Conditioning of linear-quadratic two-stage stochastic optimization problems*

Konstantin Emich† René Henrion† Werner Römisch‡

September 5, 2013

Abstract

In this paper a condition number for linear-quadratic two-stage stochastic optimization problems is introduced as the Lipschitz modulus of the multifunction assigning to a (discrete) probability distribution the solution set of the problem. Being the outer norm of the Mordukhovich coderivative of this multifunction, the condition number can be estimated from above explicitly in terms of the problem data by applying appropriate calculus rules. Here, a chain rule for the extended partial second-order subdifferential recently proved by Mordukhovich and Rockafellar plays a crucial role. The obtained results are illustrated for the example of two-stage stochastic optimization problems with simple recourse.

Keywords:
Stochastic optimization, two-stage linear-quadratic problems, conditioning, coderivative calculus, simple recourse
Mathematics Subject Classification: 90C15, 90C31, 49K40

1 Introduction

In numerical analysis, a condition number of a given mathematical problem represents an upper bound on the ratio of the (relative) solution error to the (relative) data error. Its size provides information on the difficulty of solving the problem and its reciprocal is often proportional to the perturbation distance of the problem from ill-posedness. In [2] an increasing interest in conditioning of various optimization models is detected (see, for example, [3, 10, 8, 14, 23]) and general concepts are developed for deriving condition numbers of generalized equations.

In this paper, we consider convex stochastic optimization models of the form

$$\min\Big\{\int_{\mathbb{R}^s} g(x,\xi)P(d\xi) : x \in X\Big\}, \tag{1}$$

where X is a nonempty closed convex subset of $\mathbb{R}^m$, P a probability distribution on $\mathbb{R}^s$ and g is an extended real-valued measurable function on $\mathbb{R}^m \times \mathbb{R}^s$ such that $g(\cdot,\xi)$ is convex for all ξ in the support of P. Particular cases of (1) are two-stage linear or linear-quadratic stochastic programs. Our aim is to derive results on the conditioning of such optimization models.

So far the only paper studying conditioning of such stochastic optimization models is [21]. There, the authors assumed for (1) that in addition X is polyhedral, P has finite support, $g(\cdot,\xi)$ is piecewise linear for all ξ in the support of P and that (1) has a unique solution x_0. Their approach consists in considering empirical or Monte Carlo sampling methods for solving (1) and in studying the required sample size N such that the unique (random) solution $\hat{x}_N$ of the empirical approximation

$$\min\Big\{N^{-1}\sum_{i=1}^{N} g(x,\xi^i) : x \in X\Big\}, \tag{2}$$

*This work was supported by the DFG Research Center Matheon "Mathematics for key technologies" in Berlin
†Weierstrass Institute Berlin, Germany
‡Humboldt University Berlin, Germany

satisfies problem (1) with high probability. The ξ^i, $i \in \mathbb{N}$, in (2) are independent and identically distributed $\mathbb{R}^s$-valued random samples with common distribution P. Motivated by large deviation techniques they consider the number $\beta > 0$ such that

$$\lim_{N\to\infty} N^{-1} \log\left(1 - P(\hat{x}_N = x_0)\right) = -\beta$$

as a condition measure of problem (1). More precisely, the number $(2\beta)^{-1}$ is called condition number of (1). Moreover, the authors derived an approximate formula for the condition number.

In this paper, we study linear-quadratic two-stage stochastic optimization problems (see [17]) and their conditioning. Such problems may be introduced by considering the Lagrangian (see also [16])

$$\mathcal{L}(x,z) = \langle c, x\rangle + \tfrac{1}{2}\langle x, Cx\rangle + \mathbb{E}\left(\langle z, h(\xi) - T(\xi)x\rangle - \tfrac{1}{2}\langle z, Bz\rangle\right) \quad (x \in X, z \in Z),$$

where X and Z are nonempty convex polyhedra in $\mathbb{R}^m$ and $\mathbb{R}^k$, respectively, B and C are symmetric and positive semidefinite matrices, $c \in \mathbb{R}^m$, $h(\xi)$ is a random vector in $\mathbb{R}^k$, $T(\xi)$ is a stochastic $k \times m$-matrix, and $\mathbb{E}$ denotes expectation with respect to a probability distribution P. Primal and dual problems are then associated by general duality and given by

$$\min_{x\in X}\max_{z\in Z} \mathcal{L}(x,z) \qquad \text{and} \qquad \max_{z\in Z}\min_{x\in X} \mathcal{L}(x,z).$$

The primal problem is of the form

$$\min\left\{\langle c, x\rangle + \tfrac{1}{2}\langle x, Cx\rangle + \mathbb{E}\left(\Phi(x,\xi)\right) | x \in X\right\}, \tag{3}$$

where x is the first-stage decision and

$$\Phi(x,\xi) = \max_{z\in Z}\left\{\langle z, h(\xi) - T(\xi)x\rangle - \tfrac{1}{2}\langle z, Bz\rangle\right\}. \tag{4}$$

We assume that a (k,r)-matrix W and a vector $q \in \mathbb{R}^r$ are given and consider the following explicit description of the polyhedron Z:

$$Z = \{z \in \mathbb{R}^k : W^\top z \le q\}. \tag{5}$$

As shown in the Appendix, if B is positive definite then (4) may be reformulated as

$$\Phi(x,\xi) = \inf_{y\ge 0}\left\{\langle q, y\rangle + \tfrac{1}{2}\left\langle h(\xi) - T(\xi)x - Wy, B^{-1}(h(\xi) - T(\xi)x - Wy)\right\rangle\right\}. \tag{6}$$

Hence, $\Phi(x,\xi)$ corresponds to minimal second stage (random) costs associated with a recourse decision $y \in \mathbb{R}^r$ taken upon observing $\xi \in \mathbb{R}^s$ and penalizing the violation of the equality

$$Wy = h(\xi) - T(\xi)x \tag{7}$$

by means of a quadratic penalty term instead of meeting (7) exactly as in classical two-stage linear stochastic optimization. The latter would require to assume relative complete recourse. In the context of two-stage linear-quadratic stochastic optimization we do not insist on this assumption.

As shown in [19, Theorems 9 and 23], solutions of two-stage stochastic programs do not depend in a Lipschitzian way on the underlying probability distribution in general. More precisely, the behaviour of the growth function

$$\psi_P(\tau) = \inf\left\{\int_{\mathbb{R}^s} g(x,\xi)P(d\xi) - v(P) | d(x, S(P)) \ge \tau, x \in X\right\} \quad (\tau \ge 0) \tag{8}$$

near $\tau = 0$ has to be studied. Here, $v(P)$ and $S(P)$ are the optimal value and the solution set of (1), respectively, and $d(x, S(P))$ refers to the distance of $x \in X$ to $S(P)$. Lipschitzian dependence can be concluded if the function ψ_P has linear growth close to $\tau = 0$ or if ψ_P has quadratic growth and a Lipschitz stability argument due to [20] (see also [1, Section 4.4.1]) is employed. If the support of P is finite, two-stage linear stochastic programs satisfy a linear growth condition and two-stage linear-quadratic stochastic programs a quadratic growth condition, respectively. Indeed, we provide a calmness result for solutions in the latter case (see Proposition 3.2).

Therefore, we assume that the random vector ξ has a discrete uniform probability distribution with atoms or scenarios $\xi^1, \ldots, \xi^N$. Then the optimization problem (3) can be written as

$$\min \left\{ \langle c, x\rangle + \tfrac{1}{2} \langle x, Cx\rangle + N^{-1} \textstyle\sum_{i=1}^N \Phi(x, \xi^i) | x \in X \right\}. \tag{9}$$

In order to study the dependence of solutions to (9) on the probability distribution we consider the vector $p := \left(\xi^1, \ldots, \xi^N\right)$ of scenarios and introduce the solution set mapping $S : \mathbb{R}^{Ns} \rightrightarrows \mathbb{R}^m$ as

$$S(p) := \{x \in X | x \text{ solves } (9)\}. \tag{10}$$

Our aim is to apply concepts from [2] in order to associate a condition number with the two-stage stochastic optimization problem (9).

