cases $g_\eta(\cdot, z_0)$ is continuously differentiable at x_0 by Theorem 4.3.

Now consider the case where $f(x_0, z_0) = \eta$. The proof of Lemma 4.2 shows that there is some $\hat{A}_B \in \mathcal{A}^*$ such that $(x_0, z_0) \in \operatorname{int} \mathcal{S}(\hat{A}_B)$. In particular, we have $c^\top x_0 + \hat{q}_B^\top \hat{A}_B^{-1}(Tx_0 + z_0) = \eta$ and $z_0 \notin \mathcal{L}(x_0, \eta)$ implies $\hat{q}_B^\top \hat{A}_B^{-1} = 0$. Thus, $\eta = c^\top x_0$, which contradicts the assumption.

Invoking Lemma A.2 and the above considerations, the differentiability of $\mathcal{Q}_{\mathrm{EE}_\eta}$ and the continuity of

$$\nabla \mathcal{Q}_{\mathrm{EE}_\eta}(x) = \sum_{\Delta \in D} \mu_Z \left[\overline{\mathcal{W}}(x, \Delta) \cap \{ z \in \operatorname{supp} \mu_Z \mid f(x, z) \geq \eta \} \right] \Delta$$

at x_0 can be shown by a straightforward extension of the arguments used in the proof of Theorem 4.3. ∎

The following simple example shows that the assumption $c^\top x_0 \neq \eta$ in Theorem 4.4 is necessary.

Example 4.5. Consider the function

$$f(x, z) = x + \min\left\{ 0 \mid y \in \operatorname{Argmin}\{0 \mid y = x + z,\ y \geq 0\} \right\}$$

and assume that Z is uniformly distributed on the interval $[1, 2]$. Then $x_0 = 0$ is an interior point of $F_Z = [-1, \infty)$, but

$$\mathcal{Q}_{\mathrm{EE}_0} = \mathbb{E}[\max\{f(x, z), 0\}] = \mathbb{E}[\max\{x, 0\}] = \max\{x, 0\}$$

is not differentiable at 0, although all conditions of Theorem 4.4 except $c^\top x_0 \neq \eta$ are met.

THEOREM 4.6. *Assume* dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}^1_s$, *and let* $x_0 \in \operatorname{int} F_Z$ *be such that* $\mathcal{Q}_{\mathbb{E}}(x_0) \neq 0$ *and* $\mu_Z[\mathcal{N}_{x_0} \cup \mathcal{L}(x_0, \mathcal{Q}_{\mathbb{E}}(x_0))] = 0$. *Then* $\mathcal{Q}_{\mathrm{SD}_\rho}$ *is continuously differentiable at* x_0 *for any* $\rho \in [0, 1)$.

Proof. Fix any $p \in [0, 1)$. By Theorem 4.3 and the definition of $\mathcal{Q}_{\mathrm{SD}_\rho}$ it is sufficient to show differentiability of the mapping $x \mapsto \mathcal{Q}_{\mathrm{EE}_{\mathcal{Q}_{\mathbb{E}}(x)}}(x)$. Consider the function $g : \mathbb{R}^n \times \mathbb{R}^s \to \overline{\mathbb{R}}$ defined by

$$g(x, z) := \max\{f(x, z) - \mathcal{Q}_{\mathbb{E}}(x), 0\},$$

which is finite and Lipschitz continuous on F by Lemma 2.1 and Theorem 3.1. Fix any $z_0 \in \operatorname{supp} \mu_Z \setminus \big(\mathcal{N}_{x_0} \cup \mathcal{L}(x_0, \mathcal{Q}_{\mathbb{E}}(x_0))\big)$ and suppose that $f(x_0, z_0) = \mathcal{Q}_{\mathbb{E}}(x_0)$. By the proof of Lemma 4.2 there is some $\hat{A}_B \in \mathcal{A}^*$ such that $(x_0, z_0) \in \operatorname{int} \mathcal{S}(\hat{A}_B)$. In particular, we have $c^\top x_0 + \hat{q}_B^\top \hat{A}_B^{-1}(Tx_0 + z_0) = \mathcal{Q}_{\mathbb{E}}(x_0)$ and $z_0 \notin \mathcal{L}(x_0, \mathcal{Q}_{\mathbb{E}}(x_0))$ implies $\hat{q}_B^\top \hat{A}_B^{-1} = 0$. Hence, $\mathcal{Q}_{\mathbb{E}}(x_0) = f(x_0, z_0) = c^\top x_0$, which contradicts the assumption.

Thus, $f(x_0, z_0) \neq \mathcal{Q}_{\mathbb{E}}(x_0)$ and there is a neighborhood U of x_0 such that either $g(x, z_0) = f(x, z_0) - \mathcal{Q}_{\mathbb{E}}(x_0)$ for all $x \in U$ or $g(x, z_0) = 0$ for all $x \in U$. In both cases $g(\cdot, z_0)$ is continuously differentiable at x_0 by Theorem 4.3.

Consequently, the differentiability of $\mathcal{Q}_{\mathrm{SD}_\rho}$ and the continuity of

$$\nabla \mathcal{Q}_{\mathrm{SD}_\rho}(x) = \nabla \mathcal{Q}_{\mathbb{E}}(x) + \rho \sum_{\Delta \in D} \mu_Z\big[\overline{\mathcal{W}}(x, \Delta) \cap \{z \in \operatorname{supp} \mu_Z \mid f(x, z) \geq \mathcal{Q}_{\mathbb{E}}(x)\}\big]\Delta$$

at x_0 can be shown by a straightforward extension of the arguments used in the proof of Theorem 4.3. ∎

While $\mu_Z[\mathcal{N}_{x_0}] = 0$ and the related conditions in Theorems 4.4 and 4.6 may be hard to verify in general, they automatically hold if μ_Z is absolutely continuous w.r.t. the Lebesgue measure.

COROLLARY 4.7. *Assume* dom $f \neq \emptyset$ *and that* $\mu_Z \in \mathcal{M}^1_s$ *is absolutely continuous with respect to the Lebesgue measure, then the following statements hold true:*

(a) $\mathcal{Q}_{\mathbb{E}}$ *is continuously differentiable at any* $x_0 \in \operatorname{int} F_Z$.

(b) *For any* $\eta \in \mathbb{R}$, $\mathcal{Q}_{\mathrm{EE}_\eta}$ *is continuously differentiable at any* $x_0 \in \operatorname{int} F_Z$ *satisfying* $c^\top x_0 \neq \eta$.

(c) *For any* $\rho \in [0, 1)$, $\mathcal{Q}_{\mathrm{SD}_\rho}$ *is continuously differentiable at any* $x_0 \in \operatorname{int} F_Z$ *satisfying* $\mathcal{Q}_{\mathbb{E}}(x_0) \neq c^\top x_0$.

Proof. $\mathcal{Q}_{\mathbb{E}}$: Since $\mathcal{N}_{x_0}$ is a finite union of affine hyperplanes, i.e., a set with Lebesgue measure zero, $\mu_Z[\mathcal{N}_{x_0}] = 0$ holds for all $x_0 \in \operatorname{int} F_Z$ and the statement is a direct consequence of Theorem 4.3.

$\mathcal{Q}_{\mathrm{EE}_\eta}$: By definition, $\mathcal{L}(x_0, \eta)$ is a finite union of affine hyperplanes, which implies $\mu_Z[\mathcal{N}_{x_0} \cup \mathcal{L}(x_0, \eta)] = 0$ for any $x_0 \in \operatorname{int} F_Z$ and Theorem 4.4 is applicable.

$\mathcal{Q}_{\mathrm{SD}_\rho}$: For any fixed x_0, $\mathcal{L}(x_0, \mathcal{Q}_{\mathbb{E}}(x_0))$ is a set of Lebesgue measure zero and the statement follows from Theorem 4.6. ∎

Remark 4.8. Corollary 4.7 can be used to justify a gradient descent approach. However, even for the expectation the explicit computation of gradients is a challenging task as it corresponds to calculating a high-dimensional integral or the probability of certain regions of strong stability (see (4.3)).

While the above results on continuity of the derivatives greatly depend on the fact that the derivatives of the respective integrands take only a finite number of values, at least the arguments used in the first part of the proof of Theorem 4.3 can be extended to certainty equivalents (cf. [22, Example 4.13]).

THEOREM 4.9. *Assume* dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}_s^\infty$, *and let* $u : \mathbb{R} \to \mathbb{R}$ *be a continuously differentiable function such that the derivative* u' *is positive on* $\mathbb{R}$. *Then* $\mathcal{Q}_{\mathrm{CE}_u}$ *is differentiable at any* $x_0 \in \operatorname{int} F_Z$ *satisfying* $\mu_Z[\mathcal{N}_{x_0}] = 0$.

Proof. We shall use Lemma A.2 in the appendix to show that $x \mapsto \mathbb{E}[u(\mathbb{F}(\cdot))]$ is differentiable at x_0. The desired differentiability of $\mathcal{Q}_{\mathrm{CE}_u}$ at x_0 then follows from the positivity of u'.

