



A Second-Order Sufficient Optimality Condition for Risk-Neutral Bi-level Stochastic Linear Programs

Matthias Claus¹ 

Received: 31 March 2020 / Accepted: 20 October 2020 / Published online: 10 November 2020
© The Author(s) 2020

Abstract

The expectation functionals, which arise in risk-neutral bi-level stochastic linear models with random lower-level right-hand side, are known to be continuously differentiable, if the underlying probability measure has a Lebesgue density. We show that the gradient may fail to be local Lipschitz continuous under this assumption. Our main result provides sufficient conditions for Lipschitz continuity of the gradient of the expectation functional and paves the way for a second-order optimality condition in terms of generalized Hessians. Moreover, we study geometric properties of regions of strong stability and derive representation results, which may facilitate the computation of gradients.

Keywords Bi-level stochastic linear programming · Risk-neutral model · Second-order optimality conditions · Lipschitz gradients

Mathematics Subject Classification 90C15 · 91A65

1 Introduction

We study bi-level stochastic linear programs with random right-hand side in the lower-level constraint system. The sequential nature of bi-level programming motivates a setting where the leader decides nonanticipatorily, while the follower can observe the realization of the randomness. A discussion of the related literature is provided in the recent [1]. A central result of [1] states that evaluating the leader's random outcome by taking the expectation leads to a continuously differentiable functional if the underlying probability measure is absolutely continuous w.r.t. the Lebesgue

Communicated by René Henrion.

✉ Matthias Claus
matthias.claus@uni-due.de

¹ University Duisburg-Essen, Essen, Germany

measure. This allows to formulate first-order necessary optimality conditions for the risk-neutral model. The main result of the present work provides sufficient conditions, namely boundedness of the support and uniform boundedness of the Lebesgue density of the underlying probability measure, that ensure Lipschitz continuity of the gradient of the expectation functional. Moreover, we show that the assumptions of [1] are too weak to even guarantee local Lipschitz continuity of the gradient. By the main result, second-order necessary and sufficient optimality conditions can be formulated in terms of generalized Hessians. As part of the preparatory work for the proof of the main result, we in particular show that any region of strong stability in the sense of [1, Definition 4.1] is a finite union of polyhedral cones. This representation is of independent interest, as it may facilitate the calculation or estimation of gradients of the expectation functional and thus enhance gradient descent-based approaches. The paper is organized as follows: The model and related results of [1] are discussed in Sect. 2, while the main result and a variation with weaker assumptions are formulated in Sect. 3. Sections 4 and 5 are dedicated to geometric properties of regions of strong stability and related projections that appear in the representation of the gradient. Results of these sections play an important role in the proof of the main result that is given in Sect. 6. A second-order sufficient optimality condition is formulated in Sect. 7. The paper concludes with a brief discussion of the results and an outlook in Sect. 8.

2 Model and Notation

Consider the optimistic formulation of a parametric bi-level linear program

$$\min_x \left\{ c^\top x + \min_y \{q^\top y : y \in \Psi(x, z)\} : x \in X \right\}, \tag{1}$$

where $z \in \mathbb{R}^s$ is a parameter and the data comprise a nonempty polyhedron $X \subseteq \mathbb{R}^n$, 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

$$\Psi(x, z) := \underset{y}{\text{Argmin}} \{d^\top y : Ay \leq Tx + z\}$$

with $A \in \mathbb{R}^{s \times m}$, $T \in \mathbb{R}^{s \times n}$ and $d \in \mathbb{R}^m$. By [1, Lemma 2.1], the extended real-valued mapping $f : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ given by

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

is real valued and Lipschitz continuous on the polyhedron

$$F = \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s : \exists y \in \mathbb{R}^m : Ay \leq Tx + z\}$$

if $\text{dom } f$ is nonempty. Let $Z : \Omega \rightarrow \mathbb{R}^s$ be a random vector on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote the induced Borel probability measure by $\mu_Z = \mathbb{P} \circ Z^{-1} \in$

$\mathcal{P}(\mathbb{R}^s)$. Furthermore, we introduce the set

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

If $\text{dom } f$ is nonempty and we impose the moment condition

$$\mu_Z \in \mathcal{M}_s^1 := \left\{ \mu \in \mathcal{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \|z\| \mu(dz) < \infty \right\},$$

the mapping $\mathbb{F} : F_Z \rightarrow L^1(\Omega, \mathcal{F}, \mathbb{P})$ given by $\mathbb{F}(x) := f(x, Z(\cdot))$ is well defined and Lipschitz continuous by [1, Lemma 2.4]. In a situation where the parameter z in (1) is given by a realization of the random vector Z that the follower can observe while the leader has to decide x nonanticipatorily, the upper-level outcome can be modeled by $\mathbb{F}(x)$. If we assume $X \subseteq F_Z$ and the leader’s decision is based on the expectation, we obtain the risk-neutral stochastic program

$$\min_x \{\mathbb{E}[\mathbb{F}(x)] : x \in X\}. \tag{2}$$

The following is shown in [1, Theorem 3.1, Corollary 4.7]:

Theorem 2.1 *Assume $\text{dom } f \neq \emptyset$ and that $\mu_Z \in \mathcal{M}_s^1$ is absolutely continuous w.r.t. the Lebesgue measure. Then, the mapping $Q_{\mathbb{E}} : F_Z \rightarrow \mathbb{R}$ defined by $Q_{\mathbb{E}}(x) = \mathbb{E}[\mathbb{F}(x)]$ is well defined, Lipschitz continuous and continuously differentiable at any $x_0 \in \text{int } F_Z$.*

We shall discuss some key ideas of the proof and introduce the relevant notation: Set

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

and then, f admits the representation

$$f(x, z) = c^\top x + \min_{\hat{y}} \{ \hat{q}^\top \hat{y} : \hat{y} \in \underset{\hat{y}'}{\text{Argmin}} \{ \hat{d}^\top \hat{y}' : \hat{A} \hat{y}' = Tx + z, \hat{y}' \geq 0 \} \}. \tag{3}$$

Remark 2.1 The subsequent analysis does not depend on the specific structure of \hat{q} , \hat{d} and \hat{A} and applies whenever (3) holds with some matrix \hat{A} satisfying $\text{rank } \hat{A} = s$.