2 Basic Concepts and Notation

As usual, we denote by 'gr M' the graph of some multifunction M. Denote by $\mathbb{B}_\delta(y)$ the closed ball of radius δ around some y in a metric space. We recall the following two basic properties of multifunctions $M : X \rightrightarrows Y$ between metric spaces X, Y:

Definition 2.1 *M has the Aubin property at a point $(\bar{x}, \bar{y}) \in \operatorname{gr} M$ if there exist $L, \delta > 0$ such that*

$$d\left(y, M\left(x_1\right)\right) \leq L d\left(x_1, x_2\right) \quad \forall x_1, x_2 \in \mathbb{B}_\delta(\bar{x}),\ \forall y \in M\left(x_2\right) \cap \mathbb{B}_\delta(\bar{y}). \tag{11}$$

As a weaker condition, M is said to be calm at $(\bar{x}, \bar{y}) \in \operatorname{gr} M$ if there exist $L, \delta > 0$ such that

$$d\left(y, M\left(\bar{x}\right)\right) \leq L d\left(x, \bar{x}\right) \quad \forall x \in \mathbb{B}_\delta(\bar{x}),\ \forall y \in M\left(x\right) \cap \mathbb{B}_\delta(\bar{y}).$$

The constant

$$\operatorname{lip} M(\bar{x}, \bar{y}) := \inf \left\{ L | \exists \delta > 0 : (11) \right\} \tag{12}$$

is called the *graphical modulus* of M at $(\bar{x}, \bar{y})$ [18, p.377]. It can be interpreted as the Lipschitz modulus of the multifunction M. For the following definitions and properties we refer the reader to [12] and [18].

Definition 2.2 *Let $C \subseteq \mathbb{R}^m$ be a closed subset and $\bar{x} \in C$. The Mordukhovich normal cone to C at $\bar{x}$ is defined by*

$$N_C(\bar{x}) := \left\{ x^* | \exists \left(x_n, x_n^*\right) \to (\bar{x}, x^*) : x_n \in C, x_n^* \in \left[T_C(x_n)\right]^0 \right\}.$$

Here, $\left[T_C(x_n)\right]^0$ refers to the Fréchet normal cone to C at x_n, which is the negative polar of the contingent cone

$$T_C(x) := \left\{ d \in \mathbb{R}^m \left| \exists t_k \downarrow 0, d_k \to d : x + t_k d_k \in C, \forall k \right. \right\}. \tag{13}$$

to C at x_n. For an extended-real-valued, lower semicontinuous function $f : \mathbb{R}^m \to \bar{\mathbb{R}}$ with $|f(\bar{x})| < \infty$, the Mordukhovich normal cone induces a subdifferential via

$$\partial f(\bar{x}) := \left\{ x^* | \left(x^*, -1\right) \in N_{\operatorname{epi} f}\left(\bar{x}, f(\bar{x})\right) \right\}.$$

If $f : \mathbb{R}^m \to \mathbb{R}$ is locally Lipschitz around $\bar{x}$ and $g : \mathbb{R}^m \to \bar{\mathbb{R}}$ is proper and lower semicontinuous with $|g(\bar{x})| < \infty$, then the following sum rule applies:

$$\partial\left(f + g\right)(\bar{x}) \subseteq \partial f(\bar{x}) + \partial g(\bar{x}). \tag{14}$$

Definition 2.3 *Let $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a multifunction with closed graph. The Mordukhovich coderivative $D^*M(\bar{x}, \bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ of M at some $(\bar{x}, \bar{y}) \in \operatorname{gr} M$ is defined as*

$$D^*M(\bar{x}, \bar{y})(y^*) := \left\{ x^* \in \mathbb{R}^n | \left(x^*, -y^*\right) \in N_{\operatorname{gr} M}(\bar{x}, \bar{y}) \right\}$$

When M is single-valued, i.e., $\bar{y} = M(\bar{x})$, we simply write $D^*M(\bar{x})$ instead of $D^*M(\bar{x}, M(\bar{x}))$.

If $f : \mathbb{R}^m \to \mathbb{R}$ is locally Lipschitz around $\bar{x}$, then the following scalarization formula holds true:

$$D^* f(\bar{x})(y^*) = \partial \langle y^*, f \rangle (\bar{x}). \tag{15}$$

Definition 2.4 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a lower semicontinuous function which is finite at $x \in \mathbb{R}^n$. For an element $u \in \partial f(x)$, the second-order subdifferential of f at (x, u) is a multifunction $\partial^2 f(x, u) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by*

$$\partial^2 f(x, u)(w) := (D^* \partial f)(x, u)(w) \quad \forall w \in \mathbb{R}^n.$$

If $\partial f(x)$ is single-valued, then, similar to Definition 2.3, we simply write $\partial^2 f(x)$.

Definition 2.5 *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be a lower semicontinuous function which is finite at $(x, z) \in \mathbb{R}^n \times \mathbb{R}^m$. The partial subdifferential is defined as $\partial_x f(x, z) := \partial f(\cdot, z)(x)$. Following [13], for $(x, z) \in \mathbb{R}^n \times \mathbb{R}^m$ and any $u \in \partial_x f(x, z)$, the (extended) partial second-order subdifferential of f is a multifunction $\partial_x^2 f(x, z, u) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ defined by*

$$\partial_x^2 f(x, z, u)(w) := (D^* \partial_x f)(x, z, u)(w) \quad \forall w \in \mathbb{R}^n.$$

If $\partial_x f(x, z)$ is single-valued, then, similar to Definition 2.3, we simply write $\partial_x^2 f(x, z)$.

3 A condition number for linear-quadratic two-stage stochastic optimization problems.

We consider the representation (4) of the optimal second-stage costs with the polyhedron Z defined in (5):

$$\Phi(x, \xi) = \sup_z \left\{ \langle h(\xi) - T(\xi)x, z \rangle - \tfrac{1}{2} \langle z, Bz \rangle \,|\, W^\top z \le q \right\}$$

Throughout the rest of the paper we shall make the following assumptions for Φ:

- B is symmetric and positive definite.
- The polyhedron Z is nonempty and its description (5) satisfies the Linear Independence Constraint Qualification (i.e., at each point of Z the active rows of the matrix W^T are linearly independent).
- T and h are continuously differentiable.

As a consequence of these assumptions, Φ is finite-valued and $\Phi(\cdot, \xi)$ is convex for any $\xi \in \mathbb{R}^s$. Now, the solution set to our optimization problem (9) is equivalently characterized by the generalized equation

$$0 \in \partial_x \Psi(x, p) + N_X(x), \tag{16}$$

where ∂_x and N denote the partial subdifferential and the normal cone, respectively, in the sense of convex analysis and

$$\Psi(x, p) := \langle c, x \rangle + \tfrac{1}{2} \langle x, Cx \rangle + N^{-1} \textstyle\sum_{i=1}^N \Phi(x, \xi^i) \quad (x \in \mathbb{R}^m, \quad p = (\xi^1, \ldots, \xi^N) \in \mathbb{R}^{Ns}). \tag{17}$$

Consequently, the solution set mapping S defined in (10) can also be written as

$$S(p) = \{x \in \mathbb{R}^m | (16) \text{ is satisfied}\}. \tag{18}$$

Following [2], we call $\operatorname{lip} S(\bar{p}, \bar{x})$ as defined in (12) the *condition number* of problem (9) at a point $(\bar{p}, \bar{x}) \in \operatorname{gr} S$. By definition, $\operatorname{lip} S(\bar{p}, \bar{x}) < \infty$ if and only if S has the Aubin property at $(\bar{p}, \bar{x})$ (see Def. 2.1). Moreover [18, Theorem 9.40], the condition number can be calculated as

$$\operatorname{lip} S(\bar{p}, \bar{x}) = \sup_{x^* \in \mathbb{B}} \sup_{p^* \in D^* S(\bar{p}, \bar{x})(x^*)} \|p^*\|, \tag{19}$$

where $D^* S(\bar{p}, \bar{x})$ refers to the Mordukhovich coderivative of S at $(\bar{p}, \bar{x})$ (see Def. 2.3). We also recall the well-known Mordukhovich criterion [11, Theorem 5.7] stating that S has the Aubin property at $(\bar{p}, \bar{x})$ if and only if $D^* S(\bar{p}, \bar{x})(0) = 0$.