As u is differentiable, $(u \circ f)(\cdot, z_0)$ is differentiable at x_0 for any $z_0 \in \operatorname{supp} \mu_Z \setminus \mathcal{N}_{x_0}$, which verifies assumption (a) of Lemma A.2, while the first part of assumption (b) is a direct consequence of Lemma 2.4. Moreover, by $x_0 \in \operatorname{int} F_Z$ there exists a bounded neighborhood $U \subseteq \operatorname{int} F_Z$ of x_0 and the continuity of u' implies that u is Lipschitz continuous on the bounded set $f(U \times \operatorname{supp} \mu_Z)$ with some finite Lipschitz constant L_U. Fix any $x \in U \setminus \{x_0\}$ and $z_0 \in \operatorname{supp} \mu_Z \setminus \mathcal{N}_{x_0}$. By Lemma 2.1, there exists a constant L_f such that

$$\begin{aligned} &\|x - x_0\|^{-1} \big| (u \circ f)(x, z_0) - (u \circ f)(x_0, z_0) - \nabla_x (u \circ f)(x_0, z_0)(x - x_0) \big| \\ &\leq L_u L_f + \max_{z \in \operatorname{supp} \mu_Z} |(u' \circ f)(x_0, z)| \cdot \max_{\hat{A}_B \in \mathcal{A}^*} \|\hat{q}_B^\top \hat{A}_B^{-1} B\|, \end{aligned}$$

which also verifies the second part of assumption (b) of Lemma A.2 and thus completes the proof. ◻

As the entropic risk measure is a special certainty equivalent, we obtain the following result.

COROLLARY 4.10. *Assume* dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}_s^\infty$, *and fix any* $\alpha > 0$. *Then* $\mathcal{Q}_{\mathrm{Entr}_\alpha}$ *is differentiable at any* $x_0 \in \operatorname{int} F_Z$ *satisfying* $\mu_Z[\mathcal{N}_{x_0}] = 0$.

Proof. The function $u(\kappa) = \exp(\alpha\kappa)$ is continuously differentiable with positive derivate and we have $\mathcal{Q}_{\mathrm{Entr}_\alpha} = \mathcal{Q}_{\mathrm{CE}_u}$. The result thus follows from Theorem 4.9. ◻

If expected disutility functionals of the form $D_u : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, $D_u[Y] := \mathbb{E}[u(Y)]$ with some function $u : \mathbb{R} \to \mathbb{R}$ are considered, the assumptions of Theorem 4.9 can be simplified.

COROLLARY 4.11. *Assume* dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}_s^\infty$, *and let* $u : \mathbb{R} \to \mathbb{R}$ *be a continuously differentiable function. Then* $\mathcal{Q}_{\mathrm{D}_u}$ *is differentiable at any* $x_0 \in \operatorname{int} F_Z$ *satisfying* $\mu_Z[\mathcal{N}_{x_0}] = 0$.

Proof. This is a direct consequence of the arguments detailed in the proof of Theorem 4.9. ◻

The previous results give sufficient conditions for differentiability of the objective function of problem (2.2). In the presence of differentiability, necessary optimality can be formulated in terms of directional derivatives.

PROPOSITION 4.12. *Fix any* $\mathcal{R} : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ *and assume* dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}_s^p$, *and* $X \subseteq F_Z$. *Furthermore, let* $x_0 \in X$ *be a local minimizer of problem* (2.2) *and assume that* $\mathcal{Q}_\mathcal{R}$ *is differentiable at* x_0. *Then*

$$\nabla \mathcal{Q}_\mathcal{R}(x_0) d \geq 0 \tag{4.6}$$

holds for any feasible direction

$$d \in \mathcal{D}(x_0, X) := \{d \in \mathbb{R}^n \mid \exists \varepsilon_0 > 0 : \ x_0 + \varepsilon d \in X \ \forall \varepsilon \in [0, \varepsilon_0]\}.$$

Proof. dom $f \neq \emptyset$, $\mu_Z \in \mathcal{M}_s^p$, and $X \subseteq F_Z$ imply that $\mathcal{Q}_\mathcal{R}$ is real-valued on X by Lemma 2.4. For a proof of the necessity of (4.6) we refer to [6, Proposition 2.1.2]. ◻

COROLLARY 4.13. *Assume* dom $f \neq \emptyset$*,* $X \subseteq F_Z$*, and let* $\mu_Z \in \mathcal{M}_s^1$ *be absolutely continuous w.r.t. the Lebesgue measure. Furthermore, assume*

$$-c^\top \notin \text{conv } D = \text{conv } \{\hat{q}_B^\top \hat{A}_B^{-1} T \mid \hat{A}_B \in \mathcal{A}^*\},$$

then any local minimizer of

$$\min_x \{\mathcal{Q}_{\mathbb{E}}(x) \mid x \in X\} \tag{4.7}$$

is an element of $X \setminus \text{int } X$*.*

Proof. Suppose that $x_0 \in \text{int } X$ is a local minimizer of (2.2), then Corollary 4.7 and Proposition 4.12 yield $0 = \nabla \mathcal{Q}_{\mathbb{E}}(x_0)$ as $\mathcal{D}(x_0, X) = \mathbb{R}^n$. Invoking the proof of Theorem 4.3 we have

$$0 = \nabla \mathcal{Q}_{\mathbb{E}}(x_0) = c^\top + \sum_{\Delta \in D} \mu_Z[\mathcal{W}(x_0, \Delta)]\Delta$$

and thus $-c^\top \in \text{conv } D$, which contradicts the assumptions. ∎

5. A stability result for bilevel stochastic linear problems. The aim of this section is to establish a qualitative stability result for the bilevel stochastic linear problem (2.2) w.r.t. perturbations of the underlying probability measure. Taking into account that the support of the perturbed measure may differ from the original support, we shall assume dom $f \neq \emptyset$ and $F = \mathbb{R}^n \times \mathbb{R}^s$ to ensure that the objective function of (2.2) remains well defined. In two-stage stochastic programming, a similar assumption is called complete recourse.

Remark 5.1. Sufficient conditions for dom $f \neq \emptyset$ and $F = \mathbb{R}^n \times \mathbb{R}^s$ are given in [33, Corollary 1] and [33, Corollary 2]. Note that $F = \mathbb{R}^n \times \mathbb{R}^s$ holds if and only if there is some $y \in \mathbb{R}^m$ such that $Ay < 0$. By Gordan's theorem (see [28]), the latter holds iff $u = 0$ is the only nonnegative solution to $A^\top u = 0$. Under this condition, the feasible set of the lower level is full dimensional for any leader's decision x and any parameter z.

Following the lines of [12] and [13], we shall assume that $\mathcal{R} : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ with $p \in [1, \infty)$ is law-invariant, convex, and nondecreasing.

For the sake of notational simplicity, we assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless (cf. Remark 5.2 below). Then for any $x \in X$ and $\mu \in \mathcal{M}_s^p$, Lemma 2.1 implies $(\delta_x \otimes \mu) \circ f^{-1} \in \mathcal{M}_1^p$, where $\delta_x \in \mathcal{P}(\mathbb{R}^n)$ denotes the Dirac measure at x, and the atomlessness ensures that the existence of some $Y_{(x,\mu)} \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P} \circ Y_{(x,\mu)}^{-1} = (\delta_x \otimes \mu) \circ f^{-1}$. Thus, we may consider the mapping $\mathcal{Q}_{\mathcal{R}} : X \times \mathcal{M}_s^p \to \mathbb{R}$,

$$\mathcal{Q}_{\mathcal{R}}(x, \mu) := \mathcal{R}[Y_{(x,\mu)}].$$

Note that the specific choice of $Y_{(x,\mu)}$ does not matter due to the law-invariance of $\mathcal{R}$.

Remark 5.2. The assumption that $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless does not entail a loss of generality: We may just fix an arbitrary atomless probability space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathcal{P}})$, consider a law-invariant, convex, nondecreasing mapping $\overline{\mathcal{R}} : L^p(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathcal{P}}) \to \mathbb{R}$, and define $\mathcal{R} : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ via $\mathcal{R}[Y] = \overline{\mathcal{R}}[\overline{Y}]$, where $\overline{Y}$ is an arbitrary random variable in $L^p(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathcal{P}})$ satisfying $\overline{\mathbb{P}} \circ \overline{Y}^{-1} = \mathbb{P} \circ Y^{-1}$.

Consider the parametric optimization problem

$$\min_x \{\mathcal{Q}(x, \mu) \mid x \in X\}. \tag{P$_\mu$}$$

As (P_μ) may be nonconvex, we shall pay special attention to sets of local optimal solutions. For any open set $V \subseteq \mathbb{R}^n$ we introduce the optimal value function $\varphi_V : \mathcal{M}^p_s \to \overline{\mathbb{R}}$,

$$\varphi_V(\mu) := \min_x \{\mathcal{Q}(x, \mu) \mid x \in X \cap \mathrm{cl}\, V\},$$

as well as the localized optimal solution set mapping $\phi_V : \mathcal{M}^p_s \rightrightarrows \mathbb{R}^n$,

$$\phi_V(\mu) := \operatorname*{Argmin}_x \{\mathcal{Q}(x, \mu) \mid x \in X \cap \mathrm{cl}\, V\}.$$

It is well known that additional assumptions are needed when studying stability of local solutions.

Definition 5.3. *Given $\mu \in \mathcal{M}^p_s$ and an open set $V \subseteq \mathbb{R}^n$, $\phi_V(\mu)$ is called a* complete local minimizing (CLM) set *of (P_μ) w.r.t. V if $\emptyset \neq \phi_V(\mu) \subseteq V$.*

Remark 5.4. The set of global optimal solutions $\phi_{\mathbb{R}^n}(\mu)$ and any set of isolated minimizers are CLM sets. However, in general, sets of strict local minimizers may fail to be CLM sets (cf. [54]).

In the following, we shall equip $\mathcal{P}(\mathbb{R}^s)$ with the topology of weak convergence, i.e., the topology where a sequence $\{\mu_l\}_{l\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^s)$ converges weakly to $\mu \in \mathcal{P}(\mathbb{R}^s)$, written $\mu_l \xrightarrow{w} \mu$, iff

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

holds for any bounded continuous function $h : \mathbb{R}^s \to \mathbb{R}$ (cf. [7]). The example below shows that even $\varphi_{\mathbb{R}^n}$ may fail to be weakly continuous on the entire space $\mathcal{P}(\mathbb{R}^s)$.

