

Mixed-Integer Value Functions in Stochastic Programming

Rüdiger Schultz

Institute of Mathematics, Gerhard-Mercator University Duisburg
Lotharstr. 65, D-47048 Duisburg, Germany

Abstract

We discuss the role of mixed-integer value functions in the theoretical analysis of stochastic integer programs. It is shown how the interaction of value function properties with basic results from probability theory leads to structural statements in stochastic integer programming.

Key Words. Stochastic programming, mixed-integer optimization.

AMS subject classifications. 90C15, 90C11, 90C06.

1 Stochastic Integer Programs

Stochastic programming models arise as deterministic equivalents to random optimization problems. In the present paper we confine ourselves to linear two-stage models involving integer requirements. The random optimization problem behind these models reads as follows

$$\min_{x,y,y'} \{c^T x + q^T y + q'^T y' : Tx + Wy + W'y' = h(\omega), x \in X, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'}\}. \quad (1)$$

We assume that all ingredients above have conformal dimensions, that W, W' are rational matrices, and that $X \subseteq \mathbb{R}^m$ is a nonempty closed polyhedron, possibly involving integrality constraints on components of the vector x .

Together with (1) we have a scheme of alternating decision and observation: The decision on x has to be made prior to observing the outcome of the random vector $h(\omega)$, and the vector (y, y') is selected only after having decided on x and observed $h(\omega)$. This setting corresponds to a variety of practical optimization problems under uncertainty. It readily extends to the multi-stage situation where finitely (or even infinitely) many of the above alternations occur, see [6, 12, 17] for further details on stochastic programming modelling.

As a mathematical object, problem (1) is ill-posed, since at the moment of decision on x it is not clear which vectors x are feasible, let alone optimal. As a remedy, let us proceed as follows. Rewrite (1) by separating the optimizations in x and (y, y') :

$$\min_x \left\{ c^T x + \min_{y,y'} \{ q^T y + q'^T y' : Wy + W'y' = h(\omega) - Tx, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'} \} : x \in X \right\}. \quad (2)$$

This is where the mixed-integer value function enters the scene. Indeed, in (2) we have an inner optimization problem with right-hand side parameter $h(\omega) - Tx$. Introducing the mixed-integer value function

$$\Phi(t) := \min \{ q^T y + q'^T y' : Wy + W'y' = t, y \in \mathbb{Z}_+^{\bar{m}}, y' \in \mathbb{R}_+^{m'} \}, \quad (3)$$

(2) turns into

$$\min_x \{ c^T x + \Phi(h(\omega) - Tx) : x \in X \}. \quad (4)$$

In this way, we obtain the family $f(x, \omega) := c^T x + \Phi(h(\omega) - Tx)$, $x \in X$ of real-valued random variables. The problem is still ill-posed, since in (4) the meaning of “ \min_x ” remains unclear, i.e., it is still open how to select a “best” x . In stochastic programming, scalar parameters of the random variables $f(x, \omega)$, $x \in X$ provide criteria for making the “best” selection.

The most widely used scalar parameter in this respect is the expectation. Assuming that $h(\omega) \in \mathbb{R}^s$ is a random vector on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ we obtain the well-posed optimization problem

$$\min \left\{ \int_{\Omega} (c^T x + \Phi(h(\omega) - Tx)) \mathbb{P}(d\omega) : x \in X \right\}. \quad (5)$$

In terms of the random optimization problem (1), the model (5) suggests to select, before observing $h(\omega)$, i.e., in a “here-and-now” manner, a decision x such that the expected value of the random costs $c^T x + \Phi(h(\omega) - Tx)$ becomes minimal.

When addressing risk aversion, other scalar parameters are useful. In the context of stochastic programming some first proposals have been made in [15, 16, 22]. Following the probability-based approach in [22] we introduce some threshold level $\varphi_o \in \mathbb{R}$ and consider minimization of the probability that the random costs $c^T x + \Phi(h(\omega) - Tx)$ exceed φ_o . This leads to the optimization model

$$\min \left\{ \mathbb{P}(\{\omega \in \Omega : c^T x + \Phi(h(\omega) - Tx) > \varphi_o\}) : x \in X \right\}. \quad (6)$$

As mathematical objects, (5) and (6) are optimization problems in x whose objectives we denote by $Q_{\mathbb{E}}(x)$ and $Q_{\mathbb{P}}(x)$, respectively. It is evident, that the mixed-integer value function Φ essentially determines the structure of the functions $Q_{\mathbb{E}}$ and $Q_{\mathbb{P}}$. As we will see, there is a fruitful interaction of properties of Φ with basic statements from probability theory.