The following observation follows from standard results of parametric nonlinear programming (see, e.g., [1, Remark 4.14]) via the positive definiteness of B and the Linear Independence Constraint Qualification for Z:

Proposition 3.1 *Let* $(\bar{x}, \bar{\xi}) \in \mathbb{R}^m \times \mathbb{R}^s$ *be arbitrary. Then, the function* Φ *is Fréchet differentiable with* $\nabla_x \Phi(\bar{x}, \bar{\xi}) = -T^\top(\bar{\xi}) z(h(\bar{\xi}) - T(\bar{\xi})\bar{x})$*, where* $z(v)$ *is the unique element of*

$$\operatorname*{argmax}_{W^\top z \leq q} \langle v, z \rangle - \tfrac{1}{2} \langle z, Bz \rangle . \tag{20}$$

Moreover, since $z(\cdot)$ *is locally Lipschitz,* $\nabla_x \Phi$ *is locally Lipschitz too around* $(\bar{x}, \bar{\xi})$.

Corollary 3.1 *Let* $\bar{x} \in \mathbb{R}^m$ *and* $\bar{p} = \left(\bar{\xi}^1, \ldots, \bar{\xi}^N\right) \in \mathbb{R}^{Ns}$*. Then, the partial gradient* $\nabla_x \Psi(\bar{x}, \bar{p})$ *of the function* Ψ *defined in (17) exists, is Lipschitz continuous around* $(\bar{x}, \bar{p})$ *and is given by*

$$\nabla_x \Psi(\bar{x}, \bar{p}) = c + C\bar{x} + N^{-1} \sum_{i=1}^{N} \nabla_x \Phi(\bar{x}, \bar{\xi}^i).$$

In other words, (9) is a $\mathcal{C}^{1,1}$ optimization problem. Using Corollary 3.1 we are able to show that the assumptions on Φ imply calmness of S (see Def. 2.1).

Proposition 3.2 *Let* X *be bounded. The solution set mapping* S *defined in (18) is calm at any* $(\bar{x}, \bar{p}) \in \operatorname{gr} S$ *and it holds*

$$d(x, S(\bar{p})) \leq \frac{L}{c} \|p - \bar{p}\| \quad (x \in S(p) \cap U, p \in V), \tag{21}$$

where the constant $c > 0$ *and the (bounded) neighborhood* U *of* $S(\bar{p})$ *appear in the quadratic growth condition*

$$\Psi(x, \bar{p}) - \Psi(\bar{x}, \bar{p}) \geq c\, d(x, S(\bar{p}))^2 \quad (x \in X \cap U), \tag{22}$$

V *is a bounded neighborhood of* $\bar{p}$ *and* $L > 0$ *the local Lipschitz constant of* $\nabla_x \Psi(\cdot, \cdot)$ *on* $U \times V$.

Proof. Let $(\bar{x}, \bar{p}) \in \operatorname{gr} S$. The objective function $\Psi(\cdot, \bar{p})$ is convex piecewise linear-quadratic and, hence, satisfies a quadratic growth condition due to [9, Theorem 2.7]. Hence, there exist a bounded neighborhood U of $S(\bar{p})$ and a constant $c > 0$ such that (22) holds. Next we make use of the results in [1, Section 4.4.1] on Lipschitz stability of nonlinear programs in case of a fixed feasible set and obtain

$$d(x, S(\bar{p})) \leq \frac{1}{c} \sup_{y \in U} \|\nabla_x \Psi(y, p) - \nabla_x \Psi(y, \bar{p})\|$$

for any $x \in S(p) \cap U$ and any p in some neighborhood V of $\bar{p}$. The result follows from Corollary 3.1.

■

For more general results on the stability of solutions to $\mathcal{C}^{1,1}$ problems we refer, e.g., to [6],[7].
Now we are in a position to formulate an upper estimate for the coderivative of our solution mapping S in (18) as it will be required in an upper estimation of the condition number (19):

Proposition 3.3 *Let* $(\bar{p}, \bar{x}) \in \operatorname{gr} S$*, where* $\bar{x} \in X$ *and* $\bar{p} := \left(\bar{\xi}^1, \ldots, \bar{\xi}^N\right) \in \mathbb{R}^{Ns}$*. Assume that the multifunction*

$$M(w) := \{(p, x) \,|\, w \in \nabla_x \Psi(x, p) + N_X(x)\} \tag{23}$$

is calm at $(0, \bar{p}, \bar{x})$ *(see Definition 2.1). Then,*

$$D^* S\left(\bar{p}, \bar{x}\right)(x^*) \subseteq \left\{p^* | \exists v^* : (-x^*, p^*) \in \partial_x^2 \Psi\left(\bar{x}, \bar{p}\right)(v^*) + D^* N_X(\bar{x}, -\nabla_x \Psi\left(\bar{x}, \bar{p}\right))(v^*) \times \{0\}\right\}. \tag{24}$$

Proof. By Corollary 3.1, there exists a neighbourhood $\mathcal{U}$ of $(\bar{x}, \bar{p})$ such that the solution mapping S is locally described by

$$S(p) = \{x | 0 \in f(x, p) + N_X(x)\} \quad \forall (x, p) \in \mathcal{U},$$

where $f(x, p) := \nabla_x \Psi(x, p)$ is Lipschitz on $\mathcal{U}$. From the equivalence

$$(p, x) \in \operatorname{gr} S \Longleftrightarrow g(p, x) := (x, -f(x, p)) \in \operatorname{gr} N_X \tag{25}$$

we see that $\operatorname{gr} S = g^{-1}(\operatorname{gr} N_X)$ for a locally Lipschitzian mapping g. As observed in [22, Proposition 5.2], our calmness assumption implies calmness of the multifunction

$$w := (w_1, w_2) \mapsto \{(p, x) \,|w_2 - \nabla_x \Psi(x, p) \in N_X(x + w_1)\} = \{(p, x) \,|g(p, x) + w \in \operatorname{gr} N_X\}$$

at $(0, 0, \bar{p}, \bar{x})$. This allows us to invoke [4, Theorem 4.1], in order to derive from (25) the inclusion

$$N_{\operatorname{gr} S}(\bar{p}, \bar{x}) \subseteq \bigcup \{D^* g(\bar{p}, \bar{x})(w^*) \,|w^* \in N_{\operatorname{gr} N_X}(g(\bar{p}, \bar{x}))\}. \tag{26}$$

With the partition $w^* = (u^*, v^*)$ and defining the functions $\pi(p, x) := x$ and $\tilde{f}(p, x) := -f(x, p)$ we obtain that $g = \left(\pi, \tilde{f}\right)$ and, thus,

$$\begin{aligned} D^* g(\bar{p}, \bar{x})(u^*, v^*) &= \partial \langle w^*, g\rangle(\bar{p}, \bar{x}) = \partial\left(\langle u^*, \pi\rangle + \left\langle v^*, \tilde{f}\right\rangle\right)(\bar{p}, \bar{x}) \\ &\subseteq \partial\langle u^*, \pi\rangle(\bar{p}, \bar{x}) + \partial\left\langle v^*, \tilde{f}\right\rangle(\bar{p}, \bar{x}) = (0, u^*) + D^*\tilde{f}(\bar{p}, \bar{x})(v^*). \end{aligned}$$

Here we exploited the sum rule (14) and the scalarization formula (15). Moreover, using the definition of the coderivative it is easy to see by virtue of [18, Exercise 6.7] that

$$(x^*, p^*) \in D^* f(\bar{x}, \bar{p})(-v^*) \Longleftrightarrow (p^*, x^*) \in D^*\tilde{f}(\bar{p}, \bar{x})(v^*).$$

As a consequence,

$$D^* g(\bar{p}, \bar{x})(u^*, v^*) \subseteq \{(p^*, x^*) \,|\, (x^* - u^*, p^*) \in D^* f(\bar{x}, \bar{p})(-v^*)\}.$$

Combining this with (26) yields

$$D^* S(\bar{p}, \bar{x})(x^*) \subseteq \{p^* | \exists (u^*, v^*) \in N_{\operatorname{gr} N_X}(g(\bar{p}, \bar{x})) : (-x^* - u^*, p^*) \in D^* f(\bar{x}, \bar{p})(-v^*)\}$$

which leads to (24) upon recalling the definitions of g and f as well as the fact that $D^*\nabla_x\Psi(\bar{x}, \bar{p}) = \partial_x^2\Psi(\bar{x}, \bar{p})$ (see Def. 2.5).