As the rows of \hat{A} are linearly independent, the set

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

of lower-level base matrices is nonempty. 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.,

$\hat{A}_B^{-1}(Tx + z) \geq 0$, 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

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

Set

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

and assume $\text{dom } f \neq \emptyset$, and then,

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

holds for any $(x, z) \in F$. A key concept is the region of strong stability associated with a base matrix $\hat{A}_B \in \mathcal{A}^*$ given by the set

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

on which f coincides with the affine linear mapping

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

Under the assumptions of Theorem 2.1, we have

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

and the gradient of $\mathcal{Q}_{\mathbb{E}}$ admits the representation

$$\nabla \mathcal{Q}_{\mathbb{E}}(x) = c^\top + \sum_{\Delta \in D} \mu_Z[\mathcal{W}(x, \Delta)] \Delta \quad \forall x \in \text{int } F_Z \tag{4}$$

where $D := \{\hat{q}_B^\top \hat{A}_B^{-1} T : \hat{A}_B \in \mathcal{A}^*\}$, and the set-valued aggregation mappings $\mathcal{W}, \overline{\mathcal{W}} : \mathbb{R}^n \times D \rightrightarrows \mathbb{R}^s$ are given by

$$\mathcal{W}(x, \Delta) := \left\{ z \in \mathbb{R}^s \mid (x, z) \in \bigcup_{\hat{A}_B \in \mathcal{A}^* : \hat{q}_B^\top \hat{A}_B^{-1} T = \Delta} \text{int } \mathcal{S}(\hat{A}_B) \right\}$$

and $\overline{\mathcal{W}}(x, \Delta) := \left\{ z \in \mathbb{R}^s \mid (x, z) \in \bigcup_{\hat{A}_B \in \mathcal{A}^* : \hat{q}_B^\top \hat{A}_B^{-1} T = \Delta} \text{cl int } \mathcal{S}(\hat{A}_B) \right\},$

respectively (cf. [1, Theorem 4.3, Corollary 4.7]). Continuity of the $\nabla Q_{\mathbb{E}}$ follows from the fact that the outer semicontinuity of $\overline{\mathcal{W}}$ and

$$\sum_{\Delta \in D} \mu_Z[\overline{\mathcal{W}}(x, \Delta)] = 1 \quad \forall (x, \Delta) \in \text{int } F_Z \times D$$

imply continuity of the weight functional $M_{\Delta} : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$M_{\Delta}(x) := \mu_Z[\mathcal{W}(x, \Delta)] = \mu_Z[\overline{\mathcal{W}}(x, \Delta)]. \tag{5}$$

for any $\Delta \in D$.

3 Main Result

We shall first show that the assumptions of Theorem 2.1 are too weak to guarantee Lipschitz continuity $\nabla Q_{\mathbb{E}}$.

Example 3.1 Consider the case where

$$\hat{d} = (0, 0, 0, 0)^\top, \quad \hat{A} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & \frac{3}{2} & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad T = (0, 1)^\top.$$

The feasible set of the lower-level problem is compact for any parameters in the polyhedral cone $F = \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : z_1 \geq 0, x + z_2 \geq 0\}$, which implies that $\text{dom } f$ coincides with F for any $\hat{q} \in \mathbb{R}^4$. As the objective function is constant, any feasible base matrix is optimal for the lower-level problem. Denote the elements of $\mathcal{A} = \mathcal{A}^*$ by $\hat{A}_1, \dots, \hat{A}_6$, and let

$$\Theta_i = \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s : \hat{A}_i^{-1}(Tx + z) \geq 0\}$$

be the set of parameters for which \hat{A}_i is feasible for the lower-level problem. A straightforward calculation shows that we have

$$\begin{aligned} \hat{A}_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \hat{A}_1^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Theta_1 = \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : z_1 \geq 0, x + z_2 \geq 0\}, \\ \hat{A}_2 &= \begin{pmatrix} 1 & 1 \\ 0 & \frac{3}{2} \end{pmatrix}, \quad \hat{A}_2^{-1} = \begin{pmatrix} 1 & -\frac{2}{3} \\ 0 & \frac{2}{3} \end{pmatrix}, \quad \Theta_2 = \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 \leq x + z_2 \leq \frac{3}{2}z_1\}, \\ \hat{A}_3 &= \begin{pmatrix} 1 & 1 \\ 0 & \frac{1}{2} \end{pmatrix}, \quad \hat{A}_3^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 2 \end{pmatrix}, \quad \Theta_3 = \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 \leq x + z_2 \leq \frac{1}{2}z_1\}, \\ \hat{A}_4 &= \begin{pmatrix} 0 & 1 \\ 1 & \frac{3}{2} \end{pmatrix}, \quad \hat{A}_4^{-1} = \begin{pmatrix} -\frac{3}{2} & 1 \\ 1 & 0 \end{pmatrix}, \quad \Theta_4 = \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 \leq \frac{3}{2}z_1 \leq x + z_2\}, \\ \hat{A}_5 &= \begin{pmatrix} 0 & 1 \\ 1 & \frac{1}{2} \end{pmatrix}, \quad \hat{A}_5^{-1} = \begin{pmatrix} -\frac{1}{2} & 1 \\ 1 & 0 \end{pmatrix}, \quad \Theta_5 = \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 \leq \frac{1}{2}z_1 \leq x + z_2\} \end{aligned}$$

and

$$\hat{A}_6 = \begin{pmatrix} 1 & 1 \\ \frac{3}{2} & \frac{1}{2} \end{pmatrix}, \hat{A}_6^{-1} = \begin{pmatrix} -\frac{1}{2} & 1 \\ \frac{3}{2} & -1 \end{pmatrix}, \Theta_6 = \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : \frac{1}{2}z_1 \leq x + z_2 \leq \frac{3}{2}z_1\}.$$