Example 5.5. The problem

$$\min_x \left\{ x + \int_{\mathbb{R}^s} z\ \mu(dz) \mid 0 \leq x \leq 1 \right\}$$

arises from a bilevel stochastic linear problem, where $\Psi(x, z) = \{z\}$ holds for any (x, z). Assume that $\mu = \mathbb{P} \circ Z^{-1} = \delta_0$ is the Dirac measure at 0. Then the above problem can be rewritten as $\min_x\{x \mid 0 \leq x \leq 1\}$ and its optimal value is 0.

However, while the sequence $\mu_l := (1-\frac{1}{l})\delta_0 + \frac{1}{l}\delta_l$ converges weakly to δ_0, replacing μ with μ_l yields the problem

$$\min_x \{x + 1 \mid 0 \leq x \leq 1\},$$

whose optimal value is equal to 1 for any $l \in \mathbb{N}$.

In the present work, we shall follow the approach of [13] and confine the stability analysis to locally uniformly $\|\cdot\|^p$-integrating sets.

Definition 5.6. *A set $\mathcal{M} \subseteq \mathcal{M}^p_s$ is said to be* locally uniformly $\|\cdot\|^p$-integrating *for some $p \in [1, \infty)$ iff for any $\epsilon > 0$ there exists some open neighborhood $\mathcal{N}$ of μ w.r.t. the topology of weak convergence such that*

$$\lim_{a\to\infty,\ a\geq 0}\ \sup_{\nu\in\mathcal{M}\cap\mathcal{N}} \int_{\mathbb{R}^s\setminus a\mathbb{B}} \|z\|^p\ \nu(dz) \leq \epsilon.$$

A detailed discussion of locally uniformly $\|\cdot\|^p$-integrating sets is provided in [22], [41], [42], and [43]. The following examples demonstrate the relevance of the concept.

Example 5.7. (a) Fix finite constants $\kappa, \epsilon > 0$ and $p \in [1, \infty)$. Then by [22, Corollary A.47, (c)], the set

$$\mathcal{M}(\kappa, \epsilon) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^s) \mid \int_{\mathbb{R}^s} \|z\|^{p+\epsilon} \; \mu(dz) \leq \kappa \right\}$$

of Borel probability measures with uniformly bounded moments of order $p+\epsilon$ is locally uniformly $\|\cdot\|^p$-integrating. Note that $\mathcal{M}(\infty, \epsilon)$ coincides with $\mathcal{P}(\mathbb{R}^s)$, which is not locally uniformly $\|\cdot\|$-integrating (and thus not locally uniformly $\|\cdot\|^p$-integrating for any $p \in [1, \infty)$) by Example 5.5.

(b) Fix any compact set $\Xi \subset \mathbb{R}^s$. By [22, Corollary A.47, (b)], the set

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

of Borel probability measures whose support is contained in Ξ is locally uniformly $\|\cdot\|^p$-integrating for any $p \in [1, \infty)$.

(c) Any singleton $\{\mu\} \subset \mathcal{M}^p_s$ is locally uniformly $\|\cdot\|^p$-integrating for any $p \in [1, \infty)$ by [42, Lemma 5.2].

THEOREM 5.8. *Assume* dom $f \neq \emptyset$ *and* $F = \mathbb{R}^n \times \mathbb{R}^s$. *Let* $\mathcal{M} \subseteq \mathcal{M}^p_s$ *be locally uniformly* $\|\cdot\|^p$*-integrating, then*

(a) $\mathcal{Q}_{\mathcal{R}}|_{X \times \mathcal{M}}$ *is real-valued and weakly continuous.*

(b) $\varphi_{\mathbb{R}^n}|_{\mathcal{M}}$ *is weakly upper semicontinuous.*

In addition, assume that $\mu_0 \in \mathcal{M}$ *is such that* $\phi_V(\mu_0)$ *is a CLM set of* P_{μ_0} *w.r.t. some open bounded set* $V \subset \mathbb{R}^n$. *Then the following statements hold true:*

(c) $\varphi_V|_{\mathcal{M}}$ *is weakly continuous at* μ_0.

(d) $\phi_V|_{\mathcal{M}}$ *is weakly upper semicontinuous at* μ_0 *in the sense of Berge (cf.* [5], [56, Chapter 5]*), i.e., for any open set* $\mathcal{O} \subseteq \mathbb{R}^n$ *with* $\phi|_V(\mu_0) \subseteq \mathcal{O}$ *there exists a weakly open neighborhood* $\mathcal{N}$ *of* μ_0 *such that* $\phi_V(\mu) \subseteq \mathcal{O}$ *for all* $\mu \in \mathcal{N} \cap \mathcal{M}$.

(e) *There exists some weakly open neighborhood* $\mathcal{U}$ *of* μ_0 *such that* $\phi_V(\mu)$ *is a CLM set for* (P_μ) *w.r.t.* V *for any* $\mu \in \mathcal{U} \cap \mathcal{M}$.

Proof. Fix any $x_0 \in X$. By Lemma 2.1, f is Lipschitz continuous on $X \times \mathbb{R}^s$. Thus, there exists a constant $L > 0$ such that

$$|f(x, z)| \leq L\|z\| + L\|x - x_0\| + |f(x_0, 0)|$$

hold for any $(x, z) \in X \times \mathbb{R}^s$, i.e., the growth of f is bounded linearly, and the result follows from [13, Corollary 2.4.]. ∎

Remark 5.9. Under the assumptions of Theorem 5.8 (d), any accumulation point x of a sequence local optimal solutions $x_l \in \phi_V(\mu_l)$ as $\mu_l \xrightarrow{w} \mu$, $\mu_l \in \mathcal{M}$, is a local optimal solution of (P_μ). In the broader setting of general stochastic programs, a similar qualitative stability result under perturbations w.r.t. the minimal information pseudometric has been established in [58, Theorem 5]. If $\mathcal{R}$ admits a representation as $\mathcal{R}(Y) = \mathbb{E}[h(Y)]$ for some Lipschitzian mapping h, Lemma 2.1 and [58, Theorem 5] can be used to obtain an alternate proof of Theorem 5.8 (d) by bounding the minimal information distance from above using a distance with ζ-structure that metrizes the topology of weak convergences (cf. [58, page 20]).

Remark 5.10. All results of sections 2, 3, 4, and 5 can be easily extended to the pessimistic approach to bilevel stochastic linear optimization, where f takes the form

$$f(x,z) = c^\top x - \min_y \{-q^\top y \mid y \in \Psi(x,z)\}.$$

Remark 5.11. Theorem 5.8 justifies approaches where the true underlying measure is approximated by a weakly converging sequence of finite discrete ones. As the finite discrete version of a stochastic program is usually easier to solve, many such approximation techniques have been studied in stochastic optimization literature: For instance, it is well known that approximation schemes based on discretization via empirical estimation (see [53], [63]) or conditional expectations (see [9], [36]) produce weakly converging sequences of discrete probability measures under mild assumptions. A more detailed discussion of popular discretization techniques based on sample average approximation (SAA) and their enhancements by quasi-Monte Carlo methods is provided in [61, Chapter 5]. A SAA scheme for bilevel stochastic linear programs with uncertain lower level objective function has been studied in [34, section 3], where convergence of global minimizers for the value-at-risk of the leaders's objective function value is established. In the broader framework of two-stage SMPECs, convergence results for local optimal solutions under SAA are given in [62].

6. Finite discrete distributions. Throughout this section, we shall assume that the underlying random vector Z is discrete with a finite number of realizations $Z_1, \dots, Z_K \in \mathbb{R}^s$ and respective probabilities $\pi_1, \dots, \pi_K \in (0,1]$. Let I denote the index set $\{1, \dots, K\}$, then F_Z takes the form

$$F_Z = \{x \in \mathbb{R}^n \mid \forall k \in I \; \exists y \in \mathbb{R}^m : \; Ay \le Tx + Z_k\}.$$

Suppose that $x_0 \in X$ is such that $\{y \in \mathbb{R}^m \mid Ay \le Tx_0 + Z_k\} = \emptyset$ holds for some $k \in I$. Then the probability of $f(x_0, Z(\omega)) = \infty$ is at least $\pi_k > 0$, i.e., x_0 should be considered as infeasible for problem (2.2). Consequently, $X \subseteq F_Z$ can be understood as an induced constraint. Note that $X \cap F_Z$ is a polyhedron if X is a polyhedron.

In the very first paper on bilevel stochastic programming it is shown that in the risk neutral setting, problem (2.2) reduces to a deterministic bilevel linear program, which allows to employ standard bilevel programming techniques (cf. [50]).

THEOREM 6.1 (expectation). *Assume* dom $f \neq \emptyset$ *and let* $X \subseteq F_Z$ *be a polyhedron, then the risk neutral bilevel stochastic linear problem*

$$\min_x \{\mathcal{Q}_{\mathbb{E}}(x) \mid x \in X\}$$

is equivalent to the optimistic bilevel linear program

$$\min_x \left\{ c^\top x + \min_{y_1,\dots,y_K} \left\{ \sum_{k \in I} \pi_k q^\top y_k \mid (y_1, \dots, y_K) \in \Psi_{\mathbb{E}}(x) \right\} \middle| \; x \in X \right\}, \tag{6.1}$$

where $\Psi_{\mathbb{E}} : \mathbb{R}^n \rightrightarrows \mathbb{R}^{Km}$ *is given by*

$$\Psi_{\mathbb{E}}(x) := \operatorname*{Argmin}_{y_1,\dots,y_K} \left\{ \sum_{k \in I} d^\top y_k \mid Ay_k \le Tx + Z_k \; \forall k \in I \right\}.$$

Proof. We have

$$\begin{aligned}\mathcal{Q}_{\mathbb{E}}(x) &= \sum_{k\in I} \pi_k f(x, Z_k)\\ &= c^\top x + \sum_{k\in I} \pi_k \min_{y_k}\{q^\top y_k \mid y_k \in \Psi(x, Z_k)\}\\ &= c^\top x + \min_{y_1,\dots,y_k} \left\{\sum_{k\in I} \pi_k q^\top y_k \mid y_k \in \Psi(x, Z_k)\ \forall k \in I\right\}\end{aligned}$$

and the result follows from $\Psi_{\mathbb{E}}(x) = \Psi(x, Z_1) \times \cdots \times \Psi(x, Z_K)$. ∎

Remark 6.2. The proof of Theorem 6.1 shows that the inner minimization problem in (6.1) can be decomposed into K problems of similar structure.