The article is organized as follows. In Section 2 we report on what is known about the mixed-integer value function Φ . Section 3 aims at putting together value function properties with basic probability theory. Proceeding step by step, we draw conclusions from various convergence results of probability theory. The final section is an outlook towards related issues beyond the scope of the present paper.

2 Mixed-Integer Value Functions

Studying the value function Φ is part of what is usually referred to as stability analysis of optimization problems or parametric optimization. In this area of research the accent is on properties of optimal values and optimal solution sets seen as (multi-)functions of parameters arising in the underlying optimization problems. A variety of results starting from linear programs and leading into nonlinear programming and optimal control is available, see the recent monograph [8] and references therein.

In the above stochastic programming context, the value function appears as an integrand of a suitable integral. Therefore, it is crucial to have *global* knowledge about the functional dependence of the optimal value on the respective parameter. The typical situation in *nonlinear* parametric optimization, however, is that properties of optimal values and optimal solutions are available only *locally* around given parameters. The most comprehensible class for which global results exist are mixed-integer linear programs. This explains that, so far, the models discussed in Section 1 do not go beyond the mixed-integer linear case. The stability of mixed-integer linear programs has been studied in a series of papers by Blair and Jeroslow out of which we refer to [7], and in the monographs [2, 3].

Before discussing the mixed-integer value function Φ from (3), let us have a quick look at its integer-free counterpart:

$$\Phi_{lin}(t) := \min\{q'^T y' : W' y' = t, y' \in \mathbb{R}_+^{m'}\}. \quad (7)$$

If we assume that $W'(R_+^{m'})$ is full-dimensional and that $\{u \in \mathbb{R}^s : W'^T u \leq q'\} \neq \emptyset$, then the latter set has vertices d_k , $k = 1, \dots, K$, and it holds by linear programming duality that

$$\Phi_{lin}(t) = \max\{t^T u : W'^T u \leq q'\} = \max_{k=1, \dots, K} d_k^T t \quad \text{for all } t \in W'(R_+^{m'}).$$

Hence, Φ_{lin} is convex and piecewise linear on its (conical) domain of definition. Without going into details, we mention that this convexity has far reaching consequences when setting up the expectation based stochastic programming model (5) in case integer requirements are missing in the random optimization problem (1), see [6, 12, 17] for structural and algorithmic results.

Let us now turn our attention to the mixed-integer case. In (3) we impose the basic assumptions that $W(\mathbb{Z}_+^m) + W'(R_+^{m'}) = \mathbb{R}^s$ and $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$. Then $\Phi(t)$ is a well-defined real

number for all $t \in \mathbb{R}^s$, ([14]). Moreover it holds

$$\begin{aligned}\Phi(t) &= \min\{q^T y + q'^T y' : Wy + W'y' = t, y \in \mathbb{Z}_+^{\bar{m}}, y' \in R_+^{m'}\} \\ &= \min_y \{q^T y + \min_{y'} \{q'^T y' : W'y' = t - Wy, y' \in R_+^{m'}\} : y \in \mathbb{Z}_+^{\bar{m}}\} \\ &= \min_y \{\Phi_y(t) : y \in \mathbb{Z}_+^{\bar{m}}\},\end{aligned}\tag{8}$$

where

$$\Phi_y(t) = q^T y + \max_{k=1, \dots, K} d_k^T (t - Wy) \quad \text{for all } t \in Wy + W'(R_+^{m'}).$$

Here, d_k , $k = 1, \dots, K$ denote the vertices of the polyhedron $\{u \in \mathbb{R}^s : W'^T u \leq q'\}$, and we have applied the argument about Φ_{lin} from the purely linear case. For $t \notin Wy + W'(R_+^{m'})$ the optimization problem $\min_{y'} \{q'^T y' : W'y' = t - Wy, y' \in R_+^{m'}\}$ is infeasible, and we put $\Phi_y(t) = +\infty$. It is convenient to denote $Y(t) := \{y \in \mathbb{Z}_+^{\bar{m}} : \Phi_y(t) < +\infty\}$.

According to (8) the value function Φ is made up by the pointwise minimum of a family of convex, piecewise linear functions whose domains of definition are polyhedral cones arising as shifts of the cone $W'(R_+^{m'})$. By our basic assumption $W(\mathbb{Z}_+^{\bar{m}}) + W'(R_+^{m'}) = \mathbb{R}^s$, the cone $W'(R_+^{m'})$ is full-dimensional.