■

4 Computation of $\partial_x^2\Psi$

In order to apply Proposition 3.3, we have to compute explicitly the partial second-order subdifferential $\partial_x^2\Psi$ (explicit formulae for the other term $D^* N_X$ are available from the literature, see, e.g., [5]). As a first step, we reduce the computation of $\partial_x^2\Psi$ to that of $\partial_x^2\Phi$:

Proposition 4.1 *Under the assumption of Proposition 3.3 holding at some $(\bar{p}, \bar{x}) \in \operatorname{gr} S$, where $\bar{x} \in X$ and $\bar{p} := \left(\bar{\xi}^1, \ldots, \bar{\xi}^N\right) \in \mathbb{R}^{Ns}$ one gets that, for all $v^* \in \mathbb{R}^m$,*

$$\partial_x^2\Psi(\bar{x}, \bar{p})(v^*) \subseteq \left\{\left(C^\top v^* + N^{-1}\sum\nolimits_{i=1}^N x_i^*, N^{-1}p^*\right) | \, (x_i^*, p_i^*) \in \partial_x^2\Phi\left(\bar{x}, \bar{\xi}^i\right)(v^*) \quad (i = 1, \ldots, N)\right\}.$$

Proof. Defining $p := \left(\xi^1, \ldots, \xi^N\right)$ and $\tilde{\Phi}_i(x, p) := \Phi\left(x, \xi^i\right)$ for $i = 1, \ldots, N$ and (x, p) in a neighbourhood of $(\bar{x}, \bar{p})$, we may write $\tilde{\Phi}_i = \Phi \circ \vartheta_i$, where $\vartheta_i(x, p) = \left(x, \xi^i\right)$ and infer that $\nabla_x\tilde{\Phi}_i = (\nabla_x\Phi) \circ A^i$ with a surjective matrix

$$A^i := \begin{pmatrix} I & 0 \\ 0 & B_i \end{pmatrix}; \quad B_i := \Big(0, \ldots, 0, \underset{i}{I}, 0, \ldots, 0\Big).$$

Now, the coderivative chain rule in [12, Theorem 1.66] yields that

$$D^*\nabla_x\tilde{\Phi}_i(\bar{x}, \bar{p}) = \left[A^i\right]^\top D^*\nabla_x\Phi\left(\bar{x}, \bar{\xi}^i\right) = \left[A^i\right]^\top \partial_x^2\Phi\left(\bar{x}, \bar{\xi}^i\right) \quad (i = 1, \ldots, N).$$

On the other hand, $\nabla_x \Psi(x,p) = c + Cx + N^{-1}\sum_{i=1}^N \nabla_x \tilde{\Phi}_i(x,p)$ by (17). Therefore, exploiting Definition 2.5 and the calculus rules (14) and (15), one ends up with

$$\partial_x^2 \Psi(\bar{x},\bar{p})(v^*) = D^* \nabla_x \Psi(\bar{x},\bar{p})(v^*) = \partial \langle v^*, \nabla_x \Psi\rangle (\bar{x},\bar{p}) \subseteq \left(C^\top v^*, 0\right) + N^{-1}\sum_{i=1}^N \partial \left\langle v^*, \nabla_x \tilde{\Phi}_i \right\rangle (\bar{x},\bar{p})$$
$$= \left(C^\top v^*, 0\right) + N^{-1}\sum_{i=1}^N D^* \nabla_x \tilde{\Phi}_i(\bar{x},\bar{p})(v^*) = \left(C^\top v^*, 0\right) + N^{-1}\sum_{i=1}^N \left[A^i\right]^\top \partial_x^2 \Phi\left(\bar{x},\bar{\xi}^i\right)(v^*).$$

Consequently, we arrive at the assertion of our Proposition via the inclusion

$$\partial_x^2 \Psi(\bar{x},\bar{p})(v^*) \subseteq \left\{ \left(C^\top v^*, 0\right) + N^{-1}\sum\nolimits_{i=1}^N \left(x_i^*, B_i^\top p_i^*\right) \mid (x_i^*, p_i^*) \in \partial_x^2 \Phi\left(\bar{x},\bar{\xi}^i\right)(v^*) \quad (i=1,\ldots,N) \right\}.$$

■

After reducing $\partial_x^2 \Psi$ to $\partial_x^2 \Phi$ we are now faced with the computation of the latter. In order to do so, it will be convenient to write Φ in (4) as a composition

$$\Phi(x,\xi) = \theta(\alpha(x,\xi)), \quad \alpha(x,\xi) := h(\xi) - T(\xi)x, \quad \theta(v) := \sup_{W^\top z \le q} \langle v, z\rangle - \tfrac{1}{2}\langle z, Bz\rangle \tag{27}$$

Now, a chain rule for partial second-order subdifferentials recently proved by Mordukhovich and Rockafellar [13, Theorem 3.1] allows us to derive the following further reduction:

Lemma 4.1 *Let $\bar{x} \in \mathbb{R}^m$ and $\bar{\xi} \in \mathbb{R}^s$ be such that $T(\bar{\xi})$ is surjective. Then, for all $v^* \in \mathbb{R}^m$, it holds that*

$$\begin{aligned}\partial_x^2 \Phi\left(\bar{x},\bar{\xi}\right)(v^*) \quad &= \quad -\left(0, \nabla^\top \left\langle z\left(\alpha(\bar{x},\bar{\xi})\right), T(\cdot)\, v^*\right\rangle \left(\bar{\xi}\right)\right) \\ &\quad + \left(-T(\bar{\xi}), \nabla h\left(\bar{\xi}\right) - \nabla\left(T(\cdot)\,\bar{x}\right)\left(\bar{\xi}\right)\right)^\top \partial^2\theta\left(\alpha(\bar{x},\bar{\xi})\right)\left(-T(\bar{\xi})v^*\right),\end{aligned}$$

where $z(v)$ was introduced in Proposition 3.1.

Proof. The surjectivity of $\nabla_x \alpha(\bar{x},\bar{\xi}) = -T(\bar{\xi})$ allows us to apply the above-mentioned chain rule in order to derive that

$$\begin{aligned}\partial_x^2 \Phi\left(\bar{x},\bar{\xi}\right)(v^*) \quad &= \quad \left(\nabla_{xx}^2 \langle \bar{z}, \alpha\rangle \left(\bar{x},\bar{\xi}\right) v^*, \nabla_{x\xi}^2 \langle \bar{z}, \alpha\rangle \left(\bar{x},\bar{\xi}\right) v^*\right) + \\ &\quad \left(\nabla_x \alpha\left(\bar{x},\bar{\xi}\right), \nabla_\xi \alpha\left(\bar{x},\bar{\xi}\right)\right)^\top \partial^2\theta\left(\alpha\left(\bar{x},\bar{\xi}\right), \bar{z}\right)\left(\nabla_x \alpha\left(\bar{x},\bar{\xi}\right) v^*\right),\end{aligned}$$

where $\bar{z}$ is uniquely defined by the equation

$$\nabla_x \Phi\left(\bar{x},\bar{\xi}\right) = \left[\nabla_x \alpha\left(\bar{x},\bar{\xi}\right)\right]^\top \bar{z} = -T^\top(\bar{\xi})\bar{z}.$$

Hence, $\bar{z} = z\left(\alpha(\bar{x},\bar{\xi})\right)$, where $z(v)$ was introduced in Proposition 3.1 as unique element of (20). Since also

$$\nabla_x \Phi\left(\bar{x},\bar{\xi}\right) = -T^\top(\bar{\xi})\nabla\theta(\alpha\left(\bar{x},\bar{\xi}\right))$$

by (27), the injectivity of $-T^\top(\bar{\xi})$ yields that $\bar{z} = \nabla\theta(\alpha\left(\bar{x},\bar{\xi}\right))$ which allows us to omit the argument $\bar{z}$ in the expression $\partial^2\theta\left(\alpha\left(\bar{x},\bar{\xi}\right), \bar{z}\right)$. Taking into account that

$$\begin{aligned}\nabla_{xx}^2 \langle \bar{z}, \alpha\rangle \left(\bar{x},\bar{\xi}\right) v^* \quad &= \quad 0 \\ \nabla_{x\xi}^2 \langle \bar{z}, \alpha\rangle \left(\bar{x},\bar{\xi}\right) v^* \quad &= \quad -\left[\nabla \langle \bar{z}, T(\cdot)\, v^*\rangle \left(\bar{\xi}\right)\right]^\top \\ \nabla_\xi \alpha\left(\bar{x},\bar{\xi}\right) \quad &= \quad \nabla h\left(\bar{\xi}\right) - \nabla\left(T(\cdot)\,\bar{x}\right)\left(\bar{\xi}\right),\end{aligned}$$

we arrive at the asserted formula.