Set $\hat{q} = (0, 0, -5, -3)^\top$, and let \hat{q}_i denote the part of upper-level objective function that is associated with \hat{A}_i . We have

$$\begin{aligned} \hat{q}_1^\top \hat{A}_1^{-1} T &= 0, & \hat{q}_2^\top \hat{A}_2^{-1} T &= -\frac{10}{3}, & \hat{q}_3^\top \hat{A}_3^{-1} T &= -6, \\ \hat{q}_4^\top \hat{A}_4^{-1} T &= 0, & \hat{q}_5^\top \hat{A}_5^{-1} T &= 0, & \hat{q}_6^\top \hat{A}_6^{-1} T &= -2 \end{aligned}$$

and a straightforward calculation yields

$$\begin{aligned} \mathcal{S}(\hat{A}_1) &= \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : z_1 = 0, x + z_2 \geq 0\} \\ &\cup \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : z_1 \geq 0, x + z_2 = 0\} \\ &= \text{bd } \Theta_1, \\ \mathcal{S}(\hat{A}_2) &= \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 = x + z_2 \leq \frac{3}{2}z_1\} \\ &\cup \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 \leq x + z_2 = \frac{3}{2}z_1\} \\ &= \text{bd } \Theta_2, \\ \mathcal{S}(\hat{A}_3) &= \Theta_3, \quad \mathcal{S}(\hat{A}_4) = \Theta_4, \quad \mathcal{S}(\hat{A}_6) = \Theta_6 \quad \text{and} \\ \mathcal{S}(\hat{A}_5) &= \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 = \frac{1}{2}z_1 \leq x + z_2\} \\ &\cup \{(x, z) \in \mathbb{R} \times \mathbb{R}^2 : 0 \leq \frac{1}{2}z_1 = x + z_2\} \\ &= \text{bd } \Theta_5. \end{aligned}$$

Let the density $\delta_Z : \mathbb{R}^2 \rightarrow \mathbb{R}$ of Z be given by

$$\delta_Z(t_1, t_2) = \begin{cases} \frac{1}{2\sqrt{\frac{1}{2}t_1 - t_2}}, & \text{if } 3 \leq t_1 \leq 4 \text{ and } \frac{1}{2}t_1 - 1 \leq t_2 < \frac{1}{2}t_1 \\ 0, & \text{else} \end{cases}$$

and set $c = 0$. We have $\text{supp } \mu_Z \subset \overline{\mathcal{W}}(0, -6)$, and it is easy to see that

$$\begin{aligned} \overline{\mathcal{W}}(x, -6) \cap \text{supp } \mu_Z &= \{z \in \mathbb{R}^2 : 3 \leq z_1, \leq 4, \frac{1}{2}z_1 - 1 \leq z_2 \leq \frac{1}{2}z_1 - x\} \\ \text{and } \overline{\mathcal{W}}(x, -2) \cap \text{supp } \mu_Z &= \{z \in \mathbb{R}^2 : 3 \leq z_1, \leq 4, \frac{1}{2}z_1 - x \leq z_2 < \frac{1}{2}z_1\} \end{aligned}$$

hold true whenever $x \in]0, 1]$ (Fig. 1).

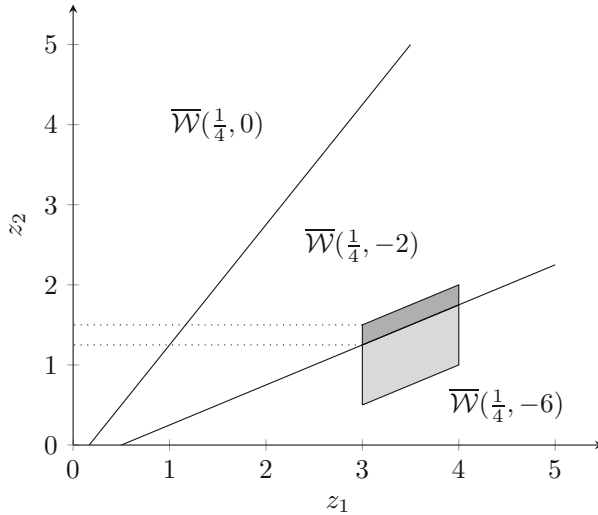


Fig. 1 In Fig. 1, the darker square depicts the intersection of $\text{supp } \mu_Z$ and $\overline{W}(\frac{1}{4}, -2)$, while the lighter square is $\overline{W}(\frac{1}{4}, -6) \cap \text{supp } \mu_Z$. The distance between the dotted lines is $x = \frac{1}{4}$

Thus, $\text{supp } \mu_Z \subset \overline{W}(x, -6) \cup \overline{W}(x, -2)$ holds for any $x \in [0, 1]$ and a simple calculation shows that

$$\begin{aligned} \nabla Q_{\mathbb{E}}(x) &= -6\mu_Z[\overline{W}(x, -6)] - 2\mu_Z[\overline{W}(x, -2)] \\ &= -6(1 - \sqrt{x}) - 2\sqrt{x} = 4\sqrt{x} - 6 \end{aligned}$$

is not locally Lipschitz continuous at $x = 0$.

Our main result is the following sufficient conditions for Lipschitz continuity of $\nabla Q_{\mathbb{E}}$:

Theorem 3.1 *Assume $\text{dom } f \neq \emptyset$ and let μ_Z be absolutely continuous w.r.t. the Lebesgue measure and have a bounded support as well as a uniformly bounded density. Then, $Q_{\mathbb{E}}$ is differentiable on $\text{int } F_Z$ with Lipschitz continuous gradient.*

Note the density in Example 3.1 is not bounded. The proof of Theorem 3.1 requires some preliminary work and will be given in Sect. 6. If the support of μ_Z is unbounded, we still obtain a weaker estimate for the gradients:

Theorem 3.2 *Assume $\text{dom } f \neq \emptyset$ and let μ_Z be absolutely continuous w.r.t. the Lebesgue measure and have a uniformly bounded density. Then, $Q_{\mathbb{E}}$ is differentiable on $\text{int } F_Z$ and for any $\epsilon > 0$ there exists a constant $L(\epsilon) > 0$ such that*

$$|\nabla Q_{\mathbb{E}}(x) - \nabla Q_{\mathbb{E}}(x')| \leq L(\epsilon)\|x - x'\| + \epsilon$$

holds for all $x, x' \in \text{int } F_Z$.

4 On the Geometry of Regions of Strong Stability

In view of (4) and (5), the gradient $\nabla Q_{\mathbb{E}}(x)$ is given by a weighted sum of the probabilities of the sets $\mathcal{W}(x, \Delta)$ or $\overline{\mathcal{W}}(x, \Delta)$ for $\Delta \in D$. As these sets are defined using regions of strong stability, we shall first study properties of the sets $\mathcal{S}(\hat{A}_B)$ with $\hat{A}_B \in \mathcal{A}^*$.

Remark 4.1 Example 3.1 shows that regions of strong stability are not convex in general.