We shall show that a similar reformulation as a standard bilevel linear program is possible for the expected excess and the mean upper semideviation.

THEOREM 6.3 (expected excess). *Assume* dom $f \neq \emptyset$ *and let* $X \subseteq F_Z$ *be a polyhedron, then for any* $\eta \in \mathbb{R}$*, the risk-averse bilevel stochastic linear problem*

$$\min_x \left\{\mathcal{Q}_{EE_\eta}(x) \mid x \in X\right\}$$

is equivalent to the optimistic bilevel linear program

$$\min_x \left\{\min_{\substack{y_1,\dots,y_K,\\ v_1,\dots,v_K}} \left\{\sum_{k\in I} \pi_k v_k \,|\, (y_1,\dots,y_K, v_1,\dots,v_K) \in \Psi_{\mathrm{EE}_\eta}(x)\right\} \;\middle|\; x \in X\right\}, \tag{6.2}$$

where $\Psi_{\mathrm{EE}_\eta} : \mathbb{R}^n \rightrightarrows \mathbb{R}^{Km+K}$ *is given by*

$$\Psi_{\mathrm{EE}_\eta}(x) := \operatorname*{Argmin}_{\substack{y_1,\dots,y_K,\\ v_1,\dots,v_K}} \left\{\sum_{k\in I} d^\top y_k \,|\, Ay_k \leq Tx + Z_k,\ v_k \geq 0,\ v_k \geq c^\top x + q^\top y_k - \eta\, \forall k \in I\right\}.$$

Proof. We have $\mathcal{Q}_{\mathrm{EE}_\eta}(x) = \sum_{k\in I} \pi_k g_k(x)$, where

$$\begin{aligned}g_k(x) &= \max\{0, f(x, Z_k) - \eta\}\\ &= \min_{v_k} \{v_k \mid v_k \geq 0,\ v_k \geq f(x, Z_k) - \eta\}\\ &= \min_{v_k} \left\{v_k \;\middle|\; v_k \geq 0,\ v_k \geq c^\top x + \min_{y_k}\{q^\top y_k \mid y_k \in \Psi(x, Z_k)\} - \eta\right\}\\ &= \min_{y_k, v_k} \left\{v_k \mid v_k \geq 0,\ v_k \geq c^\top x + q^\top y_k - \eta,\ y_k \in \Psi(x, Z_k)\right\}\end{aligned}$$

holds for any $(x, k) \in X \times I$. Thus,

$$\begin{aligned}\mathcal{Q}_{\mathrm{EE}_\eta}(x) &= \sum_{k=1}^{K} \pi_k \inf_{y_k, v_k} \left\{v_k \mid v_k \geq 0,\ v_k \geq c^\top x + q^\top y_k - \eta,\ y_k \in \Psi(x, Z_k)\right\}\\ &= \inf_{\substack{y_1,\dots,y_K,\\ v_1,\dots,v_K}} \left\{\sum_{k=1}^{K} \pi_k v_k \mid v_k \geq 0,\ v_k \geq c^\top x + q^\top y_k - \eta,\ y_k \in \Psi(x, Z_k)\ \forall k \in I\right\}\\ &= \inf_{\substack{y_1,\dots,y_K,\\ v_1,\dots,v_K}} \left\{\sum_{k=1}^{K} \pi_k v_k \mid (y_1,\dots,y_K, v_1,\dots,v_K) \in \Psi_{\mathrm{EE}_\eta}(x)\right\},\end{aligned}$$

which completes the proof. ∎

Remark 6.4. Let $\Psi_{\mathrm{EE}_{\eta,k}} : X \rightrightarrows \mathbb{R}^{m+1}$ be given by

$$\Psi_{\mathrm{EE}_{\eta,k}}(x) := \operatorname*{Argmin}_{y_k, v_k} \left\{ d^\top y_k \mid Ay_k \leq Tx + Z_k,\ v_k \geq 0,\ v_k \geq c^\top x + q^\top y_k - \eta \right\},$$

then $\Psi_{\mathrm{EE}_\eta}(x)$ admits the representation

$$\Psi_{\mathrm{EE}_\eta}(x) = \{(y_1, \dots, y_K, v_1, \dots, v_K) \mid (y_k, v_k) \in \Psi_{\mathrm{EE}_{\eta,k}}(x)\ \forall k \in I\}.$$

Thus, the inner minimization problem in (6.2) decomposes into K problems of similar structure.

THEOREM 6.5 (mean upper semideviation). *Assume* $\operatorname{dom} f \neq \emptyset$ *and let* $X \subseteq F_Z$ *be a polyhedron, then for any* $\rho \in [0, 1)$, *the risk-averse bilevel stochastic linear problem*

$$\min_x \left\{ \mathcal{Q}_{\mathrm{SD}_\rho}(x) \mid x \in X \right\}$$

is equivalent to the optimistic bilevel linear program

$$\min_x \left\{ c^\top x + \min_{\substack{y_1,\dots,y_K,\\ v_1,\dots,v_K}} \left\{ \begin{array}{c} (1-\rho) \sum_{k \in I} \pi_k q^\top y_k + \rho \sum_{k \in I} \pi_k v_k \\ s.t.\ (y_1, \dots, y_K, v_1, \dots, v_K) \in \Psi_{\mathrm{SD}_\rho}(x) \end{array} \right\} \mid x \in X \right\}, \tag{6.3}$$

where $\Psi_{\mathrm{SD}_\rho} : \mathbb{R}^n \rightrightarrows \mathbb{R}^{Km+K}$ *is given by*

$$\Psi_{\mathrm{SD}_\rho}(x) := \operatorname*{Argmin}_{\substack{y_1,\dots,y_K,\\ v_1,\dots,v_K}} \left\{ \sum_{k \in I} d^\top y_k \mid Ay_k \leq Tx + Z_k,\ v_k \geq q^\top y_k,\ v_k \geq \sum_{j \in I} \pi_j q^\top y_j\ \forall k \in I \right\}.$$

Proof. By

$$\begin{aligned}
\mathcal{Q}_{\mathrm{SD}_\rho}(x) &= \mathcal{Q}_{\mathbb{E}}(x) + \rho \mathcal{Q}_{\mathrm{EE}_{\mathcal{Q}_{\mathbb{E}}(x)}}(x) \\
&= \mathcal{Q}_{\mathbb{E}}(x) + \rho \sum_{k \in I} \pi_k \max\{0, f(x, Z_k) - \mathcal{Q}_{\mathbb{E}}(x)\} \\
&= (1-\rho)\mathcal{Q}_{\mathbb{E}}(x) + \rho \sum_{k \in I} \pi_k \max\{f(x, Z_k), \mathcal{Q}_{\mathbb{E}}(x)\} \\
&= (1-\rho)\mathcal{Q}_{\mathbb{E}}(x) + \rho c^\top x + \rho \sum_{k \in I} \pi_k \min_{v_k} \{v_k - c^\top x | v_k \geq f(x, Z_k), v_k \geq \mathcal{Q}_{\mathbb{E}}(x)\}
\end{aligned}$$

and the representation of $\mathcal{Q}_{\mathbb{E}}$ that was established in the proof of Theorem 6.1, we have

$$\begin{aligned}
\mathcal{Q}_{\mathrm{SD}_\rho}(x) &= c^\top x + (1-p) \sum_{k \in I} \pi_k \min_{y_k} \left\{ q^\top y_k \mid y_k \in \Psi(x, Z_k)\ \forall k \in I \right\} \\
&\quad + \min_{v_1,\dots,v_K} \left\{ \rho \sum_{k \in I} \pi_k v_k \;\middle|\; \begin{array}{l} v_k \geq \min_{y_k} \{q^\top y_k \mid y_k \in \Psi(x, Z_k)\}, \\ v_k \geq \sum_{k \in I} \pi_k \min_{y_k} \left\{ q^\top y_k \mid y_k \in \Psi(x, Z_k) \right\} \end{array} \right\} \\
&= c^\top x + \min_{\substack{y_1,\dots,y_K,\\ v_1,\dots,v_K}} \left\{ (1-\rho) \sum_{k \in I} \pi_k q^\top y_k + \rho \sum_{k \in I} \pi_k v_k \;\middle|\; \begin{array}{ll} y_k \in \Psi(x, Z_k),\ v_k \geq q^\top y_k & \forall k \in I \\ v_k \geq \sum_{j \in I} \pi_j q^\top y_j & \forall k \in I \end{array} \right\},
\end{aligned}$$

which completes the proof. ∎

Remark 6.6. The inner minimization problem in (6.3) does not decompose scenariowise due to the K coupling constraints $v_k \geq \sum_{j \in I} \pi_j q^\top y_j$ for $k \in I$ in the description of $\mathcal{Q}_{\mathrm{SD}_\rho}(x)$.

Finally, we shall consider models involving the conditional value-at-risk.