Some first conclusions about the continuity of Φ may be drawn from the above observations:

1. Suppose that $t \in \mathbb{R}^s$ does not belong to any boundary of any of the sets $Wy + W'(\mathbb{R}_+^{m'})$, $y \in \mathbb{Z}_+^{\bar{m}}$. Then the same is true for all points τ in some open ball B around t . Hence, $Y(\tau) = Y(t)$ for all $\tau \in B$. With an enumeration $(y_n)_{n \in \mathbb{N}}$ of $Y(t)$ we consider the functions $\Phi^\kappa(\tau) := \min\{\Phi_{y_n}(\tau) : n \leq \kappa\}$ for all $\tau \in B$. Then $\lim_{\kappa \rightarrow \infty} \Phi^\kappa(\tau) = \Phi(\tau)$ for all $\tau \in B$. Since, for any function Φ_y , its ‘‘slopes’’ are determined by the same, finitely many vectors d_k , $k = 1, \dots, K$, the functions Φ^κ , $\kappa \in \mathbb{N}$ are all Lipschitz continuous on B with a uniform Lipschitz constant. Thus, the family of functions Φ^κ , $\kappa \in \mathbb{N}$ is equicontinuous on B and has a pointwise limit there. Consequently, this pointwise limit Φ is continuous on B , in fact Lipschitz continuous with the mentioned uniform constant.
2. Any discontinuity point of Φ must be located at the boundary of some set $Wy + W'(\mathbb{R}_+^{m'})$, $y \in \mathbb{Z}_+^{\bar{m}}$. Hence, the set of discontinuity points of Φ is contained in a countable union of hyperplanes. Since $W'(\mathbb{R}_+^{m'})$ has only finitely many facets, this union of hyperplanes subdivides into finitely many classes, such that, in each class, the hyperplanes are parallel. By the rationality of the matrices W and W' , within each class, the pairwise distance of the hyperplanes is uniformly bounded below by some positive number.
3. Let $t_n \rightarrow t$ and $y \in \mathbb{Z}_+^{\bar{m}}$ such that $t_n \in Wy + W'(\mathbb{R}_+^{m'})$ for all sufficiently large n . Since the set $Wy + W'(\mathbb{R}_+^{m'})$ is closed, this yields $t \in Wy + W'(\mathbb{R}_+^{m'})$. Therefore, for sufficiently large n , $Y(t_n) \subseteq Y(t)$. This paves the way for showing that $\liminf_{t_n \rightarrow t} \Phi(t_n) \geq \Phi(t)$, which is the lower semicontinuity of Φ at t .

The above analysis has been refined in [2, 3, 7]. In particular, it is shown in Theorem 3.3 of [7] that, for each $t \in \mathbb{R}^s$, the minimization in (8) can be restricted to a finite set (depending on t , in general). Lemma 5.6.1. and Lemma 5.6.2. of [2] provide the representation of the continuity sets of Φ to be displayed in the subsequent proposition. The global proximity result in part (iv) of the subsequent proposition is derived in Theorem 1 at page 115 of [3] and in Theorem 2.1 of [7]. Altogether, we have the following statement about the mixed-integer value function Φ .

Proposition 2.1 *Let W, W' be rational matrices and assume that $W(\mathbb{Z}_+^{\bar{m}}) + W'(R_+^{m'}) = \mathbb{R}^s$ as well as $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$. Then it holds*

- (i) Φ is real-valued and lower semicontinuous on \mathbb{R}^s ,
- (ii) there exists a countable partition $\mathbb{R}^s = \cup_{i=1}^{\infty} \mathcal{T}_i$ such that the restrictions of Φ to \mathcal{T}_i are piecewise linear and Lipschitz continuous with a uniform constant $L > 0$ not depending on i ,
- (iii) each of the sets \mathcal{T}_i has a representation $\mathcal{T}_i = \{t_i + \mathcal{K}\} \setminus \cup_{j=1}^N \{t_{ij} + \mathcal{K}\}$ where \mathcal{K} denotes the polyhedral cone $W'(R_+^{m'})$ and t_i, t_{ij} are suitable points from \mathbb{R}^s , moreover, N does not depend on i ,
- (iv) there exist positive constants β, γ such that $|\Phi(t_1) - \Phi(t_2)| \leq \beta \|t_1 - t_2\| + \gamma$ whenever $t_1, t_2 \in \mathbb{R}^s$.

3 Implications of Probability Theory

Essential properties of the objective functions $Q_{\mathcal{E}}$ and $Q_{\mathcal{P}}$ in the optimization problems (5) and (6), respectively, have their roots in properties of Φ . We will study these interrelations by employing some basic tools from probability theory as can be found, for instance, in the textbooks [5, 10].

Throughout this section we will impose the basic assumptions that W, W' are rational matrices, that $W(\mathbb{Z}_+^m) + W'(R_+^{m'}) = \mathbb{R}^s$, and that $\{u \in \mathbb{R}^s : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$.