Proposition 4.1 Assume $\text{dom } f \neq \emptyset$, then

$$\mathcal{S}(\hat{A}_B) = \mathcal{S}(\hat{A}_B) + \ker(T, I_s)$$

holds for any $\hat{A}_B \in \mathcal{A}^*$.

Proof The above result immediately follows from the fact that the quantities involved in the definition of $\mathcal{S}(\hat{A}_B)$ only depend on $Tx + z$. □

Corollary 4.1 Assume $\text{dom } f \neq \emptyset$ and $n \geq 1$, then no region of strong stability has any extremal points.

Proof Let (x, z) be an arbitrary point of some region of strong stability $\mathcal{S}(\hat{A}_B)$. The n -dimensional kernel of (T, I_s) contains some nonzero element (x_0, z_0) , and we have $(x - x_0, z - z_0), (x + x_0, z + z_0) \in \mathcal{S}(\hat{A}_B)$ by Proposition 4.1. Thus, $(x, z) = \frac{1}{2}(x - x_0, z - z_0) + \frac{1}{2}(x + x_0, z + z_0)$ is no extremal point of $\mathcal{S}(\hat{A}_B)$. □

Our main result on the structure of $\mathcal{S}(\hat{A}_B)$ is the following:

Theorem 4.1 Assume $\text{dom } f \neq \emptyset$, then any region of strong stability is a union of at most $(s + 1)^{|\mathcal{A}^*|}$ polyhedral cones and at most $(s + 1)^{|\mathcal{A}^*| - 1}$ of these cones have a nonempty interior. Moreover, the multifunction $\mathcal{S} : \mathcal{A}^* \rightrightarrows \mathbb{R}^n \times \mathbb{R}^s$ is polyhedral, i.e., $\text{gph } \mathcal{S}$ is a finite union of polyhedra.

Before we get to proof of Theorem 4.1, we will establish the following auxiliary result:

Lemma 4.1 Let $\mathcal{W} := \{\xi \in \mathbb{R}^k : V\xi < 0\}$ with $V \in \mathbb{R}^{l \times k}$ be nonempty, then

$$\text{cl } \mathcal{W} = \overline{\mathcal{W}} := \{\xi \in \mathbb{R}^k : V\xi \leq 0\}.$$

Proof The inclusion $\text{cl } \mathcal{W} \subseteq \overline{\mathcal{W}}$ is trivial. Moreover, for any $\xi_0 \in \mathcal{W} = \text{int } \overline{\mathcal{W}}$ and $\xi \in \overline{\mathcal{W}}$ the line segment principle (cf. [6, Lemma 2.1.6]) implies $[\xi_0, \xi] \subseteq \mathcal{W}$ and thus $\xi \in \text{cl } \mathcal{W}$. □

We are now ready to prove Theorem 4.1.

Proof (Proof of Theorem 4.1) Denote the elements of the finite set \mathcal{A}^* by $\hat{A}_1, \dots, \hat{A}_l$ and the associated parts of the objective function by $\hat{q}_1, \dots, \hat{q}_l$. Fix any index $i \in \{1, \dots, l\}$; then for any $(x, z) \in F$ satisfying $\hat{A}_i^{-1}(Tx + z) \geq 0$, the constraint $c^\top x + \hat{q}_i^\top \hat{A}_i^{-1}(Tx + z) = f(x, z)$ in the definition of $\mathcal{S}(\hat{A}_i)$ can be reformulated as

$$\left((\hat{q}_i^\top \hat{A}_i^{-1} - \hat{q}_j^\top \hat{A}_j^{-1})(Tx + z) \leq 0 \quad \vee \quad \hat{A}_j^{-1}(Tx + z) \not\geq 0 \right) \quad \forall j = 1, \dots, l.$$

Introducing the sets

$$\begin{aligned} \Theta_i &:= \{(x, z) \in F : \hat{A}_i^{-1}(Tx + z) \geq 0\} \\ &:= \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s : \hat{A}_i^{-1}(Tx + z) \geq 0\}, \\ \Gamma_{ij0} &:= \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s : (\hat{q}_i^\top \hat{A}_i^{-1} - \hat{q}_j^\top \hat{A}_j^{-1})(Tx + z) \leq 0\} \text{ and} \\ \Gamma_{ijk} &:= \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s : e_k^\top \hat{A}_j^{-1}(Tx + z) < 0\} \end{aligned}$$

with indices $j = 1, \dots, l$ and $k = 1, \dots, s$ and using the fact that $\mathcal{S}(\hat{A}_i)$ is closed by the Lipschitz continuity of f , we obtain the representation

$$\begin{aligned} \mathcal{S}(\hat{A}_i) &= \text{cl} \left(\Theta_i \cap \bigcap_{j=1, \dots, l} \bigcup_{k=0, \dots, s} \Gamma_{ijk} \right) \\ &= \bigcup_{\substack{\alpha \in \{0, \dots, s\}^l, \\ \Theta_i \cap \bigcap_{j=1, \dots, l} \Gamma_{ij\alpha_j} \neq \emptyset}} \text{cl} \left(\Theta_i \cap \bigcap_{\substack{j=1, \dots, l, \\ \alpha_j=0}} \Gamma_{ij0} \cap \bigcap_{\substack{j=1, \dots, l, \\ \alpha_j \neq 0}} \Gamma_{ij\alpha_j} \right). \end{aligned}$$

As $\Theta_i \cap \bigcap_{j=1, \dots, l, \alpha_j=0} \Gamma_{ij0}$ is convex and closed, while $\bigcap_{j=1, \dots, l, \alpha_j \neq 0} \Gamma_{ij\alpha_j}$ is convex and open, [6, Proposition 2.1.10] yields

$$\mathcal{S}(\hat{A}_i) = \bigcup_{\substack{\alpha \in \{0, \dots, s\}^l, \\ \Theta_i \cap \bigcap_{j=1, \dots, l} \Gamma_{ij\alpha_j} \neq \emptyset}} \left(\Theta_i \cap \bigcap_{\substack{j=1, \dots, l, \\ \alpha_j=0}} \Gamma_{ij0} \cap \text{cl} \bigcap_{\substack{j=1, \dots, l, \\ \alpha_j \neq 0}} \Gamma_{ij\alpha_j} \right)$$