THEOREM 6.7 (conditional value-at-risk). *Assume* dom $f \neq \emptyset$ *and let* $X \subseteq F_Z$ *be a polyhedron, then for any* $\alpha \in (0,1)$*, the risk-averse bilevel stochastic linear problem*

$$\min_x \{\mathcal{Q}_{\mathrm{CVaR}_\alpha}(x) \mid x \in X\}$$

is equivalent to

$$\min_x \left\{ \min_{\eta \in \mathbb{R}} \left\{ \eta + \min_{\substack{y_1,\ldots,y_K,\\ v_1,\ldots,v_K}} \left\{ \begin{array}{c} \dfrac{1}{1-\alpha} \displaystyle\sum_{k \in I} \pi_k v_k \\ s.t.\ (y_1,\ldots,y_K,v_1,\ldots,v_K) \in \Psi_{\mathrm{EE}_\eta}(x) \end{array} \right\} \right\} \;\middle|\; x \in X \right\}. \tag{6.4}$$

Proof. As

$$\mathcal{Q}_{\mathrm{CVaR}_\alpha}(x) = \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{1-\alpha} \mathcal{Q}_{\mathrm{EE}_\eta}(x) \right\},$$

the result follows directly from the representation of $\mathcal{Q}_{\mathrm{EE}_\eta}(x)$ that was established in the proof Theorem 6.3. ◻

Remark 6.8. Every evaluation of the objective function in (6.4) corresponds to solving a bilevel linear problem with scalar upper level variable η.

7. A regularization scheme for bilevel linear problems. In the setting of Theorems 6.1, 6.3 and 6.5, the risk-averse bilevel stochastic linear problem may be reformulated as a standard optimistic bilevel linear problem of the form

$$\min_u \{g^\top u + \min_w \{h^\top w \mid w \in \Psi(u)\} \mid u \in U\}, \tag{7.1}$$

where $\Psi : \mathbb{R}^k \rightrightarrows \mathbb{R}^l$ is given by

$$\Psi(u) = \operatorname*{Argmin}_w \{t^\top w \mid Ww \leq Bu + b\}$$

for vectors $g \in \mathbb{R}^k$, $h, t \in \mathbb{R}^l$, and $b \in \mathbb{R}^r$, matrices $W \in \mathbb{R}^{r \times l}$ and $B \in \mathbb{R}^{r \times k}$, and a nonempty polyhedron $U \subseteq \mathbb{R}^k$.

We shall discuss a solution approach for (7.1) that relies on replacing it with a regularized single level problem involving the Karush–Kuhn–Tucker (KKT) conditions of the lower level problem.

THEOREM 7.1 (cf. [30, Theorem 3.7], [46]). *Assume that* $\operatorname{Argmin}_w\{h^\top w \mid w \in \Psi(u)\}$ *is nonempty for any* $u \in U$*. Then the following statements hold true:*

(a) *The optimal values of* (7.1) *and*

$$\min_{u,w} \{g^\top u + h^\top w \mid u \in U,\ w \in \Psi(u)\} \tag{7.2}$$

coincide.

(b) $\overline{u}$ *is a global minimizer of* (7.1) *iff there exists some* $\overline{w}$ *such that* $(\overline{u}, \overline{w})$ *is a global minimizer of* (7.2).

(c) $\overline{u}$ *is a local minimizer of* (7.1) *iff there exists some* $\overline{w}$ *such that* $(\overline{u}, \overline{w})$ *is a local minimizer of* (7.2).

Proof. By assumption, the mapping $\varphi_o : U \to \mathbb{R}$

$$\varphi_o(u) := \min_w \{h^\top w \mid w \in \Psi(u)\}$$

is well defined and for any $\tilde{u} \in U$ there exists some $\tilde{w} \in \Psi(\tilde{u})$ such that $h^\top \tilde{w} = \varphi_o(\tilde{u})$. Furthermore, $\varphi_o(\tilde{u}) \le h^\top w$ holds for any $w \in \Psi(\tilde{u})$, which implies (a), (b), and the "only if" part of (c).

To show the "if" part of (c), suppose that $(\overline{u}, \overline{w})$ is a local minimizer of (7.2). Then there exist some $\epsilon > 0$ such that

$$g^\top u + h^\top w \ge g^\top \overline{u} + h^\top \overline{w} \tag{7.3}$$

holds for any $(u, w) \in B_\epsilon(\overline{u}, \overline{w})$ satisfying $u \in U$ and $w \in \Psi(u)$. In particular, we have $g^\top \overline{u} + h^\top w \ge g^\top \overline{u} + h^\top \overline{w}$ for any $w \in \Psi(\overline{u}) \cap B_\epsilon(\overline{w})$, which implies that $\overline{w}$ is a local and thus global minimizer of the linear program

$$\min_w \{h^\top w \mid w \in \Psi(\overline{u})\}.$$

Consider the mapping $M : U \rightrightarrows \mathbb{R}^l$ defined by

$$M(u) := \operatorname*{Argmin}_w \{h^\top w \mid w \in \Psi(u)\} = \{w \mid Ww \le Bu + b,\ h^\top w \le \varphi_o(u)\}.$$

As φ_o is Lipschitz continuous by Theorem A.1 in the appendix, Lipschitz continuity of M follows from the same result. Suppose that $\overline{u}$ is not a local minimizer of (7.1), then there exist a sequence $\{u_n\}_{n\in\mathbb{N}}$ such that and $u_n \in U$ and

$$g^\top u_n + \varphi_o(u_n) < g^\top \overline{u} + \varphi_o(\overline{u}) \tag{7.4}$$

hold for any $n \in \mathbb{N}$ and we have $\lim_{n\to\infty} u_n = u$. The Lipschitz continuity of M and $\overline{w} \in M(\overline{u})$ imply

$$\lim_{n\to\infty} \inf_{w\in M(u_n)} \|w - \overline{w}\| = 0.$$

Thus, there exists a sequence $\{w_n\}_{n\in\mathbb{N}}$ satisfying $\lim_{n\to\infty} w_n = \overline{w}$ and $w_n \in M(u_n)$ for all $n \in \mathbb{N}$. Consequently, by (7.4), there is some $N \in \mathbb{N}$ such that for any $n \ge N$, we have $(u_n, w_n) \in B_\epsilon(\overline{u}, \overline{w})$ and

$$g^\top u_n + h^\top w_n = g^\top u_n + \varphi_o(u_n) < g^\top \overline{u} + \varphi_o(\overline{u}) = g^\top \overline{u} + h^\top \overline{w},$$

which contradicts (7.3). Thus, $\overline{u}$ is a local minimizer of (7.1). ∎

Next, we use the KKT conditions of the lower level problem to replace (7.2) with the single-level problem

$$\min_{u,w,v} \left\{ g^\top u + h^\top w \;\middle|\; \begin{array}{l} Ww \le Bu + b,\ W^\top v = t,\ v \le 0, \\ v^\top (Ww - Bu - b) = 0,\ u \in U \end{array} \right\}. \tag{7.5}$$

The relationship between bilevel problems and mathematical programs with complementarity constraints arising from the lower level KKT system has been investigated in [16]. In the special case of bilevel linear problems, the following holds.

THEOREM 7.2 (cf. [16, Theorem 3.2]).

(a) *The optimal values of* (7.2) *and* (7.5) *coincide*

(b) $(\overline{u},\overline{w})$ *is a global minimizer of* (7.2) *iff there exists some* $\overline{v}$ *such that* $(\overline{u},\overline{w},\overline{v})$ *is a global minimizer of* (7.5).

(c) $(\overline{u},\overline{w})$ *is a local minimizer of* (7.2) *iff* $(\overline{u},\overline{w},\overline{v})$ *is a local minimizer of* (7.5) *for any* $\overline{v} \leq 0$ *satisfying* $W^\top \overline{v} = t$ *and* $\overline{v}^\top (W\overline{w} - B\overline{u} - b) = 0$.

Proof. As the lower level problem is linear, its KKT conditions are necessary and sufficient for optimality. Thus, we have $w \in \Psi(u)$ iff there exists some $v \leq 0$ such that $W^\top v = t$ and $v^\top (Ww - Bu - b) = 0$, which implies (a), (b) and the "only if" part of (c).

To show the "if" part of (c), let $(\overline{u},\overline{w},\overline{v})$ be a local minimizer of (7.5) for any $\overline{v} \leq 0$ satisfying $W^\top \overline{v} = t$ and $\overline{v}^\top (W\overline{w} - B\overline{u} - b) = 0$ and suppose that $(\overline{u},\overline{w})$ is not a local minimizer of (7.2). Then there exist sequences $\{u_n\}_{n\in\mathbb{N}} \subseteq U$ and $\{w_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} u_n = \overline{u}$, $\lim_{n\to\infty} w_n = \overline{w}$, and for any $n \in \mathbb{N}$ we have $w_n \in \Psi(u_n)$ and

$$g^\top u_n + h^\top w_n < g^\top \overline{u} + h^\top \overline{w}. \tag{7.6}$$

As the mapping $\Lambda : \mathrm{gph}\,\Psi \rightrightarrows \mathbb{R}^r$ given by

$$\Lambda(u, w) := \{v \in \mathbb{R}^r \mid W^\top v = t,\ v \leq 0,\ v^\top (Ww - Bu - b) = 0\} \tag{7.7}$$

is outer semicontinuous by Lemma A.3 in the appendix, there exists some $N \in \mathbb{N}$ such that

$$\Lambda(u_n, w_n) \subseteq \Lambda(\overline{u},\overline{w}) \tag{7.8}$$

holds for all $n \geq N$. Fix any converging sequence $\{v_n\}_{n\in\mathbb{N}}$ such that $v_n \in \Lambda(u_n, w_n)$ holds for any $n \in \mathbb{N}$. By (7.8) we have $\overline{v} = \lim_{n\to\infty} v_n \in \Lambda(\overline{u},\overline{w})$. Thus, $(\overline{u},\overline{w},\overline{v})$ is a local minimizer of (7.5). In particular, there exists some $\overline{N} \in \mathbb{N}$ such that $g^\top u_n + h^\top w_n \geq g^\top \overline{u} + h^\top \overline{w}$ for all $n \geq \overline{N}$, which contradicts (7.6). ∎

A logical Benders decomposition algorithm for (7.5) is proposed in [31]. However, it is known that Mangasarian–Fromovitz constraint qualification or Slater's constraint qualification are violated at every feasible point of (7.5) (cf. [59]). To overcome the difficulties related with this property, we propose replacing (7.5) by

$$\mathrm{P}(\varepsilon) \qquad \min_{u,w,v} \left\{ g^\top u + h^\top w \;\middle|\; \begin{array}{l} Ww \leq Bu + b,\ W^\top v = t,\ v \leq 0, \\ v^\top (Ww - Bu - b) \leq \varepsilon,\ u \in U \end{array} \right\}$$

and solve this problem for $\varepsilon \downarrow 0$. This approach and its use to solve general mathematical programs with equilibrium constraints (MPECs) has been investigated in [60]. For the special case of the bilevel linear optimization problem (7.2) we can prove the following result.