For convenience we denote by μ the image measure $\mathcal{P} \circ h^{-1}$ on \mathbb{R}^s . With this notation, the functions $Q_{\mathcal{E}}$ and $Q_{\mathcal{P}}$ read

$$Q_{\mathcal{E}}(x) = \int_{\mathbb{R}^s} (c^T x + \Phi(h - Tx)) \mu(dh) \quad \text{and} \quad Q_{\mathcal{P}}(x) = \mu(\{h \in \mathbb{R}^s : c^T x + \Phi(h - Tx) > \varphi_o\}).$$

3.1 Measurability

Since $Q_{\mathcal{E}}(x)$ is essentially determined by an integral over Φ and $Q_{\mathcal{P}}(x)$ involves a level set of Φ , it has to be assured that the integral and the probability are taken over a measurable function and a measurable set, respectively.

Proposition 3.1 *For any $x \in \mathbb{R}^m$, $f(x, h) := c^T x + \Phi(h - Tx)$ is a measurable function of h , implying in particular that $Q_{\mathcal{P}}(x)$ is well-defined for all $x \in \mathbb{R}^m$.*

Proof: Φ being lower semicontinuous on \mathbb{R}^s , f is measurable as a superposition of measurable functions. Then, $\{h \in \mathbb{R}^s : f(x, h) > \varphi_o\}$ is a measurable subset of \mathbb{R}^s , and $Q_{\mathcal{P}}(x)$ is well-defined for all $x \in \mathbb{R}^m$. \square

3.2 Integrability

A measurable function into the reals is called integrable, in case its positive and negative parts both are. Integrability is often established via an integrable majorant of the absolute value of the function in question. In the present context, integrability is important for assuring that $Q_{\mathcal{E}}(x)$ is well-defined for all $x \in \mathbb{R}^m$.

Proposition 3.2 *If μ has a finite first moment, i.e., if $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$, then $Q_{\mathcal{E}}(x)$ is well-defined for all $x \in \mathbb{R}^m$.*

Proof: Our basic assumptions imply that $\Phi(0) = 0$. Together with Proposition 2.1(iv) this provides the following estimate

$$\begin{aligned} \int_{\mathbb{R}^s} |\Phi(h - Tx)| \mu(dh) &= \int_{\mathbb{R}^s} |\Phi(h - Tx) - \Phi(0)| \mu(dh) \\ &\leq \int_{\mathbb{R}^s} (\beta \|h - Tx\| + \gamma) \mu(dh) \\ &\leq \beta \int_{\mathbb{R}^s} \|h\| \mu(dh) + \beta \|Tx\| + \gamma. \end{aligned}$$

This implies that $Q_{\mathcal{E}}(x) \in \mathbb{R}$ for all $x \in \mathbb{R}^m$, and the proof is complete. \square

3.3 Continuity of the Probability Measure

Given a sequence $(M_n)_{n \in \mathbb{N}}$ of measurable sets in \mathbb{R}^s , the limes inferior $\liminf_{n \rightarrow \infty} M_n$ and the limes superior $\limsup_{n \rightarrow \infty} M_n$ are defined as the sets of all points belonging to all but a finite number of the M_n , and to infinitely many M_n , respectively. If μ is some probability measure on \mathbb{R}^s , then it holds

$$\mu(\liminf_{n \rightarrow \infty} M_n) \leq \liminf_{n \rightarrow \infty} \mu(M_n) \leq \limsup_{n \rightarrow \infty} \mu(M_n) \leq \mu(\limsup_{n \rightarrow \infty} M_n). \quad (9)$$

This will be our main tool to deduce (semi-)continuity of the function $Q_{\mathcal{P}}$ from the properties of Φ . With $x \in \mathbb{R}^m$ we introduce the notation

$$\begin{aligned} M(x) &:= \{h \in \mathbb{R}^s : c^T x + \Phi(h - Tx) > \varphi_o\}, \\ M_e(x) &:= \{h \in \mathbb{R}^s : c^T x + \Phi(h - Tx) = \varphi_o\}, \\ M_d(x) &:= \{h \in \mathbb{R}^s : \Phi \text{ is discontinuous at } h - Tx\}. \end{aligned}$$

Proposition 3.3 *The function $Q_{\mathcal{P}}$ is lower semicontinuous on \mathbb{R}^m . If, for some $x \in \mathbb{R}^m$, it holds $\mu(M_e(x) \cup M_d(x)) = 0$, then $Q_{\mathcal{P}}$ is continuous at x . The latter assumption is fulfilled for all $x \in \mathbb{R}^m$ if μ has a density.*

Proof: Let us first verify that for all $x \in \mathbb{R}^m$

$$M(x) \subseteq \liminf_{x_n \rightarrow x} M(x_n) \subseteq \limsup_{x_n \rightarrow x} M(x_n) \subseteq M(x) \cup M_e(x) \cup M_d(x). \quad (10)$$

Let $h \in M(x)$. The lower semicontinuity of Φ (Proposition 2.1(i)) yields

$$\liminf_{x_n \rightarrow x} (c^T x_n + \Phi(h - Tx_n)) \geq c^T x + \Phi(h - Tx) > \varphi_o.$$

Therefore, there exists an $n_o \in \mathbb{N}$ such that $c^T x_n + \Phi(h - Tx_n) > \varphi_o$ for all $n \geq n_o$, implying $h \in M(x_n)$ for all $n \geq n_o$. Hence, $M(x) \subseteq \liminf_{x_n \rightarrow x} M(x_n)$.