The sets $\Theta_i \cap \bigcap_{j=1, \dots, l, \alpha_j=0} \Gamma_{ij0}$ are obviously polyhedral cones, and Lemma 4.1 implies

$$\begin{aligned} &\text{cl} \bigcap_{\substack{j=1, \dots, l, \\ \alpha_j \neq 0}} \Gamma_{ij\alpha_j} \\ &= \left\{ (x, z) \in \mathbb{R}^n \times \mathbb{R}^s : e_{\alpha_j}^\top \hat{A}_j^{-1}(Tx + z) \leq 0 \quad \forall j = 1, \dots, l : \alpha_j \neq 0 \right\} \end{aligned}$$

$$= \bigcap_{\substack{j=1, \dots, l, \\ \alpha_j \neq 0}} \text{cl } \Gamma_{ij\alpha_j}.$$

Moreover, for any $\alpha_i \in \{1, \dots, s\}$ we have $e_{\alpha_i}^\top \hat{A}_i^{-1} \neq 0$ and thus

$$\begin{aligned} & \text{int} \left(\Theta_i \cap \bigcap_{j=1, \dots, l} \text{cl } \Gamma_{ij\alpha_j} \right) \\ & \subseteq \text{int} \left(\Theta_i \cap \text{cl } \Gamma_{ii\alpha_i} \right) \\ & \subseteq \text{int} \left\{ (x, z) \in \mathbb{R}^n \times \mathbb{R}^s : (e_{\alpha_i}^\top \hat{A}_i^{-1} T, e_{\alpha_i}^\top \hat{A}_i^{-1}) \begin{pmatrix} x \\ z \end{pmatrix} = 0 \right\} \\ & = \emptyset. \end{aligned}$$

The second part of the theorem is an immediate consequence of the finiteness of \mathcal{A}^* . \square

Corollary 4.2 *Assume $\text{dom } f \neq \emptyset$, then any region of strong stability is star shaped and contains the n -dimensional kernel of (T, I_s) .*

Proof Radial convexity is an immediate consequence of Theorem 4.1, as any region of strong stability contains the line segments from the origin to any feasible point. The second statement directly follows from Proposition 4.1. \square

Two-stage stochastic programming can be understood as the special case of bi-level stochastic programming where the objectives of leader and follower coincide. In this case, any region of strong stability is a polyhedral cone and thus convex:

Proposition 4.2 *Assume $\text{dom } f \neq \emptyset$ and $\hat{q} = \alpha \hat{d}$ for some $\alpha > 0$. Then, any region of strong stability is a polyhedral cone.*

Proof We shall use the notation of the proof of Theorem 4.1 and denote the part of \hat{d} associated with \hat{A}_i by \hat{d}_i . Fix any $(x, z) \in F$ and consider any base matrices $\hat{A}_i, \hat{A}_j \in \mathcal{A}^*$ that are feasible and thus optimal for the lower-level problem. As

$$\hat{q}_i^\top \hat{A}_i^{-1} (Tx + z) = \alpha \hat{d}_i^\top \hat{A}_i^{-1} (Tx + z) = \alpha \hat{d}_j^\top \hat{A}_j^{-1} (Tx + z) = \hat{q}_j^\top \hat{A}_j^{-1} (Tx + z),$$

both base matrices are also optimal with respect to the upper-level objective function. Thus, $\mathcal{S}(\hat{A}_i)$ coincides with the polyhedral cone Θ_i . \square

Remark 4.2 As $\hat{d} = (0, 0, 0, 0)^\top$ holds in Example 3.1, we see the assumption $\hat{q} = \alpha \hat{d}$ for some $\alpha \in \mathbb{R}$ in Proposition 4.2 cannot be replaced with the weaker condition that $\{\hat{q}, \hat{d}\}$ is linearly dependent.

5 Properties of the Aggregation Mappings

We shall now study the aggregation mappings \mathcal{W} and $\overline{\mathcal{W}}$ defined in Sect. 2. The following result is the counterpart of Theorem 4.1:

Theorem 5.1 *Assume $\text{dom } f \neq \emptyset$, then the multifunction $\overline{\mathcal{W}}$ is polyhedral. Moreover, $\overline{\mathcal{W}}(x, \Delta)$ is a finite union of polyhedra for any $(x, \Delta) \in \mathbb{R}^n \times D$.*

The proof of Theorem 5.1 will be based on the following auxiliary result:

Lemma 5.1 *Let $C_1, \dots, C_l \subseteq \mathbb{R}^k$ be closed and convex. Then,*

$$\text{cl int } \bigcup_{i=1, \dots, l} C_i = \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i.$$

Proof As the sets C_1, \dots, C_l are closed and the interior of a union is contained in the union of the interiors, we have

$$\begin{aligned} \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i &\supseteq \text{cl int } \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i \supseteq \text{cl } \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} \text{int } C_i \\ &= \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} \text{cl int } C_i = \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i, \end{aligned}$$

where the first equality is due to the fact that the closure of a union equals the union of the closures and the second equation is a direct consequence of the line segment principle. Thus,

$$\text{cl int } \bigcup_{i=1, \dots, l} C_i \supseteq \text{cl int } \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i = \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i.$$

For the reverse inclusion, suppose that there is some

$$x \in \left(\text{cl int } \bigcup_{i=1, \dots, l} C_i \right) \setminus \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i.$$

By definition, there are sequences $\{x_n\}_{n \in \mathbb{N}} \subset \mathbb{R}^k$ and $\{\epsilon_n\}_{n \in \mathbb{N}} \subset \mathbb{R}_{>0}$ satisfying $x_n \rightarrow x$ and $B_{\epsilon_n}(x_n) \subseteq \bigcup_{i=1, \dots, l} C_i$ for all $n \in \mathbb{N}$. As $\bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i$ is closed, there exists some $N \in \mathbb{N}$ such that $x_n \notin \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i$ for all $n \geq N$. Together with the previous considerations, the strong separation theorem (cf. [9, Theorem 11.4]) yields the existence of some $\delta_N \in (0, \epsilon_N]$ such that

$$B_{\delta_N}(x_N) \subseteq \left(\bigcup_{i=1, \dots, l} C_i \right) \setminus \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i = \bigcup_{i=1, \dots, l: \text{int } C_i = \emptyset} C_i.$$