THEOREM 7.3. *Here, we let* $(\overline{u},\overline{w},\overline{v})$ *be an accumulation point of a sequence* $\{(u_n, w_n, v_n)\}_{n\in\mathbb{N}}$ *of local minimizers of problem* $\mathrm{P}(\varepsilon_n)$ *for* $\varepsilon_n \downarrow 0$. *Then* $(\overline{u},\overline{w})$ *is a local minimizer of* (7.2).

Proof. Without loss of generality, we may assume that $\{(u_n, w_n, v_n)\}_{n\in\mathbb{N}}$ converges. Suppose that $(\overline{u},\overline{w})$ is not a local minimizer of (7.2). Then, since U is a polyhedron and $\mathrm{gph}\,\Psi$ is polyhedral (cf. [15, Theorem 3.1]), i.e., equal to the union of a finite number of polyhedra, there exists a direction $(d_u, d_w) \in \mathbb{R}^k \times \mathbb{R}^l$ and a sequence $\alpha_m \downarrow 0$ such that $\overline{u} + \alpha_m d_u \in U$, $\overline{w} + \alpha_m d_w \in \Psi(\overline{u} + \alpha_m d_u)$, and

$$g^\top (\overline{u} + \alpha_m d_u) + h^\top (\overline{w} + \alpha_m d_w) < g^\top \overline{u} + h^\top \overline{w} \tag{7.9}$$

hold for any $m \in \mathbb{N}$. As the mapping Λ defined by (7.7) is outer semicontinuous, there exists a constant $N \in \mathbb{N}$ such that $\Lambda(\overline{u} + \alpha_m d_u, \overline{w} + \alpha_m d_w) \subseteq \Lambda(\overline{u}, \overline{w})$ for any $m \geq N$. In particular, there exists some vertex $\tilde{v}$ of $\Lambda(\overline{u}, \overline{w})$ such that $\tilde{v}$ is a vertex of $\Lambda(\overline{u} + \alpha_m d_u, \overline{w} + \alpha_m d_w)$ for any $m \geq N$. We shall prove that there exists some $\overline{N} \in \mathbb{N}$ such that

$$W(w_n + \alpha_m d_w) - B(u_n + \alpha_m d_u) - b \leq 0, \quad u_n + \alpha_m d_u \in U \tag{7.10}$$

holds for any $m, n \geq \overline{N}$.

For any $i \in \{1, \ldots, r\}$ with $e_i^\top \tilde{v} < 0$ and any $m \geq N$, we have

$$e_i^\top \big(W\overline{w} - B\overline{u} - b + \alpha_m(W d_w - B d_u)\big) = e_i^\top \big(W(\overline{w} + \alpha_m d_w) - B(\overline{u} + \alpha_m d_u) - b\big) = 0.$$

As $e_i^\top (W\overline{w} - B\overline{u} - b) = 0$ and $\alpha_m > 0$, this implies $e_i^\top (W d_w - B d_u) = 0$. Furthermore, since (u_n, w_n) is feasible for $\mathrm{P}(\varepsilon_n)$, we conclude that

$$e_i^\top \big(W(w_n + \alpha_m d_w) - B(u_n + \alpha_m d_u) - b\big) \leq 0$$

for any $m, n \in \mathbb{N}$.

Similarly, for any $i \in \{1, \ldots, r\}$ such that $e_i^\top \tilde{v} = 0$ and $e_i^\top (W\overline{w} - B\overline{u} - b) = 0$, we obtain $e_i^\top (W d_w - B d_u) \leq 0$ and thus

$$e_i^\top \big(W(w_n + \alpha_m d_w) - B(u_n + \alpha_m d_u) - b\big) \leq 0$$

for any $m, n \in \mathbb{N}$.

Finally, for any $i \in \{1, \ldots, r\}$ such that $e_i^\top \tilde{v} = 0$ and $e_i^\top (W\overline{w} - B\overline{u} - b) < 0$, the existence of some $N' \in \mathbb{N}$ such that

$$e_i^\top \big(W(w_n + \alpha_m d_w) - B(u_n + \alpha_m d_u) - b\big) \leq 0$$

for any $m, n \geq N'$ is a direct consequence of the continuity of the mapping

$$(u, w, \alpha) \mapsto W(w + \alpha d_w) - B(u + \alpha d_u) - b.$$

The second part of (7.10) can be shown using similar arguments. Without loss of generality, we may assume that $U = \{u \in \mathbb{R}^k \mid Su \leq \vartheta\}$ for some $S \in \mathbb{R}^{s \times k}$ and $\vartheta \in \mathbb{R}^s$. For any $i \in \{1, \ldots, s\}$ satisfying $e_i^\top (S\bar{u}) = e_i^\top \vartheta$, $\overline{u} + \alpha_m d_u \in U$ implies $e_i^\top S d_u \leq 0$ and thus

$$e_i^\top \big(S(u_n + \alpha_m d_u)\big) \leq e_i^\top \big(S(u_n)\big) \leq e_i^\top \vartheta.$$

Moreover, for any $i \in \{1, \ldots, s\}$ with $e_i^\top (S\bar{u}) < e_i^\top \vartheta$, the existence of some $\overline{N} \geq N'$ such that

$$e_i^\top \big(S(u_n + \alpha_m d_u)\big) < e_i^\top \vartheta$$

for all $n, m \geq \overline{N}$ follows from the continuity of the mapping $(u, \alpha) \mapsto S(u + \alpha d_u)$.

By the above considerations, we have

$$\tilde{v}^\top \big(W(w_n + \alpha_m d_w) - B(u_n + \alpha_m d_u) - b\big) = \tilde{v}^\top \big(W w_n - B u_n - b\big) \tag{7.11}$$

and $\lim_{n \to \infty} \tilde{v}^\top (W w_n - B u_n - b) = 0$. Furthermore, as $\varepsilon_n \downarrow 0$, (u_n, w_n, v_n) is feasible for $\mathrm{P}(\varepsilon_{n'})$ for any $n' \geq n$. Thus, we may assume that

$$\tilde{v}^\top \big(W w_n - B u_n - b\big) \leq \frac{\varepsilon_n}{2} \tag{7.12}$$

holds for any $n \in \mathbb{N}$ without loss of generality. Equations (7.10), (7.11), and (7.12) imply that $(u_n + \alpha_m d_u, w_n + \alpha_m d_w, \tilde{v})$ is feasible for $\mathrm{P}(\varepsilon_n)$ for any $m, n \geq \overline{N}$.

Fix $n \geq \overline{N}$. We shall prove that for any $\lambda \in (0,1]$, there is some $M_\lambda \geq \overline{N}$ such that

$$\begin{aligned}&\lambda(u_n + \alpha_m d_u, w_n + \alpha_m d_w, \tilde{v}) + (1-\lambda)(u_n, w_n, v_n)\\ &= (u_n + \lambda\alpha_m d_u, w_n + \lambda\alpha_m d_w, \lambda\tilde{v} + (1-\lambda)v_n)\end{aligned}$$

is feasible for $\mathrm{P}(\varepsilon_n)$ whenever $m \geq M_\lambda$. As $\lim_{m\to\infty}(1-\lambda)\lambda\alpha_m v_n^\top\big(Wd_w - Bd_u\big) = 0$, there exists some $M_\lambda \geq \overline{N}$ such that

$$(1-\lambda)\lambda\alpha_m v_n^\top\big(Wd_w - Bd_u\big) \leq \lambda\frac{\varepsilon_n}{2}$$

for all $m \geq M_\lambda$. By (7.11), (7.12), and the feasibility of (u_n, w_n, v_n) for $\mathrm{P}(\varepsilon_n)$, we have

$$\begin{aligned}&\big(\lambda\tilde{v} + (1-\lambda)v_n\big)^\top\big(W(w_n + \lambda\alpha_m d_w) - B(u_n + \lambda\alpha_m d_u) - b\big)\\ &= \lambda\tilde{v}^\top\big(W(w_n + \lambda\alpha_m d_w) - B(u_n + \lambda\alpha_m d_u) - b\big)\\ &+ (1-\lambda)v_n^\top\big(Ww_n - Bu_n - b\big) + (1-\lambda)\lambda\alpha_m v_n^\top\big(Wd_w - Bd_u\big)\\ &\leq \lambda\frac{\varepsilon_n}{2} + (1-\lambda)\varepsilon_n + \lambda\frac{\varepsilon_n}{2} = \varepsilon_n\end{aligned}$$

for any $m \geq M_\lambda$ and feasibility follows from the linearity of the remaining restrictions.