Let $h \in \limsup_{x_n \rightarrow x} M(x_n) \setminus M(x)$. Then there exists an infinite subset \tilde{N} of \mathbb{N} such that

$$c^T x_n + \Phi(h - Tx_n) > \varphi_o \quad \forall n \in \tilde{N} \quad \text{and} \quad c^T x + \Phi(h - Tx) \leq \varphi_o.$$

Now two cases are possible. First, Φ is continuous at $h - Tx$. Passing to the limit in the first inequality then yields that $c^T x + \Phi(h - Tx) \geq \varphi_o$, and $h \in M_e(x)$. Secondly, Φ is discontinuous at $h - Tx$. In other words, $h \in M_d(x)$, and (10) is established.

By (9) we have for all $x \in \mathbb{R}^m$

$$Q_{\mathcal{P}}(x) = \mu(M(x)) \leq \mu(\liminf_{x_n \rightarrow x} M(x_n)) \leq \liminf_{x_n \rightarrow x} \mu(M(x_n)) = \liminf_{x_n \rightarrow x} Q_{\mathcal{P}}(x_n),$$

verifying the asserted lower semicontinuity. In case $\mu(M_e(x) \cup M_d(x)) = 0$ this argument extends:

$$\begin{aligned} Q_{\mathcal{P}}(x) &= \mu(M(x)) = \mu(M(x) \cup M_e(x) \cup M_d(x)) \geq \mu(\limsup_{x_n \rightarrow x} M(x_n)) \\ &\geq \limsup_{x_n \rightarrow x} \mu(M(x_n)) = \limsup_{x_n \rightarrow x} Q_{\mathcal{P}}(x_n), \end{aligned}$$

and $Q_{\mathcal{P}}$ is continuous at x .

According to our discussion preceding Proposition 2.1 the set $M_d(x)$ is contained in a countable union of hyperplanes. In view of (8) the same is true for $M_e(x)$. Thus $M_e(x) \cup M_d(x)$ is contained in a set of Lebesgue measure zero, and $\mu(M_e(x) \cup M_d(x)) = 0$ by the absolute continuity of μ . \square

3.4 Fatou's Lemma

For a sequence $(g_n)_{n \in \mathbb{N}}$ of measurable functions from \mathbb{R}^s to \mathbb{R} with an integrable function $g \leq g_n \quad \forall n \in \mathbb{N}$, Fatou's Lemma asserts

$$\int_{\mathbb{R}^s} \liminf_{n \rightarrow \infty} g_n(h) \mu(dh) \leq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^s} g_n(h) \mu(dh).$$

Together with the lower semicontinuity of Φ this will provide the lower semicontinuity of $Q_{\mathcal{E}}$.

Proposition 3.4 *The function $Q_{\mathcal{E}}$ is lower semicontinuous on \mathbb{R}^m provided that $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$.*

Proof: Let $x \in \mathbb{R}^m$ and $x_n \rightarrow x$. We will apply Fatou's Lemma essentially to the functions $g_n(h) := \Phi(h - Tx_n)$. Denote $r := \max_{n \in \mathbb{N}} \|x_n\|$. Proposition 2.1(iv) and $\Phi(0) = 0$ then imply

$$\begin{aligned} \Phi(h - Tx_n) &\geq \Phi(0) - |\Phi(h - Tx_n) - \Phi(0)| \\ &\geq -\beta \|h - Tx_n\| - \gamma \\ &\geq -\beta \|h\| - \beta r \|T\| - \gamma \end{aligned}$$

yielding an integrable minorant g for the family of functions g_n . By Fatou's Lemma and the lower semicontinuity of Φ we have

$$\begin{aligned} Q_{\mathcal{E}}(x) &= \int_{\mathbb{R}^s} (c^T x + \Phi(h - Tx)) \mu(dh) \\ &\leq \int_{\mathbb{R}^s} \liminf_{n \rightarrow \infty} (c^T x_n + \Phi(h - Tx_n)) \mu(dh) \\ &\leq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^s} (c^T x_n + \Phi(h - Tx_n)) \mu(dh) \\ &= \liminf_{n \rightarrow \infty} Q_{\mathcal{E}}(x_n). \end{aligned}$$