As any C_i with $\text{int } C_i = \emptyset$ is contained in an affine subspace of dimension strictly smaller than k (cf. [2, Section 2.5.2]), we obtain the contradiction

$$\emptyset \neq \text{int } B_{\delta_N}(x_N) \subseteq \text{int } \bigcup_{i=1, \dots, l: \text{int } C_i = \emptyset} C_i = \emptyset.$$

Thus,

$$\left(\text{cl int } \bigcup_{i=1, \dots, l} C_i \right) \setminus \bigcup_{i=1, \dots, l: \text{int } C_i \neq \emptyset} C_i = \emptyset,$$

which completes the proof. □

Corollary 5.1 *Let $C \subseteq \mathbb{R}^k$ be a finite union of polyhedra (polyhedral cones). Then, $\text{cl int } C$ is a finite union of polyhedra (polyhedral cones).*

Proof The above statement is an immediate consequence of Lemma 5.1. □

Proof (Proof of Theorem 5.1) As D is finite, it is sufficient to consider the multifunctions $\overline{\mathcal{W}}(\cdot, \Delta) : \mathbb{R}^n \rightrightarrows \mathbb{R}^s$ for fixed $\Delta \in D$. We have

$$\text{gph } \overline{\mathcal{W}}(\cdot, \Delta) = \bigcup_{\hat{A}_B \in \mathcal{A}^*: \hat{q}_B^\top \hat{A}_B^{-1} T = \Delta} \text{cl int } \mathcal{S}(\hat{A}_B),$$

which is a finite union of polyhedra by Corollary 5.1. Similarly, $\overline{\mathcal{W}}(x, \Delta)$ admits the representation

$$\overline{\mathcal{W}}(x, \Delta) = \bigcup_{\hat{A}_B \in \mathcal{A}^*: \hat{q}_B^\top \hat{A}_B^{-1} T = \Delta} \{z \in \mathbb{R}^s : (x, z) \in \text{cl int } \mathcal{S}(\hat{A}_B)\}.$$

By Theorem 4.1 and Corollary 5.1, the set

$$\{z \in \mathbb{R}^s : (x, z) \in \text{cl int } \mathcal{S}(\hat{A}_B)\}$$

is the intersection of a finite union of polyhedral cones and the affine subspace $\{(x', z') \in \mathbb{R}^n \times \mathbb{R}^s : x' = x\}$ and thus a finite union of polyhedral cones for any $x \in \mathbb{R}^n$ and any $\hat{A}_B \in \mathcal{A}^*$. □

The following result on \mathcal{W} is a simple consequence of the fact that the constraint system describing a region of strong stability only imposes conditions on $(Tx + z)$.

Proposition 5.1 *Assume $\text{dom } f \neq \emptyset$, then*

$$\mathcal{W}(x, \Delta) = \mathcal{W}(x', \Delta) + \{T(x - x')\}$$

holds for any $x, x' \in \mathbb{R}^n$ and $\Delta \in D$.

Proof Fix any $x, x' \in \mathbb{R}^n, z \in \mathbb{R}^s$ and set $z' = z + T(x - x')$, then $Tx + z = Tx' + z'$ and thus

$$\begin{aligned} f(x, z) - c^\top x &= \min_y \{q^\top y : y \in \underset{y'}{\text{Argmin}}\{d^\top y' : Ay' \leq Tx + z\}\} \\ &= f(x', z') - c^\top x'. \end{aligned}$$

Similarly, for any $\hat{A}_B \in \mathcal{A}^*, (x, z) \in \mathcal{S}(\hat{A}_B)$ holds if and only if

1. there exists some $y \in \mathbb{R}^m$ such that $Ay \leq Tx + z = Tx' + z'$,
2. $\hat{A}_B^{-1}(Tx' + z') = \hat{A}_B^{-1}(Tx + z) \geq 0$ and
3. $\hat{q}_B^\top \hat{A}_B^{-1}(Tx' + z') = \hat{q}_B^\top \hat{A}_B^{-1}(Tx + z) = f(x, z) - c^\top x = f(x', z') - c^\top x'$,

i.e., if and only if $(x', z') \in \mathcal{S}(\hat{A}_B)$. We conclude that

$$\begin{aligned} z \in \mathcal{W}(x, \Delta) &\Leftrightarrow \exists \hat{A}_B \in \mathcal{A}^* : \Delta = \hat{q}_B^\top \hat{A}_B^{-1} T, (x, z) \in \text{int } \mathcal{S}(\hat{A}_B) \\ &\Leftrightarrow \exists \hat{A}_B \in \mathcal{A}^* : \Delta = \hat{q}_B^\top \hat{A}_B^{-1} T, (x', z') \in \text{int } \mathcal{S}(\hat{A}_B) \\ &\Leftrightarrow z' \in \mathcal{W}(x', \Delta) \end{aligned}$$

holds for any $\Delta \in D$, which completes the proof. □

6 Proof of the Main Result

We are finally ready to prove Theorem 3.1 based on the results of Sects. 4 and 5 as well as the two following auxiliary results:

Lemma 6.1 *Assume $\text{dom } f \neq \emptyset$, and let $\mu_Z \in \mathcal{P}(\mathbb{R}^s)$ be absolutely continuous w.r.t. the Lebesgue measure, then*

$$\mu_Z [\mathcal{W}(x, \Delta) \setminus (\mathcal{W}(x, \Delta) + \{t\})] \leq \mu_Z [\overline{\mathcal{W}}(x, \Delta) \setminus (\overline{\mathcal{W}}(x, \Delta) + \{t\})]$$

holds for any $x \in \mathbb{R}^n, \Delta \in D$ and $t \in \mathbb{R}^s$.