As (7.10) implies $g^\top d_u + h^\top d_w < 0$,

$$g^\top(u_n + \lambda\alpha_m d_u) + h^\top(w_n + \lambda\alpha_m d_w) < g^\top u_n + h^\top w_n$$

holds for any $\lambda \in (0,1]$ and $m \geq M_\lambda$, which, by

$$\lim_{m\to\infty}\lim_{\lambda\to 0}\lambda(u_n + \lambda\alpha_m d_u, w_n + \lambda\alpha_m d_w, \lambda\tilde{v} + (1-\lambda)v_n) = (u_n, w_n, v_n),$$

yields a contradiction to the local optimality of (u_n, w_n, v_n) for $\mathrm{P}(\varepsilon_n)$. ∎

Remark 7.4. By Theorem 7.1 (c), the local minizers of the bilevel linear problem (7.1) are exactly the projections of local minimizers of its single-level reformulation (7.2). Unfortunately, projections of local minimizers of the KKT formulation (7.5) may fail to be local minimizers of (7.2) (see [16] for a counterexample with quadratic objective function). While $(\overline{u}, \overline{w})$ is a local minimizer of (7.2) if and only if $(\overline{u}, \overline{w}, \overline{v})$ is a local minimizer of (7.5) for any $\overline{v} \leq 0$ satisfying $W^\top\overline{v} = t$ and $\overline{v}^\top(W\overline{w} - B\overline{u} - b) = 0$ by Theorem 7.2 (c), the latter condition is hard to verify and poses immense challenges from an algorithmic point of view. A major advantage of relaxation $(\mathrm{P}(\varepsilon))$ is the fact that the projection of any accumulation point of a sequence of local minimizers of $(\mathrm{P}(\varepsilon))$ is guaranteed to be a local minimizer of (7.2) by Theorem 7.3 (c). Moreover, $(\mathrm{P}(\varepsilon))$ can be tackled by interior point methods that fail for (7.5) due to the violation of Slater's constraint qualification. In addition, the numerical tests carried out in [60] indicate that relaxing the complementarity constraint as detailed above has a stabilizing effect when standard nonlinear programming solvers are applied to MPECs and mathematical problems with complementarity constraints (MPCCs).

Remark 7.5. The smoothing method proposed in [25] replaces the complementarity constraint $v^\top(Ww - Bu - b) = 0$ in (7.5) with $v^\top(Ww - Bu - b) = \varepsilon$ for some

$\varepsilon > 0$. Note that this approach leads to perturbation rather than a relaxation. In contrast, the feasible set of $(\mathrm{P}(\varepsilon))$ encompasses the feasible set of the original problem (7.5), which may allow identify active parts of the complementarity constraints early on.

Remark 7.6. In [33], Ivanov details a Big-M approach where the disjunctive constraints induced by the complementarity condition $v^\top(Ww - Bu - b) = 0$ are modeled using binary variables. This formulation leads to linear subproblems and can be used to determine global minimizers (cf. [15, subsection 3.6.2]). In bilevel stochastic linear programming, the number of subproblems scales exponentially w.r.t. the number of scenarios. Thus, this approach heavily suffers from the curse of dimensionality and is in particular challenging, when an approximation of a high-dimensional distribution is employed. In contrast, the present approach allows us to approximate local solutions for the bilevel program by calculating local minimizers for the regularized problem and seems to be more suitable when the number of scenarios is large.

Combining the ideas in the proof of [17, Theorem 2.1] with [16] it is also possible to obtain convergence of global minimizers.

THEOREM 7.7. *Here, we let* $(\overline{u}, \overline{w}, \overline{v})$ *be an accumulation point of a sequence* $\{(u_n, w_n, v_n)\}_{n\in\mathbb{N}}$ *of global minimizers of problem* $\mathrm{P}(\varepsilon_n)$ *for* $\varepsilon_n \downarrow 0$. *Then* $(\overline{u}, \overline{w})$ *is a global minimizer of* (7.2).

Remark 7.8. An alternative regularization scheme employs a so-called NCP (nonlinear complementarity problem) function $\phi : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}^r$ (cf. [47]) to replace the conditions

$$Ww \leq Bu + b, \quad v \leq 0, \quad v^\top(Ww - Bu - b) = 0 \tag{7.13}$$

in (7.5) by the single constraint $\|\phi(-v, Bu + b - Ww)\|_1 = 0$. Popular choices of ϕ include the Fischer–Burmeister function $(\phi_{FB}(a,b))_i = \sqrt{a_i^2 + b_i^2} - a_i - b_i$ and the natural residual $(\phi_{NR}(a,b))_i = \min\{a_i, b_i\}$ (see [27]). Note that, a regularization of the type $\phi(-v, Bu+b-Ww) \leq \varepsilon$ with $\varepsilon > 0$ actually corresponds to a relaxation of all the conditions in (7.13). In contrast with the proposed relaxation $(\mathrm{P}(\varepsilon))$, feasibility of (u, w, v) for the regularized NCP reformulation does not even guarantee feasibility of w for the lower level problem.

Appendix A. We shall recall some technical results used throughout the paper.

THEOREM A.1 (see [38, Theorem 4.2]). *If D is positive semidefinite, the set-valued mapping* $C : \mathbb{R}^k \rightrightarrows \mathbb{R}^m$ *given by*

$$C(t) := \operatorname*{Argmin}_y \{y^\top D y + d_0^\top y \mid Ay \leq t\}$$

is Lipschitz continuous on $\operatorname{dom} C := \{t \in \mathbb{R}^k \mid C(t) \neq \emptyset\}$, *i.e., there exists a constant* $\Lambda > 0$ *such that* $d_\infty(C(t), C(t)) \leq \Lambda\|t - t\|$ *holds for any* $t, t \in \operatorname{dom} C$.

The following result is a well-known direct consequence of Lebesgue's dominated convergence theorem.

LEMMA A.2. *Let μ be a Borel-probability measure on* $\mathbb{R}^s$, $V \subseteq \mathbb{R}^n \times \mathbb{R}^s$ *open, and* $g : V \to \mathbb{R}$ *such that the following conditions are satisfied:*

(a) $g(\cdot, z)$ *is differentiable at* $x_0 \in V_\mu := \{x \mid (x,z) \in V\ \forall z \in \operatorname{supp}\, \mu\}$ *for μ-almost all* $z \in \mathbb{R}^s$ *and the derivative* $g(x_0, z)$ *is measurable w.r.t. z.*

(b) *There exists a neighborhood $U \subseteq V_\mu$ of x_0 such that*
   (i) *the integral $\int_{\mathbb{R}^s} g(x,z)\ \mu(dz)$ is well defined and finite for all $x \in U$ and*
   (ii) *there is an integrable function $m : U \to \mathbb{R}$ such that $|e(x,z)| \leq m(z)$ holds for all $x \in U \setminus \{x_0\}$ and μ-almost all $z \in \mathbb{R}^s$, where*

$$e(x,z) = \frac{1}{\|x - x_0\|}\Big(g(x,z) - g(x_0,z) - \nabla_x g(x_0,z)(x - x_0)\Big).$$

Then $h : V_\mu \to \overline{\mathbb{R}}$, $h(x) = \int_{\mathbb{R}^s} g(x,z)\ \mu(dz)$ is differentiable at x_0 and

$$\nabla h(x_0) = \int_{\mathbb{R}^s} \nabla_x g(x_0,z)\ \mu(dz).$$

Proof. Set

$$\varepsilon(x) := \frac{1}{\|x - x_0\|}\Big(h(x) - h(x_0) - \int_{\mathbb{R}^s} \nabla_x g(x_0,z)(x - x_0)\ \mu(dz)\Big).$$

By assumption, we have $\lim_{x\to x_0} |e(x,z)| = 0$ for μ-almost all $z \in \mathbb{R}^s$ and Lebesgue's dominated convergence theorem implies

$$\lim_{x\to x_0} |\varepsilon(x)| \leq \lim_{x\to x_0} \int_{\mathbb{R}^s} |e(x,z)|\ \mu(dz) = \int_{\mathbb{R}^s} \lim_{x\to x_0} |e(x,z)|\ \mu(dz) = 0,$$

which completes the proof. ∎

LEMMA A.3. *Let $\mathcal{C} \subseteq \mathbb{R}^{k\times s}$ be closed, then the set-valued mapping $\mathcal{T} : \mathbb{R}^k \rightrightarrows \mathbb{R}^l$,*

$$\mathcal{T}(t) = \{z \in \mathbb{R}^s \mid (t,z) \in \mathcal{C}\}$$

is outer semicontinuous (cf. [56]*), i.e.,* $\limsup_{t\to t_0} \mathcal{T}(t) \subseteq \mathcal{T}(t_0)\ \forall t_0 \in \mathbb{R}^k$.

Proof. By definition of the outer limit, $z \in \limsup_{t\to t_0} \mathcal{T}(t)$ holds iff there are sequences $\{t_n\}_{n\in\mathcal{N}} \subset \mathbb{R}^k$ and $\{z_n\}_{n\in\mathcal{N}} \subset \mathbb{R}^s$ such that

$$\lim_{n\to\infty} t_n = t_0, \quad \lim_{n\to\infty} z_n = z, \quad \text{and} \quad z_n \in \mathcal{T}(t_n)\ \forall n \in \mathbb{N}.$$

For any such sequences we have $(t_n, z_n) \in \mathcal{C}$ for all $n \in \mathcal{N}$ and thus $(t_0, z) \in \mathcal{C}$. Consequently, $z \in \mathcal{T}(t_0)$. ∎

REFERENCES

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

[2] K. Beer, *Lösung großer linearer Optimierungsaufgaben*, Deutscher Verlag der Wissenschaften, Berlin, 1977.

[3] D. Belomestny and V. Krätschmer, *Central limit theorems for law-invariant coherent risk measures*, J. Appl. Probab., 49 (2012), pp. 1–21.