■

Now, it remains to provide an explicit formula for the second order subdifferential $\partial^2\theta$. Before doing so, we recall the following

Proposition 4.2 *[5, Corollary 3.5] Consider a polyhedron $P := \{u|Au \leq b\}$. Fix arbitrary $\bar{u} \in P$ and $\bar{w} \in N_P(\bar{u})$. Denote by $I := \{i|\,\langle a_i, \bar{u}\rangle = b_i\}$ the index set of active rows of A at $\bar{u}$. Assume that these active rows are linearly independent. Moreoever, let $J := \{i \in I|\lambda_i > 0\}$ be the index set of strictly positive multipliers, where λ is the unique solution of $\sum_{i\in I} \lambda_i a_i = \bar{w}$. Then,*

$$D^*N_P(\bar{u},\bar{w})(s^*) = \begin{cases} \text{pos}\,\{a_i|i \in I : \langle a_i, s^*\rangle > 0\} + \text{span}\,\{a_i|i \in I : \langle a_i, s^*\rangle = 0\} & \text{if } s^* \in \bigcap\limits_{i\in J} a_i^\perp \\ \emptyset & \text{else} \end{cases}.$$

Here, 'pos' and 'span' refer to the convex cone and linear subspace, respectively, generated by the elements in the corresponding set.

Proposition 4.3 *For any $\bar{v}, w^* \in \mathbb{R}^r$, the second-order subdifferential of the function θ in (27) calculates as*

$$\begin{aligned} \partial^2\theta(\bar{v})(w^*) &= \{z^*|Bz^* - w^* \in D^*N_Z(z(\bar{v}), \bar{v} - B\,z(\bar{v}))(-z^*)\} \\ &= \begin{cases} \{z^*|Bz^* - w^* \in \text{pos}\,\{w_i|i \in I : \langle w_i, z^*\rangle < 0\} + \text{span}\,\{w_i|i \in I : \langle w_i, z^*\rangle = 0\}\} & \text{if } z^* \in \bigcap\limits_{i\in J} w_i^\perp \\ \emptyset & \text{else} \end{cases} \end{aligned}$$

where $z(\bar{v})$ refers to the unique element of (20) and - with respect to the notation introduced in (5) - the w_i represent the columns of the matrix W. Moreover $I := \{i|\,\langle w_i, z(\bar{v})\rangle = q_i\}$ is the index set of active rows of $W^\top$ at $z(\bar{v})$ and $J := \{i \in I|\lambda_i > 0\}$ is the index set of strictly positive multipliers, where λ denotes the unique solution of $\sum_{i\in I} \lambda_i w_i = \bar{v} - B\,z(\bar{v})$.

Proof. Given the definition of θ in (27) and applying Proposition 3.1 to the special case $h(\xi) = 0$ and $T(\xi) = -I$ for all ξ, we see that θ is strictly differentiable with $\nabla\theta(v)$ being the unique element of (20), i.e., $\nabla\theta(v) = z(v)$. Moreover, $\nabla\theta$ is locally Lipschitz. With Z defined in (5), we deduce from (20) the equivalence

$$(v, z) \in \text{gr}\,\nabla\theta \iff v - Bz \in N_Z(z) \iff (z, v - Bz) \in \text{gr}\,N_Z.$$

Hence $\text{gr}\,\nabla\theta = L^{-1}\text{gr}\,N_Z$, where $L(v, z) = (z, v - Bz)$ is a surjective linear mapping. Then, recalling the symmetry of B, [18, Exercise 6.7] yields that

$$N_{\text{gr}\,\nabla\theta}(\bar{v}, \nabla\theta(\bar{v})) = \begin{pmatrix} 0 & I \\ I & -B \end{pmatrix} N_{\text{gr}\,N_Z}(\nabla\theta(\bar{v}), \bar{v} - B\,\nabla\theta(\bar{v})).$$

Exploiting the corresponding definitions, this last relation entails the first equality asserted in this proposition. Now, with Z defined in (5) satisfying the Linear Independence Constraint Qualification (see basic assumptions imposed at the beginning of Section 3), the assertion of the proposition follows immediately from Proposition 4.2.

■

5 An upper estimate for the condition number

5.1 An upper estimate for D^*S

Collecting the results of Proposition 3.3, Proposition 4.1 and Lemma 4.1, we arrive at the following upper estimate for the coderivative of the solution mapping S in (18):

Theorem 5.1 *Let $(\bar{p},\bar{x})\in\operatorname{gr}S$, where $\bar{x}\in X$ and $\bar{p}:=\left(\bar{\xi}^1,\ldots,\bar{\xi}^N\right)\in\mathbb{R}^{Ns}$. Assume that the multifunction (23) is calm at $(0,\bar{p},\bar{x})$ and that the matrices $T(\bar{\xi}^i)$ are surjective for $i=1,\ldots,N$. Then,*

$$D^*S\left(\bar{p},\bar{x}\right)(x^*)\subseteq\left\{p^*\;\middle|\;\begin{array}{l}\exists v^*\exists u^*\in D^*N_X(\bar{x},-\nabla_x\Psi\left(\bar{x},\bar{p}\right))\left(v^*\right),\\ \exists z_i^*\in\partial^2\theta\left(h(\bar{\xi}^i)-T(\bar{\xi}^i)\bar{x}\right)\left(-T(\bar{\xi}^i)v^*\right)\quad(i=1,\ldots,N):\\ N^{-1}\displaystyle\sum_{i=1}^N\left[T(\bar{\xi}^i)\right]^\top z_i^*=C^\top v^*+x^*+u^*,\\ p_i^*=N^{-1}\Big(-\nabla^\top\langle\bar{z}_i,T(\cdot)v^*\rangle(\bar{\xi}^i)+\left[\nabla h\left(\bar{\xi}^i\right)-\nabla(T(\cdot)\bar{x})(\bar{\xi}^i)\right]^\top z_i^*\Big)\\ \hfill(i=1,\ldots,N),\end{array}\right\}$$

where the $\bar{z}_i$ are the unique elements of

$$\operatorname*{argmax}_{W^\top z\leq q}\left\langle h(\bar{\xi}^i)-T(\bar{\xi}^i)\bar{x},z\right\rangle-\tfrac{1}{2}\left\langle z,Bz\right\rangle\quad(i=1,\ldots,N).$$

In the following Proposition we provide an instance under which the calmness assumption of the previous Theorem is satisfied:

Proposition 5.1 *If T is a constant mapping, i.e. $T(\xi)\equiv T$, and h is an affine linear mapping, i.e. $h(\xi)=A\xi+b$, then the calmness condition of Proposition 3.3 is satisfied.*

Proof. Putting $z=\left(z^1,\ldots,z^N\right)$ and, as before, $p=\left(\xi^1,\ldots,\xi^N\right)$, we introduce the sets

$$\begin{aligned}\Lambda_1\quad&:=\left\{(y,p,x,z)\,|\left(x,y-c-Cx-N^{-1}T^\top\sum\nolimits_{i=1}^N z^i\right)\in\operatorname{gr}N_X\right\}\\ \Lambda_2^i\quad&:=\left\{(y,p,x,z)\,|\left(A\xi^i+b-Tx,z^i\right)\in\operatorname{gr}\nabla\theta\right\}\quad(i=1,\ldots,N).\end{aligned}$$

Then, with M defined in (23), one has that $\operatorname{gr}M=\pi\left(\Lambda_1\cap\Lambda_2^1\cap\cdots\cap\Lambda_2^N\right)$, where π denotes the projection onto the first three coordinates. Indeed, by definition of M and by Corollary 3.1,

$$(y,p,x)\in\operatorname{gr}M\iff y-c-Cx-N^{-1}\sum_{i=1}^N\nabla_x\Phi(x,\xi^i)\in N_X(x).$$