Proof By the arguments used in the proof of [1, Lemma 4.2], we have

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

where $\mathcal{N}_x \subset \mathbb{R}^s$ is contained in a finite union of hyperplanes. Consequently,

$$\begin{aligned} &\mathcal{W}(x, \Delta) \setminus (\mathcal{W}(x, \Delta) + \{t\}) \\ &\subseteq \left[\overline{\mathcal{W}}(x, \Delta) \setminus (\overline{\mathcal{W}}(x, \Delta) + \{t\}) \right] \cup \left[(\overline{\mathcal{W}}(x, \Delta) + \{t\}) \setminus (\mathcal{W}(x, \Delta) + \{t\}) \right] \\ &= \left[\overline{\mathcal{W}}(x, \Delta) \setminus (\overline{\mathcal{W}}(x, \Delta) + \{t\}) \right] \cup \left[(\overline{\mathcal{W}}(x, \Delta) \setminus \mathcal{W}(x, \Delta)) + \{t\} \right] \\ &\subseteq \left[\overline{\mathcal{W}}(x, \Delta) \setminus (\overline{\mathcal{W}}(x, \Delta) + \{t\}) \right] \cup \left[\mathcal{N}_x + \{t\} \right] \end{aligned}$$

and the above statement is a direct consequence of the fact that the Lebesgue measure of \mathcal{N}_x equals zero. □

Lemma 6.2 *Assume $\text{dom } f \neq \emptyset$ and let μ_Z be absolutely continuous w.r.t. the Lebesgue measure and have a bounded support as well as a uniformly bounded density. Then, the weight functional M_Δ is Lipschitz continuous on $\text{int } F_Z$ for any $\Delta \in D$.*

Proof By definition of $\mathcal{W}(x, \Delta)$, Proposition 5.1 and Lemma 6.1,

$$\begin{aligned} & |\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x', \Delta)]| \\ &= |\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x, \Delta) + \{T(x' - x)\}]| \\ &\leq \mu_Z[\mathcal{W}(x, \Delta) \setminus (\mathcal{W}(x, \Delta) + \{T(x' - x)\})] \\ &\quad + \mu_Z[(\mathcal{W}(x, \Delta) + \{T(x' - x)\}) \setminus \mathcal{W}(x, \Delta)] \\ &\leq \mu_Z[\overline{\mathcal{W}}(x, \Delta) \setminus (\overline{\mathcal{W}}(x, \Delta) + \{T(x' - x)\})] \\ &\quad + \mu_Z[(\overline{\mathcal{W}}(x, \Delta) + \{T(x' - x)\}) \setminus \overline{\mathcal{W}}(x, \Delta)] \end{aligned}$$

holds for any fixed $\Delta \in D$. As both

$$\overline{\mathcal{W}}(x, \Delta) \setminus (\overline{\mathcal{W}}(x, \Delta) + \{T(x' - x)\}) \quad \text{and} \quad (\overline{\mathcal{W}}(x, \Delta) + \{T(x' - x)\}) \setminus \overline{\mathcal{W}}(x, \Delta)$$

are contained in

$$\mathcal{H}_{x,x'} := \{v + l \cdot T(x' - x) : v \in \text{bd } \overline{\mathcal{W}}(x, \Delta), l \in [-1, 1]\}$$

and there exists a finite upper bound $\alpha \in \mathbb{R}$ for the Lebesgue density of μ_Z , we have

$$|\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x', \Delta)]| \leq 2\alpha\lambda^s[\mathcal{H}_{x,x'} \cap \text{supp } \mu_Z],$$

where λ^s denotes the s -dimensional Lebesgue measure. By Theorem 5.1, the boundary of $\overline{\mathcal{W}}(x, \Delta)$ is contained in a finite union of lower-dimensional polyhedra. Let \mathbb{H}_x denote a set of such cones with minimal cardinality. It is a straightforward conclusion from the proofs of Theorem 4.1, Theorem 5.1 and Lemma 5.1 that the cardinality of \mathbb{H}_x can be bounded by a constant $K \in \mathbb{N}$ that does not depend on x . Moreover, as any $H \in \mathbb{H}_x$ is contained in some hyperplane, the $s - 1$ -dimensional Lebesgue measure of $H \cap \text{supp } \mu_Z$ is at most $\text{diam}(\text{supp } \mu_Z)^{s-1}$. Thus,

$$\begin{aligned} & |\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x', \Delta)]| \\ &\leq 2\alpha \sum_{H \in \mathbb{H}_x} \lambda^s[\{v + l \cdot T(x' - x) : v \in H, l \in [-1, 1]\} \cap \text{supp } \mu_Z] \\ &\leq 4\alpha K \cdot \text{diam}(\text{supp } \mu_Z)^{s-1} \cdot \|T\|_{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^s)} \|x' - x\|, \end{aligned}$$

by Cavalieri’s principle, which completes the proof. □

Proof (Proof of Theorem 3.1) Continuous differentiability on $\text{int } F_Z$ is a direct consequence of [1, Corollary 4.7]. Fix any $x, x' \in \text{int } F_Z$; then, (4) and Lemma 6.2 yield

$$\begin{aligned} & \|\nabla Q_{\mathbb{E}}(x) - \nabla Q_{\mathbb{E}}(x')\| \\ & \leq \sum_{\Delta \in D} |\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x', \Delta)]| \cdot \|\Delta\| \\ & \leq 4\alpha K |D| \cdot \text{diam}(\text{supp } \mu_Z)^{s-1} \cdot \max_{\Delta \in D} \|\Delta\| \cdot \|T\|_{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^s)} \|x' - x\| \end{aligned}$$

and thus the desired Lipschitz continuity. □

Proof (Proof of Theorem 3.2) Fix any $\kappa > 0$. As μ_Z is tight by [3, Theorem 1.3], there exists a compact set $C(\kappa) \subset \mathbb{R}^s$ such that $\mu_Z[\mathbb{R}^s \setminus C(\kappa)] < \kappa$. Combining this with the estimate from the first part of the proof of Lemma 6.2 and using the same notation established therein, we see that

$$\begin{aligned} |\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x', \Delta)]| & \leq 2\mu_Z[\mathcal{H}_{x,x'}] \\ & = 2\mu_Z[\mathcal{H}_{x,x'} \cap C(\kappa)] + 2\mu_Z[\mathcal{H}_{x,x'} \setminus C(\kappa)] \\ & \leq 2\alpha \lambda^s [\mathcal{H}_{x,x'} \cap C(\kappa) \cap \text{supp } \mu_Z] + 2\kappa \end{aligned}$$

holds for any $\Delta \in D$. Thus,

$$\begin{aligned} & |\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x', \Delta)]| \\ & \leq 2\kappa + 2\alpha \sum_{H \in \mathbb{H}_x} \lambda^s [\{v + l \cdot T(x' - x) : v \in H, l \in [-1, 1]\} \cap C(\kappa) \cap \text{supp } \mu_Z] \\ & \leq 2\kappa + 4\alpha K \cdot \text{diam}(C(\kappa) \cap \text{supp } \mu_Z)^{s-1} \cdot \|T\|_{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^s)} \|x' - x\|. \end{aligned}$$