□

3.5 Lebesgue's Dominated Convergence Theorem

Let g_n, g ($n \in \mathbb{N}$) be measurable functions from \mathbb{R}^s to \mathbb{R} such that $\lim_{n \rightarrow \infty} g_n = g$, μ -almost surely. If there exists an integrable function $\bar{g} \geq |g_n| \forall n \in \mathbb{N}$, μ -almost surely, then Lebesgue's Dominated Convergence Theorem asserts that g_n, g ($n \in \mathbb{N}$) are integrable and that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^s} g_n(h) \mu(dh) = \int_{\mathbb{R}^s} g(h) \mu(dh).$$

This theorem will lead us to the continuity of $Q_{\mathcal{E}}$.

Proposition 3.5 *If $\int_{\mathbb{R}^s} \|h\| \mu(dh) < \infty$ and $\mu(M_d(x)) = 0$ then $Q_{\mathcal{E}}$ is continuous at x . The latter assumption is fulfilled for all $x \in \mathbb{R}^m$ if μ has a density.*

Proof: Let $x \in \mathbb{R}^m$, $x_n \rightarrow x$, and $r := \max_{n \in \mathbb{N}} \|x_n\|$. Again we employ Proposition 2.1(iv) and $\Phi(0) = 0$, and we obtain

$$|\Phi(h - Tx_n)| = |\Phi(h - Tx_n) - \Phi(0)| \leq \beta \|h\| + \beta r \|T\| + \gamma$$

providing us with an integrable majorant. By $\mu(M_d(x)) = 0$, we have

$$\lim_{n \rightarrow \infty} (c^T x_n + \Phi(h - Tx_n)) = c^T x + \Phi(h - Tx) \quad \text{for } \mu\text{-almost all } h \in \mathbb{R}^s.$$

Now Lebesgue's Dominated Convergence Theorem completes the proof:

$$\begin{aligned} \lim_{n \rightarrow \infty} Q_{\mathcal{E}}(x_n) &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}^s} (c^T x_n + \Phi(h - Tx_n)) \mu(dh) \\ &= \int_{\mathbb{R}^s} (c^T x + \Phi(h - Tx)) \mu(dh) = Q_{\mathcal{E}}(x). \end{aligned}$$

□

3.6 Convergence of Probability Measures - Rubin's Theorem

The dependence of the optimization problems (5) and (6) on the underlying probability measure, although seemingly a theoretical issue, has practical relevance in various respects. When building models like (5) and (6) the probability measure often enters in a subjective way or results from an approximation based on statistical data. Moreover, the integrals in (5) and (6) are typically multivariate, and their integrands are given only implicitly. This poses insurmountable numerical difficulties if the probability distribution is continuous. A possible remedy then consists of approximation via discrete distributions. All this motivates considerations on whether "small" perturbations of the underlying probability measure result in "small" perturbations of optimal values and optimal solutions to (5) and (6). The mathematical machinery for addressing these issues is provided by the stability analysis of stochastic programs (for surveys see [11, 21]).

A first and crucial step towards stability analysis is to study $Q_{\mathcal{E}}$ and $Q_{\mathcal{P}}$ as functions jointly in x and μ . Again the value function Φ will provide some valuable insights. Let us consider $Q_{\mathcal{E}}$ and $Q_{\mathcal{P}}$ as

functions on $\mathbb{R}^m \times \mathcal{P}(\mathbb{R}^s)$ where $\mathcal{P}(\mathbb{R}^s)$ denotes the set of all (Borel) probability measures on \mathbb{R}^s . As an essential prerequisite some convergence notion is needed on $\mathcal{P}(\mathbb{R}^s)$. Here, weak convergence of probability measures has proven both sufficiently general to cover relevant applications and sufficiently specific to enable substantial statements. A sequence $\{\mu_n\}$ in $\mathcal{P}(\mathbb{R}^s)$ is said to converge weakly to $\mu \in \mathcal{P}(\mathbb{R}^s)$, written $\mu_n \xrightarrow{w} \mu$, if for any bounded continuous function $g : \mathbb{R}^s \rightarrow \mathbb{R}$ we have

$$\int_{\mathbb{R}^s} g(\xi) \mu_n(d\xi) \rightarrow \int_{\mathbb{R}^s} g(\xi) \mu(d\xi) \quad \text{as } n \rightarrow \infty. \quad (11)$$

A basic reference for weak convergence of probability measures is Billingsley's book [4].