Since $\nabla_x\Phi(x,\xi^i)=-T^\top\nabla\theta\left(h\left(\xi^i\right)-Tx\right)$ for $i=1,\ldots,N$ by (27), it follows that

$$(y,p,x)\in\operatorname{gr}M\iff\exists z:(y,p,x,z)\in\Lambda_1\cap\Lambda_2^1\cap\cdots\cap\Lambda_2^N,$$

which amounts to the asserted identity. Now, the graph of the normal cone mapping to a polyhedron such as $\operatorname{gr}N_X$ can be represented as a finite union of polyhedra. Hence Λ_1 as a preimage of such a set under an affine linear mapping is a finite union of polyhedra itself. Moreover, with the same argument, the relation $\operatorname{gr}\nabla\theta=L^{-1}\operatorname{gr}N_Z$ used in the proof of Proposition 4.3 reveals that $\operatorname{gr}\nabla\theta$ too is a finite union of polyhedra and, hence, so are the sets $\Lambda_2^1,\ldots,\Lambda_2^N$ as preimages of $\operatorname{gr}\nabla\theta$ under affine linear mappings. It follows that the intersection $\Lambda_1\cap\Lambda_2^1\cap\cdots\cap\Lambda_2^N$ is also a finite union of polyhedra. Consequently, $\operatorname{gr}M$ is a finite union of polyhedra (recall that the projection of a polyhedron is a polyhedron). Now, calmness (actually: upper Lipschitz continuity) of M at any point of its graph is a result of Robinson's Theorem [15].

■

Combining Proposition 5.1 with Theorem 5.1 and Proposition 4.3, we may draw the following conclusion for a simplified setting:

Corollary 5.1 *Let $(\bar{p},\bar{x}) \in \operatorname{gr} S$, where $\bar{x} \in X$ and $\bar{p} := \left(\bar{\xi}^1,\ldots,\bar{\xi}^N\right) \in \mathbb{R}^{Ns}$. Assume that $T(\xi) \equiv T$, and $h(\xi) = A\xi + b$. Moreover, let T be surjective. Then,*

$$D^*S(\bar{p},\bar{x})(x^*) \subseteq \left\{ p^* \left| \begin{array}{l} \exists v^* \, \exists u^* \in D^*N_X(\bar{x}, -c - C\bar{x} + N^{-1}T^\top \sum_{i=1}^N z(\bar{v}_i))(v^*) \\ \exists z_i^* : Bz_i^* + Tv^* \in D^*N_Z(z(\bar{v}_i), \bar{v}_i - Bz(\bar{v}_i))(-z_i^*) \quad (i = 1,\ldots,N) \\ N^{-1}T^\top \sum\limits_{i=1}^N z_i^* = C^\top v^* + x^* + u^* \\ p_i^* = N^{-1}A^\top z_i^*, \, \bar{v}_i = A\bar{\xi}^i + b - T\bar{x} \quad (i = 1,\ldots,N) \end{array} \right. \right\} \tag{28}$$

where $z(v)$ is defined in (20).

Hence, $D^*S(\bar{p},\bar{x})(x^*)$ is contained in a set which is given in terms of the data of the stochastic program and of the Mordukhovich coderivative of the normal cone mappings to the polyhedra X and Z, respectively. The latter may be computed by Proposition 4.2.

5.2 Application to conditioning in the case of simple recourse

We apply the result of the previous section to the special setting of so-called *simple recourse*. More precisely, we assume that our two-stage stochastic optimization problem has the following (primal) form:

$$\min_{x \in X} \langle c, x\rangle + \frac{\sigma}{2}\|x\|^2 + N^{-1}\sum_{i=1}^N \Phi(x,\xi^i),$$

where $\xi^i \in \mathbb{R}^s$ $(i = 1,\ldots,N)$ are realizations of the random vector ξ and where

$$\begin{array}{rl} X := & \{x \in \mathbb{R}^m | Dx \le f\} \\ \Phi(x,\xi) := & \sup\limits_{-q^- \le z \le q^+} \langle A\xi + b - Tx, z\rangle - \dfrac{\tau}{2}\|z\|^2 . \end{array}$$

We assume that $\tau, \sigma > 0$. Note that this problem differs from a standard problem of simple recourse type as much as our general problem (9) differs from a standard two-stage problem by admitting violation of recourse at the cost of a penalty. The reason to use the term 'simple recourse' here, is the rectangular shape of the set Z in (5). Clearly, this problem fits the model (9) with

$$q := \left(q^+, q^-\right), \quad W := (I| - I), \quad B := \tau I, \quad C := \sigma I, \quad h(\xi) := A\xi + b, \quad T(\xi) \equiv T. \tag{29}$$

in (6). As mentioned in the introduction, the matrix $B^{-1} = \tau^{-1}I$ induces a penalty on violating the constraint (7), hence we may interpret τ^{-1} as a penalty parameter. We assume that the second stage costs are strictly positive ($q_j^+, q_j^- > 0$ for all j) implying that the rectangle $[-q^-, q^+]$ satisfies our basic assumption of nondegeneracy imposed on the polyhedron Z in the beginning of section 3. Our first observation relates to the elements z_i^* in (28):

Lemma 5.1 *Let T be surjective and let $\xi, z^* \in \mathbb{R}^r$, $x, v^* \in \mathbb{R}^m$ be such that*

$$Bz^* + Tv^* \in D^*N_Z(z(v), v - B\,z(v))(-z^*).$$

Here, $v := A\xi + b - Tx$ and $z(v)$ is the unique element of (20). Then,

$$\begin{array}{ll} \left|z_j^*\right| \le \tau^{-1}\|t_j\|\,\|v^*\| & \text{if } j \in \{1,\ldots,r\} \\ z_j^* = 0 & \text{if } j \in J_1 \cup J_2 \end{array},$$

where, t_j denotes the jth row of T and

$$\begin{array}{rcl} J_1 & := & \left\{j \in \{1,\ldots,r\}\,|\,z_j(v) = q_j^+, \langle a_j,\xi\rangle + b_j - \langle t_j, x\rangle > \tau q_j^+\right\} \\ J_2 & := & \left\{j \in \{1,\ldots,r\}\,|\,z_j(v) = -q_j^-, \langle a_j,\xi\rangle + b_j - \langle t_j, x\rangle < -\tau q_j^-\right\}, \end{array}$$

with a_j referring to the jth row of A.

Proof. Specifying the matrix W in Proposition 4.3 to our setting, we have that its columns are given by $w_j = e_j$ and $w_{j+r} = -e_j$ for $j = 1, \ldots, r$, where e_j refers to the jth canonical vector in $\mathbb{R}^r$. Therefore, the index set I introduced in Proposition 4.3 takes in our setting the form

$$I = \left\{ j \in \{1, \ldots, r\} \,|z_j (v) = q_j^+ \right\} \cup \left\{ j \in \{r+1, \ldots, 2r\} \,|z_{j-r} (v) = -q_{j-r}^- \right\}.$$

Similarly, the index set J introduced in Proposition 4.3 takes the form

$$J = \{ j \in I | \lambda_j > 0 \},$$

where λ is the unique solution of

$$\sum_{j \in I \cap \{1, \ldots, r\}} \lambda_j e_j - \sum_{j \in I \cap \{r+1, \ldots, 2r\}} \lambda_j e_{j-r} = v - B\, z(v). \tag{30}$$

Observe that one cannot have $j \in I$ and $j + r \in I$ simultaneously for the same index $j \in \{1, \ldots, r\}$ due to $q_j^+ > 0 > -q_{j-r}^-$. Consequently, recalling that $B = \tau I$, (30) yields

$$\begin{array}{ll} \lambda_j = v_j - \tau z_j (v) = v_j - \tau q_j^+ & \text{if } j \in I \cap \{1, \ldots, r\} \\ -\lambda_j = v_{j-r} - \tau z_{j-r} (v) = v_{j-r} + \tau q_{j-r}^- & \text{if } j \in I \cap \{r+1, \ldots, 2r\} \end{array}.$$