We therefore have

$$\begin{aligned} & \|\nabla Q_{\mathbb{E}}(x) - \nabla Q_{\mathbb{E}}(x')\| \\ & \leq \sum_{\Delta \in D} |\mu_Z[\mathcal{W}(x, \Delta)] - \mu_Z[\mathcal{W}(x', \Delta)]| \cdot \|\Delta\| \\ & \leq 4\alpha K |D| \text{diam}(C(\kappa) \cap \text{supp } \mu_Z)^{s-1} \max_{\Delta \in D} \|\Delta\| \cdot \|T\|_{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^s)} \|x' - x\| + 2|D|\kappa \end{aligned}$$

and choosing $\kappa = \frac{\epsilon}{2|D|}$ yields the desired estimate. □

Remark 6.1 The constant $L(\epsilon)$ derived in the proof of Theorem 3.2 depends on ϵ . If the support of μ_Z is unbounded, we have $L(\epsilon) \rightarrow \infty$ as $\epsilon \downarrow 0$.

7 A Sufficient Second-Order Optimality Condition

Under the conditions of Theorem 3.1, $\nabla Q_{\mathbb{E}}$ is Lipschitz continuous on $\text{int } F_Z$ and thus differentiable almost everywhere on $\text{int } F_Z$ by Rademacher’s theorem. Let $D \subseteq \text{int } F_Z$

denote the set of points at which $\nabla Q_{\mathbb{E}}$ is differentiable, then generalized Clarke’s Hessian of $Q_{\mathbb{E}}$ at some $x \in \text{int } F_Z$ is the nonempty, convex and compact set

$$\partial^2 Q_{\mathbb{E}}(x) = \text{conv} \left\{ H \in \mathbb{R}^{n \times n} : \exists \{x_k\}_{k \in \mathbb{N}} \subseteq \mathcal{D} : x_k \rightarrow x, \nabla^2 Q_{\mathbb{E}}(x_k) \rightarrow H \right\}.$$

We have $\partial^2 Q_{\mathbb{E}}(x) = \{\nabla^2 Q_{\mathbb{E}}(x)\}$ whenever $x \in \mathcal{D}$.

Let the feasible set of (2) be given by $X = \{x \in \mathbb{R}^n : Bx \leq b\}$ with some $B \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$. The following second-order sufficient condition is based on [7]:

Theorem 7.1 *Assume $\text{dom } f \neq \emptyset$, $X \subseteq \text{int } F_Z$ and let μ_Z be absolutely continuous w.r.t. the Lebesgue measure and have a bounded support as well as a uniformly bounded density. Moreover, let (\bar{x}, \bar{u}) be a KKT point of (2), i.e.,*

$$\nabla Q_{\mathbb{E}}(\bar{x}) + B^T \bar{u} = 0, \quad B\bar{x} \leq b, \quad \bar{u}^T (B\bar{x} - b) = 0, \quad \bar{u} \geq 0$$

and assume that any $H \in \partial^2 Q_{\mathbb{E}}(\bar{x})$ is positive definite on

$$\left\{ h \in \mathbb{R}^n : \begin{array}{l} e_i^T B h = 0 \quad \forall i : \bar{u}_i > 0 \\ e_j^T B h \leq 0 \quad \forall j : \bar{u}_j = e_j^T B \bar{x} = 0 \end{array} \right\}.$$

Then, \bar{x} is a strict local minimizer with order 2 of (2), i.e., there exist a neighborhood U of \bar{x} and a constant $L > 0$ such that

$$Q_{\mathbb{E}}(x) > Q_{\mathbb{E}}(\bar{x}) + L \|x - \bar{x}\|^2$$

holds for any $x \in X \cap U$.

Proof This is a straightforward conclusion from [7, Theorem 1]. □

Remark 7.1 There are various other approaches for optimization problems with data in the class $C^{1,1}$, which consists of differentiable functions with locally Lipschitzian gradients. For instance, second-order optimality conditions can also be formulated based on Dini (cf. [5, Section 4.4]) or Riemann (cf. [8]) derivatives.

8 Conclusions

We have derived sufficient conditions for Lipschitz continuity of the gradient of the expectation functional arising from a bi-level stochastic linear program with random right-hand side in the lower-level constraint system. Invoking the structure of the upper level constraints, we used this result to formulate a second-order sufficient optimality condition for the risk-neutral bi-level stochastic program in terms of the generalized Hessian of $Q_{\mathbb{E}}$. Moreover, the main result on the geometry of regions of strong stability and its counterpart for the aggregation mapping $\overline{\mathcal{W}}$ may facilitate the computation or sample-based estimation of gradients of the expectation functional, which enhances

gradient descent-based methods. As any region of strong stability is a finite union of polyhedral cones, a promising approach is to employ spherical radial decomposition techniques to calculate $\nabla Q_{\mathbb{E}}$ (cf. [4, Chapter 4]). The details are beyond the scope of this paper but shall be addressed in future research.

Acknowledgements The author thanks the Deutsche Forschungsgemeinschaft for its support via the Collaborative Research Center TRR 154.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Statement Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Burtscheidt, J., Claus, M., Dempe, S.: Risk-averse models in bilevel stochastic linear programming. *SIAM J. Optim.* **30**(1), 377–406 (2020)
2. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
3. Billingsley, P.: *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, 2nd edn. Wiley, New York (1999)
4. Genz, A., Bretz, F.: Computation of multivariate normal and t probabilities. In: *Lecture Notes in Statistics*, vol. 195. Springer, Heidelberg (2009)
5. Ginchev, I., La Torre, D., Rocca, M.: $C^{k,1}$ functions, characterization, Taylor's formula and optimization: a survey. *Real Anal. Exch.* **35**(2), 311–342 (2009/2010)
6. Hiriart-Urruty, J.-B., Lemaréchal, C.: *Fundamentals of Convex Analysis*. Springer, Berlin (2001)
7. Klatté, D., Tammer, K.: On second-order sufficient optimality conditions for $c^{1,1}$ -optimization problems. *Optimization* **19**(2), 169–179 (1988)
8. La Torre, D., Rocca, M.: $C^{1,1}$ functions and optimality conditions. *J. Concr. Appl. Math.* **3**(1), 41–54 (2005)
9. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.