We are heading for sufficient conditions for the the continuity of Q_E and Q_P jointly in x and μ . Beside properties of the value function, a theorem on weak convergence of image measures attributed in [4] to Rubin will turn out most useful. The theorem states the following: Given a sequence of probability measures $\mu_n \xrightarrow{w} \mu$ and measurable functions g_n, g ($n \in \mathbb{N}$) from \mathbb{R}^s to \mathbb{R} such that $\mu(E) = 0$ where $E := \{h \in \mathbb{R}^s : \exists h_n \rightarrow h \text{ such that } g_n(h_n) \not\rightarrow g(h)\}$, it holds $\mu_n \circ g_n^{-1} \xrightarrow{w} \mu \circ g^{-1}$.

Proposition 3.6

- (i) Let $\mu \in \mathcal{P}(\mathbb{R}^s)$ be such that $\mu(M_e(x) \cup M_d(x)) = 0$. Then $Q_P : \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^s) \rightarrow \mathbb{R}$ is continuous at (x, μ) .
- (ii) Fix arbitrary $p > 1$ and $K > 0$, and denote $\Delta_{p,K}(\mathbb{R}^s) := \{\nu \in \mathcal{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \|h\|^p \nu(dh) \leq K\}$. Let $\mu \in \Delta_{p,K}(\mathbb{R}^s)$ be such that $\mu(M_d(x)) = 0$. Then $Q_E : \mathbb{R}^m \times \Delta_{p,K}(\mathbb{R}^s) \rightarrow \mathbb{R}$ is continuous at (x, μ) .

Proof: To prove (i), let $x_n \rightarrow x$ (in \mathbb{R}^m) and $\mu_n \xrightarrow{w} \mu$ (in $\mathcal{P}(\mathbb{R}^s)$) be arbitrary sequences. By $\chi_n, \chi : \mathbb{R}^s \rightarrow \{0, 1\}$ we denote the indicator functions of the sets $M(x_n), M(x), n \in \mathbb{N}$. With these functions we consider the exceptional set E from above:

$$E := \{h \in \mathbb{R}^s : \exists h_n \rightarrow h \text{ such that } \chi_n(h_n) \not\rightarrow \chi(h)\}.$$

To see that $E \subseteq M_e(x) \cup M_d(x)$, assume that $h \in (M_e(x) \cup M_d(x))^c = (M_e(x))^c \cap (M_d(x))^c$ where the superscript c denotes the set-theoretic complement. Then Φ is continuous at $h - Tx$, and either $c^T x + \Phi(h - Tx) > \varphi_o$ or $c^T x + \Phi(h - Tx) < \varphi_o$. Thus, for any sequence $h_n \rightarrow h$ there exists an $n_o \in \mathbb{N}$ such that for all $n \geq n_o$ either $c^T x_n + \Phi(h_n - Tx_n) > \varphi_o$ or $c^T x_n + \Phi(h_n - Tx_n) < \varphi_o$. Hence, $\chi_n(h_n) \rightarrow \chi(h)$ as $h_n \rightarrow h$, implying $h \in E^c$.

In view of $E \subseteq M_e(x) \cup M_d(x)$ and $\mu(M_e(x) \cup M_d(x)) = 0$ we obtain that $\mu(E) = 0$. Rubin's Theorem now yields $\mu_n \circ \chi_n^{-1} \xrightarrow{w} \mu \circ \chi^{-1}$.

Since $\mu_n \circ \chi_n^{-1}, \mu \circ \chi^{-1}, n \in \mathbb{N}$ are probability measures on $\{0, 1\}$, their weak convergence particularly implies that

$$\mu_n \circ \chi_n^{-1}(\{1\}) \rightarrow \mu \circ \chi^{-1}(\{1\}).$$

This is nothing but $\mu_n(M(x_n)) \rightarrow \mu(M(x))$ or $Q_P(x_n, \mu_n) \rightarrow Q_P(x, \mu)$, and (i) is shown.

For proving (ii) let again $x_n \rightarrow x$ in \mathbb{R}^m , but $\mu_n \xrightarrow{w} \mu$ in $\Delta_{p,K}(\mathbb{R}^s)$. Introduce measurable functions g_n, g by $g_n(h) := \Phi(h - Tx_n), g(h) := \Phi(h - Tx), n \in \mathbb{N}$. For the corresponding exceptional set E a simple continuity argument provides $M_d(x)^c \subseteq E^c$ or, equivalently, $E \subseteq M_d(x)$. Hence, $\mu(E) = 0$, and Rubin's Theorem provides

$$\mu_n \circ g_n^{-1} \xrightarrow{w} \mu \circ g^{-1}. \quad (12)$$

To prove the assertion it is sufficient to show that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^s} g_n(h) \mu_n(dh) = \int_{\mathbb{R}^s} g(h) \mu(dh).$$

Changing variables yields the equivalent statement

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} t \mu_n \circ g_n^{-1}(dt) = \int_{\mathbb{R}} t \mu \circ g^{-1}(dt).$$