[4] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton University Press, Princeton, NJ, 2009.

[5] C. Berge, *Espaces topologiques: Fonctions multivoques*, Coll. Universitaire de mathématiques 3, Dunod, Paris, 1959.

[6] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 1999.

[7] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[8] S. I. Birbil, G. Gürkan, and O. Listes, *Simulation-based Solution of Stochastic Mathematical Problems with cOmplementarity Constraints: Sample-path Analysis*, Technical report, ERIM Report Series Research in Management, ERS-2004-016-LIS, Rotterdam, The Netherlands, 2014.

[9] J. R. Birge and R. J.-B. Wets, *Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse*, Math. Programming Stud., 27 (1986), pp. 54–102.

[10] M. Carrion, J. M. Arroyo, and A. J. Conejo, *A bilevel stochastic programming approach for retailer futures market trading*, IEEE Trans. Power Syst., 24 (2009), pp. 1446–1456.

[11] P. Cheridito and T. Li, *Risk measures on Orlicz hearts*, Math. Finance, 19 (2009), pp. 189–214.

[12] M. Claus, *Advancing Stability Analysis of Mean-risk Stochastic Programs: Bilevel and Two-stage Models*, Ph.D. thesis, University of Duisburg-Essen, Duisburg, Germany, 2016.

[13] M. Claus, V. Krätschmer, and R. Schultz, *Weak continuity of risk functionals with applications to stochastic programming*, SIAM J. Optim., 27 (2017), pp. 91–108, https://doi.org/10.1137/15M1048689.

[14] V. DeMiguel and H. Xu, *A stochastic multiple-leader Stackelberg model: Analysis, computation, and application*, Oper. Res., 57 (2009), pp. 1220–1235.

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

[16] S. Dempe and J. Dutta, *Is bilevel programming a special case of a mathematical program with complementarity constraints?*, Math. Program., 131 (2012), pp. 37–48.

[17] S. Dempe and S. Franke, *Solution of Bilevel Optimization Problems using the KKT Approach*, Optimization, 68 (2019), pp. 1471–1489.

[18] S. Dempe, S. V. Ivanov, and A. Naumov, *Reduction of the bilevel stochastic optimization problem with quantile objective function to a mixed-integer problem*, Appl. Stoch. Models Bus. Ind., 33 (2017), pp. 544–554.

[19] B. C. Eaves, *On quadratic programming*, Management Sci., 17 (1971), pp. 698–711.

[20] I. Ekeland and R. Temam, *Analyse convexe et problèmes variationnels*, Dunod, Paris, 1974.

[21] H. Föllmer and A. Schied, *Convex measures of risk and trading constraints*, Finance Stoch., 6 (2002), pp. 429–447.

[22] H. Föllmer and A. Schied, *Stochastic Finance*, 3rd ed., Walter de Gruyter, Berlin, 2011.

[23] M. Frittelli and E. R. Gianin, *Law invariant convex risk measures*, Adv. Math. Econ. 7, Springer, Tokyo, 2005, pp. 42–53.

[24] M. Frittelli and E. R. Gianin, *Putting order in risk measures*, J. Bank. Finance, 26 (2002), pp. 1473–1486.

[25] M. Fukushima and J. S. Pang, *Convergence of a smoothing continuation method for mathematical programs with equilibrium constraints*, in Ill-posed Variational Problems and Regularization Techniques, M. Théra and R. Tichatschke, eds., Lecture Notes in Econom. and Math. Systems 477, Springer, New York, 1999, pp. 99–110.

[26] A. A. Gaivoronski and A. Werner, *Stochastic programming perspective on the agency problems under uncertainty*, Managing Safety of Heterogeneous Systems, Springer, 2012, pp. 137–167.

[27] A. Galántai, *Properties and construction of NCP functions*, Comput. Optim. Appl., 52 (2012), pp. 805–824.

[28] P. Gordan, *Über die Auflösungen linearer Gleichungen mit reellen Coefficienten*, Math. Annalen, 6 (1873), pp. 23–28.

[29] L. Hellemo, P. I. Barton, and A. Tomasgard, *Decision-dependent probabilities in stochastic programs with recourse*, Comput. Manag. Sci., 15 (2018), pp. 369–395.

[30] C. Henkel, *An Algorithm for the Global Resolution of Linear Stochastic Bilevel Programs*, Ph.D. thesis, University of Duisburg-Essen, Duisburg, Germany, 2014.

[31] J. Hu, J. E. Mitchell, J.-S. Pang, and B. Yu, *On linear programs with linear complementarity constraints*, J. Global Optim., 53 (2012), pp. 29–51.

[32] A. Inoue, *On the worst case conditional expectation*, J. Math. Anal. Appl., 286 (2003), pp. 237–247.

[33] S. V. Ivanov, *Bilevel stochastic linear programming problems with quantile criterion*, Autom. Remote Control, 75 (2014), pp. 107–118.

[34] S. V. Ivanov, *A bilevel stochastic programming problem with random parameters in the follower's objective function*, J. Appl. Ind. Math., 12 (2018), pp. 658–667.

[35] T. W. Jonsbråten, R. J.-B. Wets, and D. L. Woodruff, *A class of stochastic programs with decision dependent random elements*, Ann. Oper. Res., 82 (1998), pp. 83–106.

[36] P. Kall, A. Ruszczyński, and K. Frauendorfer, *Approximation techniques in stochastic programming*, Numerical Techniques for Stochastic Optimization, Y. Ermoliev and R. J.-B. Wets, eds., Springer Ser. Comput. Math. 10, Springer, Berlin, 1988, pp. 33–64.

[37] D. Klatte and B. Kummer, *Stability properties of infima and optimal solutions of parametric optimization problems*, Nondifferentiable Optimization: Motivations and Applications, Proceedings of the IIASA Workshop, Sopron, Lect. Notes Econ. Math. Syst. 255, Springer,

Berlin, 1984, pp. 215–229.

[38] D. Klatte and G. Thiere, *Error bounds for solutions of linear equations and inequalities*, ZOR–Math. Methods Oper. Res., 41 (1995), pp. 191–214.

[39] S. Kosuch, P. Le Bodic, J. Leung, and A. Lisser, *On a stochastic bilevel programming problem*, Networks, 59 (2012), pp. 107–116.

[40] R. M. Kovacevic and G. C. Pflug, *Electricity swing option pricing by stochastic bilevel optimization: A survey and new approaches*, European J. Oper. Res., 237 (2013), pp. 389–403.

[41] V. Krätschmer, A. Schied, and H. Zähle, *Comparative and qualitative robustness for law-invariant risk measures*, Finance Stoch., 18 (2014), pp. 271–295.

[42] V. Krätschmer, A. Schied, and H. Zähle, *Domains of weak continuity of statistical functionals with a view on robust statistics*, J. Multivariate Anal., 158 (2017), pp. 1–19.

[43] V. Krätschmer, A. Schied, and H. Zähle, *Qualitative and infinitesimal robustness of tail-dependent statistical functionals*, J. Multivariate Anal., 103 (2012), pp. 35–47.

[44] G.-H. Lin and M. Fukushima, *Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey*, Pac. J. Optim., 6 (2010), pp. 455–482.

[45] Y. Liu, H. Xu, and G.-H. Lin, *Stability analysis of two-stage stochastic mathematical programs with complementarity constraints via NLP regularization*, SIAM J. Optim., 21 (2011), pp. 669–705, https://doi.org/10.1137/100785685.

[46] B. Luderer, *Über die Äquivalenz nichtlinearer Optimierungsaufgaben*, Wiss. Z. Tech. Hochsch. Karl-Marx-Stadt, 26 (1983), pp. 257–258.

[47] O. L. Mangasarian, *Equivalence of the complementarity problem to a system of nonlinear equations*, SIAM J. Appl. Math., 31 (1976), pp. 89–92, https://doi.org/10.1137/0131009.

[48] A. Müller, *Certainty equivalents as risk measures*, Braz. J. Probab. Stat., 21 (2007), pp. 1–12.

[49] 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*, Transpor. Res. B-Meth., 42 (2008), pp. 843–860.

[50] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Oper. Res. Lett., 25 (1999), pp. 159–167.

[51] G. C. Pflug, *Some remarks on the value-at-risk and the conditional value-at-risk*, in Probabilistic Constrained Opimization - Methodology and Applications, Kluwer Academic Publishers, Dordrecht, 2000, pp. 272–281.

[52] A. Pichler, *A quantitative comparison of risk measures*, Ann. Oper. Res., 254 (2017), pp. 251–275.

[53] D. Pollard, *Convergence of Stochastic Processes*, Springer-Verlag, New York, 1984.

[54] S. M. Robinson, *Local epi-continuity and local optimization*, Math. Programming, 37 (1987), pp. 208–222.

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

[56] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 2009.

[57] E. Roghanian, S. J. Sadjadi, and M.-B. Aryanezhad, *A probabilistic bi-level linear multi-objective programming problem to supply chain planning*, Appl. Math. Comput., 188 (2007), pp. 786–800.

[58] W. Römisch, *Stability of Stochastic Programming Problems*, in Stochastic Programming, Handbooks Oper. Res. Management Sci. 10, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, 2003, pp. 483–554.

[59] H. Scheel and S. Scholtes, *Mathematical programs with equilibrium constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22.

[60] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), pp. 918–936, https://doi.org/10.1137/S1052623499361233.

[61] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, 2nd ed., MPS SIAM Ser. Optim. 9, SIAM, Philadelphia, 2014, https://doi.org/10.1137/1.9780898718751.

[62] A. Shapiro and H. Xu, *Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation*, Optimization, 57 (2008), pp. 395–418.

[63] G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York, 1986.

[64] A. Werner, *Bilevel Stochastic Programming Problems: Analysis and Application to Telecommunications*, Ph.D. thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2005.