It follows that

$$\begin{array}{rcl} J & = & \left\{ j \in \{1, \ldots, r\} \,|z_j (v) = q_j^+, \; \langle a_j, \xi \rangle + b_j - \langle t_j, x \rangle > \tau q_j^+ \right\} \cup \\ & & \left\{ j \in \{r+1, \ldots, 2r\} \,|z_{j-r} (v) = -q_{j-r}^-, \; \langle a_{j-r}, \xi \rangle + b_{j-r} - \langle t_{j-r}, x \rangle < -\tau q_{j-r}^- \right\} \end{array}$$

Now, by Proposition 4.3, $\langle z^*, w_j \rangle = 0$ for all $j \in J$. With respect to the index sets J_1, J_2 introduced in the statement of this Lemma, the following holds true: If $j \in J_1$, then j belongs to the first set in the union above, hence $j \in J$. Then, $0 = \langle z^*, w_j \rangle = z_j^*$. Similarly, if $j \in J_2$, then $j + r$ belongs to the second set in the union above, hence $j + r \in J$. Then, $0 = \langle z^*, w_{j+r} \rangle = -z_j^*$. This proves the second statement in the assertion of this Lemma. Next, let $j \in \{1, \ldots, r\}$ be arbitrary. The relation $Bz^* + Tv^* \in D^* N_Z (z(v), v - B\, z(v)) (-z^*)$ translates by Proposition 4.3 in our setting to

$$\tau z^* + Tv^* \in \text{pos } \{ w_j | j \in I : \langle w_j, z^* \rangle < 0 \} + \text{span } \{ w_j | j \in I : \langle w_j, z^* \rangle = 0 \}$$

or to

$$\tau z^* + Tv^* = \sum_{k \le r, k \in I, z_k^* < 0} \lambda_k^a e_k - \sum_{k \le r, k+r \in I, z_k^* > 0} \lambda_k^b e_k + \sum_{k \le r, k \in I, z_k^* = 0} \mu_k^a e_k + \sum_{k \le r, k+r \in I, z_k^* = 0} \mu_k^b e_k \tag{31}$$

for certain coefficients $\lambda_k^a, \lambda_k^b \ge 0$ and $\mu_k^a, \mu_k^b \in \mathbb{R}$. Now, if $z_j^* = 0$, then the estimate in the first statement in the assertion of our Lemma is trivially satisfied. Otherwise, if $z_j^* \ne 0$, then by (31),

$$\tau z_j^* + \langle t_j, v^* \rangle = \begin{cases} \lambda_j^a \ge 0 & \text{if } j \in I, z_j^* < 0 \\ -\lambda_j^b \le 0 & \text{if } j + r \in I, z_j^* > 0 \\ 0 & \text{else} \end{cases}.$$

In the first case, one has that $0 > z_j^* \ge -\tau^{-1} \langle t_j, v^* \rangle$ which directly implies the asserted estimate $\left| z_j^* \right| \le \tau^{-1} \|t_j\| \|v^*\|$. The second case follows analogously. The third case is evident as well. This proves the first statement in the assertion of this Lemma.

■

Observe that the index sets J_1, J_2 introduced in Lemma 5.1 represent those components j of the solution $z(v)$ of problem (20) for $v := A\xi + b - Tx$ which are strongly active (i.e., which are active with respect to the constraints $-q^- \le z \le q^+$ and for which the associated Lagrange multiplier is strictly positive). This Lemma eventually allows us to calculate an upper estimate for the condition number in case of simple recourse. To this aim, we fix some $\bar{x} \in X$ and $\bar{p} := \left(\bar{\xi}^1, \ldots, \bar{\xi}^N \right) \in \mathbb{R}^{Ns}$ such that $\bar{x} \in S(\bar{p})$, i.e., $0 \in \nabla_x \Psi(\bar{x}, \bar{p}) + N_X(\bar{x})$ for Ψ defined in (17). With d_i referring to the rows of D in the description $Dx \leq f$ of the polyhedron X, this implies that

$$\nabla_x \Psi(\bar{x}, \bar{p}) = \sum\nolimits_{i \in \tilde{I}} \lambda_i d_i \quad \left(\tilde{I} := \{i | \langle d_i, \bar{x} \rangle = f_i\}\right) \tag{32}$$

for certain $\lambda_i \leq 0$ $\left(i \in \tilde{I}\right)$. For each $i = 1, \ldots, N$ we put $\bar{v}_i := A\bar{\xi}^i + b - T\bar{x}$ and introduce the index sets

$$\begin{aligned} J_1(i) \quad &: \quad = \left\{j \in \{1, \ldots, r\} \,| z_j\left(\bar{v}^i\right) = q_j^+, \langle a_j, \bar{\xi}^i \rangle + b_j - \langle t_j, \bar{x} \rangle > \tau q_j^+ \right\} \\ J_2(i) \quad &: \quad = \left\{j \in \{1, \ldots, r\} \,| z_j\left(\bar{v}^i\right) = -q_j^-, \langle a_j, \bar{\xi}^i \rangle + b_j - \langle t_j, \bar{x} \rangle < -\tau q_j^- \right\}, \end{aligned}$$

i.e., the same index sets characterizing strongly active components in the solution of problem (20) as in Lemma 5.1 but now related to the different scenarios $\bar{\xi}^i$. This allows us to define the following quantity

$$\Delta(T) := \sum\nolimits_{i=1}^{N} \Delta_i(T), \quad \Delta_i(T) := \left(\sum_{j \in \{1, \ldots, r\} \setminus (J_1(i) \cup J_2(i))} \|t_j\|^2 \right)^{1/2} \quad (i = 1, \ldots, N).$$

Observe that $\Delta(T)$ increases not only with increasing elements of the matrix T but also with decreasing number of strongly active components in the scenario-dependent solutions $z(\bar{v}_i)$ of the problems

$$\max_{-q^- \leq z \leq q^+} \langle \bar{v}_i, z \rangle - \frac{\tau}{2} \|z\|^2. \tag{33}$$

Clearly, $0 \leq \Delta_i(T) \leq \|T\|_F$, where $\|\cdot\|_F$ refers to the Frobenius norm. Here, the minimum is attained if all components of $z(\bar{v}_i)$ are strongly active (i.e., $z(\bar{v}_i)$ equals a corner of the rectangle $[-q^-, q^+]$ and all Lagrange multipliers are strictly positive). In contrast, the maximum is attained if no component is strongly active (e.g., $z(\bar{v}_i)$ lies in the interior of the rectangle $[-q^-, q^+]$ or it lies on the boundary of this rectangle but all Lagrange multipliers equal zero). We have the following upper estimate for the condition number:

Theorem 5.2 *In the setting specified above, assume that even* $\lambda_i < 0$ $\left(i \in \tilde{I}\right)$ *in (32) , i.e., strict complementarity holds at* $\bar{x}$*. Moreover, let* T *be surjective. Finally, let the parameters* σ *and* τ *in (29) satisfy the relation*

$$\tau\sigma > N^{-1} \|T\| \Delta(T). \tag{34}$$

Then, the condition number $\operatorname{lip} S(\bar{p}, \bar{x})$ *as introduced in (19), can be estimated by*

$$\operatorname{lip} S(\bar{p}, \bar{x}) \leq \frac{\|A\|}{\left([\Delta(T)]^{-1} N\sigma\tau - \|T\|\right)}.$$

Proof. In order to estimate $\operatorname{lip} S(\bar{p}, \bar{x})$, fix an arbitrary x^* with $\|x^*\| \leq 1$ and an arbitrary $p^* \in D^*S(\bar{p}, \bar{x})(x^*)$. Our assumptions allow us to apply Corollary 5.1. Accordingly, there exist u^*, v^* and z_i^* satisfying the relations in (28). In particular, $u^* \in D^*N_X\left(\bar{x}, -c - C\bar{x} + N^{-1}T^\top \sum_{i=1}^N z(\bar{v}_i)\right)(v^*)$. The assumption of strict complementarity yields that $v^* \in \operatorname{Ker} D_I$ and $u^* \in \operatorname{Im} D_I^\top$, where D_I is the reduction of D to its active rows (see, e.g., [5, Corollary 3.7]). This entails that $\langle u^*, v^* \rangle = 0$ which may be exploited in order to reduce the first equation in (28) to

$$N^{-1}T^\top \sum_{i=1}^{N} \langle z_i^*, v^* \rangle = \sigma \|v^*\|^2 + \langle x^*, v^* \rangle,$$

where z_i^* is such that

$$Bz_i^* + Tv^* \in D^*N_Z(z(\bar{v}_i), \bar{v}_i - Bz(\bar{v}_i))(-z_i^*) \quad (i = 1, \ldots, N).$$

From here, we get the estimate

$$\sigma \|v^*\| \le 1 + N^{-1} \|T\| \sum_{i=1}^{N} \|z_i^*\| . \tag{35}$$

Now, for each such z_i^* with components $z_{i,j}^*$ we have by Lemma 5.1, that

$$\|z_i^*\|^2 = \sum_{j\in\{1,\dots,r\}\setminus(J_1(i)\cup J_2(i))} \left(z_{i,j}^*\right)^2 \le \tau^{-2} \|v^*\|^2 \sum_{j\in\{1,\dots,r\}\setminus(J_1(i)\cup J_2(i))} \|t_j\|^2 ,$$

whence, with $\Delta_i(T)$ as introduced in the statement of this Theorem,

$$\|z_i^*\| \le \tau^{-1} \|v^*\| \Delta_i(T) \quad \text{and} \quad \sum_{i=1}^{N} \|z_i^*\| \le \tau^{-1} \|v^*\| \sum\nolimits_{i=1}^{N} \Delta_i(T) = \tau^{-1} \|v^*\| \Delta(T).$$

Combining this with (35) leads along with (34) to $\|v^*\| \le \left(\sigma - N^{-1}\tau^{-1} \|T\| \Delta(T)\right)^{-1}$. Now, the second equation in (28) may be exploited to derive

$$\|p_i^*\| \le N^{-1} \|A\| \|z_i^*\| \le N^{-1}\tau^{-1}\Delta_i(T) \|A\| \|v^*\| \le N^{-1}\tau^{-1}\Delta_i(T) \|A\| \left(\sigma - N^{-1}\tau^{-1} \|T\| \Delta(T)\right)^{-1} .$$

Hence,

$$\|p^*\| = \left(\sum_{i=1}^{N} \|p_i^*\|^2\right)^{1/2} \le \frac{\|A\|}{(N\sigma\tau - \|T\| \Delta(T))} \left(\sum_{i=1}^{N} \Delta_i^2(T)\right)^{1/2} \le \frac{\|A\| \Delta(T)}{(N\sigma\tau - \|T\| \Delta(T))}.$$

Since x^* with $\|x^*\| \le 1$ and $p^* \in D^*S\left(\bar p, \bar x\right)(x^*)$ were arbitrarily chosen, the asserted estimate for the condition number follows. ■

The result of the Theorem can be roughly interpreted as follows: the condition number decreases with σ but increases with the norms $\|T\|, \|A\|$, with the penalty parameter τ^{-1} and with $\Delta(T)$ (i.e., with a decreasing number of strongly active components in the solutions of problems (33)). At the first glance one might have the impression that the condition number decreases also with an increasing number N of scenarios. One has to take into account, however, that the quantity $\Delta(T)$ itself depends on N (the number of terms in the sum), hence it is a better idea to interpret the expression $[\Delta(T)]^{-1} N = [\Delta(T)/N]^{-1}$ as a mean number of non strongly active components in the solutions of problems (33).

Acknowledgment

The authors would like to thank three anonymous referees for their valuable comments.

Appendix

Equivalence between (4) and (6): We consider the second-stage costs as given in (4) with $Z = \{z \in \mathbb{R}^k : W^\top z \le q\}$. From [18, Example 11.43], one derives by duality that

$$\begin{aligned}\Phi(x,\xi) &= \sup_{W^\top z \le q} \left\{\langle h(\xi) - T(\xi)x, z\rangle - \tfrac{1}{2}\langle z, Bz\rangle\right\} \\ &= \sup_{z} \left\{\langle h(\xi) - T(\xi)x, v\rangle - \tfrac{1}{2}\langle z, Bz\rangle - \left\{\sup\nolimits_{y\ge 0} \left\langle W^\top z - q, y\right\rangle\right\}\right\}\end{aligned}$$

Consequently, we may rewrite $\Phi(x,\xi)$ as

$$\Phi(x,\xi) = \inf_{y\ge 0} \left\{\langle q, y\rangle + \sup_{z} \left\{\langle h(\xi) - T(\xi)x - Wy, z\rangle - \tfrac{1}{2}\langle z, Bz\rangle\right\}\right\}.$$

If one assumes that B is positive definite, it follows that

$$\sup_z \left\{ \langle h(\xi) - T\xi)x - Wy, v\rangle - \tfrac{1}{2}\langle z, Bz\rangle \right\} = \tfrac{1}{2}\left\langle h(\xi) - T(\xi)x - Wy, B^{-1}\left(h(\xi) - T(\xi)x - Wy\right)\right\rangle$$

and, hence,

$$\Phi(x,\xi) = \inf_{y\geq 0} \left\{ \langle q, y\rangle + \tfrac{1}{2}\left\langle h(\xi) - T(\xi)x - Wy, B^{-1}(h(\xi) - T(\xi)x - Wy)\right\rangle \right\}.$$

References

[1] J.F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer, New York, 2000.

[2] A.L. Dontchev and R.T. Rockafellar, Regularity and conditioning of solution mappings in variational analysis, Set-Valued Analysis 12 (2004), 79–109.

[3] R.M. Freund and J.R. Vera, Some characterizations and properties of the distance to ill-posedness and the condition measure of a conic linear system, Math. Program. 86 (1999), 225–260.

[4] R. Henrion, A. Jourani and J. Outrata, On the calmness of a class of multifunctions, SIAM Journal on Optimization 13 (2002), 603-618.

[5] R. Henrion and W. Römisch, On M-stationary points for a stochastic equilibrium problem under equilibrium constraints in electricity spot market modeling, Applications of Mathematics 52 (2007), 473–494.

[6] D. Klatte, Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs, Math. Program. 88 (2000), 285–311.

[7] B. Kummer, Lipschitzian and pseudo-Lipschitzian inverse functions and applications to nonlinear programming, in: A.V. Fiacco (ed.), Mathematical Programming with Data Perturbations, pp. 201222, Marcel Dekker, New York, 1997.

[8] Wu Li, The sharp Lipschitz constants for feasible and optimal solutions of a perturbed linear program, Linear Algebra Appl. 187 (1993), 15–40.

[9] Wu Li, Error bounds for piecewise convex quadratic programs and applications, SIAM J. Control Optim. 33 (1995), 1510–1529.

[10] O.L. Mangasarian, A condition number for differentiable convex inequalities, Math. Oper. Res. 10 (1985), 175-179.

[11] B.S. Mordukhovich, Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions, Trans. Amer. Math. Soc. 340 (1993), 1–35.

[12] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation. Vol. 1: Basic Theory, Springer, Berlin, 2006.

[13] B.S. Mordukhovich and R.T. Rockafellar, Second-Order subdifferential calculus with applications to tilt stability in optimization, SIAM J. Optim. 22 (2012), 953–986.

[14] J. Renegar, Condition numbers, the barrier method, and the conjugate-gradient method, SIAM J. Optim. 6 (1996), 879–912.

[15] S.M. Robinson, Some continuity properties of polyhedral multifunctions, Mathematical Programming Studies, 14, (1976) 206-214.

[16] R.T. Rockafellar, Linear-quadratic programming and optimal control, SIAM J. Control Optim. 25 (1987), 781–814.

[17] R.T. Rockafellar and R. J.-B. Wets, A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming, Math. Program. Study 28 (1986), 63–93.

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

[19] W. Römisch: Stability of stochastic programming problems, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.

[20] A. Shapiro, Perturbation analysis of optimization problems in Banach spaces, Numer. Funct. Anal. Optim. 13 (1992), 97–116.

[21] A. Shapiro, T. Homem-de-Mello and J. Kim, Conditioning of convex piecewise linear stochastic programs, Math. Program. 94 (2002), 1–19.

[22] T. Surowiec, Explicit Stationarity Conditions and Solution Characterization for Equilibrium Problems with Equilibrium Constraints, PhD thesis, Humboldt University Berlin, 2010.

[23] M.H. Wright, Ill-conditioning and computational error in interior methods for nonlinear programming, SIAM J. Optim. 9 (1999), 84–111.