For fixed $a \in \mathbb{R}_+$, consider the truncation $\kappa_a : \mathbb{R} \rightarrow \mathbb{R}$ with

$$\kappa_a(t) := \begin{cases} t & , \quad |t| < a \\ 0 & , \quad |t| \geq a. \end{cases}$$

Now

$$\begin{aligned}
\left| \int_{\mathbb{R}} t \mu_n \circ g_n^{-1}(dt) - \int_{\mathbb{R}} t \mu \circ g^{-1}(dt) \right| &\leq \left| \int_{\mathbb{R}} (t - \kappa_a(t)) \mu_n \circ g_n^{-1}(dt) \right| \\
&+ \left| \int_{\mathbb{R}} \kappa_a(t) \mu_n \circ g_n^{-1}(dt) - \int_{\mathbb{R}} \kappa_a(t) \mu \circ g^{-1}(dt) \right| \\
&+ \left| \int_{\mathbb{R}} (\kappa_a(t) - t) \mu \circ g^{-1}(dt) \right|. \tag{13}
\end{aligned}$$

The proof is completed by showing that, for a given $\varepsilon > 0$, each of the three expressions on the right becomes less than $\varepsilon/3$ provided that n and a are sufficiently large.

For the first expression we obtain

$$\begin{aligned}
\left| \int_{\mathbb{R}} (t - \kappa_a(t)) \mu_n \circ g_n^{-1}(dt) \right| &\leq \int_{\{t: |t| \geq a\}} |t| \mu_n \circ g_n^{-1}(dt) \\
&= \int_{\{h: |g_n(h)| \geq a\}} |g_n(h)| \mu_n(dh). \tag{14}
\end{aligned}$$

Since $p > 1$,

$$\begin{aligned}
\int_{\mathbb{R}^s} |g_n(h)|^p \mu_n(dh) &\geq \int_{\{h: |g_n(h)| \geq a\}} |g_n(h)| \cdot |g_n(h)|^{p-1} \mu_n(dh) \\
&\geq a^{p-1} \int_{\{h: |g_n(h)| \geq a\}} |g_n(h)| \mu_n(dh). \tag{15}
\end{aligned}$$

Therefore, the estimate in (14) can be continued by

$$\leq a^{1-p} \int_{\mathbb{R}^s} |g_n(h)|^p \mu_n(dh). \tag{16}$$

Proposition 2.1(iv) and $g_n(0) = 0$ imply

$$|g_n(h)|^p \leq (\beta \|h\| + \beta \|x_n\| \cdot \|T\| + \gamma)^p.$$

Since $(x_n)_{n \in \mathbb{N}}$ is bounded and all μ_n belong to $\Delta_{p,K}(\mathbb{R}^s)$, there exists a positive constant c such that

$$\int_{\mathbb{R}^s} |g_n(h)|^p \mu_n(dh) \leq c \quad \text{for all } n \in \mathbb{N}.$$

Hence, (16) can be estimated above by c/a^{p-1} which becomes less than $\varepsilon/3$ if a is sufficiently large.

We now turn to the second expression in (13). Since every probability measure on the real line has at most countably many atoms, we have $\mu \circ g^{-1}(\{t : |t| = a\}) = 0$ for (Lebesgue-)almost all $a \in \mathbb{R}$. Therefore, κ_a is a measurable function whose set of discontinuity points D_{κ_a} has $\mu \circ g^{-1}$ -measure zero for almost all $a \in \mathbb{R}$. We apply Rubin's Theorem to the weakly convergent sequence $\mu_n \circ g_n^{-1} \xrightarrow{w} \mu \circ g^{-1}$, cf. (12), and the identical sequence of functions κ_a . The role of the exceptional set then is taken by D_{κ_a} , and Rubin's Theorem is working due to $\mu \circ g^{-1}(D_{\kappa_a}) = 0$. This yields the conclusion

$$\mu_n \circ g_n^{-1} \circ \kappa_a^{-1} \xrightarrow{w} \mu \circ g^{-1} \circ \kappa_a^{-1} \quad \text{for almost all } a \in \mathbb{R}. \tag{17}$$

Consider the bounded continuous function $\eta : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$\eta(t') := \begin{cases} -a & , \quad t' \leq -a \\ t' & , \quad -a \leq t' \leq a \\ a & , \quad t' \geq a. \end{cases}$$

By the weak convergence in (17), we obtain

$$\int_{\mathbb{R}} \eta(t') \mu_n \circ g_n^{-1} \circ \kappa_a^{-1}(dt') \rightarrow \int_{\mathbb{R}} \eta(t') \mu \circ g^{-1} \circ \kappa_a^{-1}(dt') \quad \text{as } n \rightarrow \infty. \tag{18}$$

Changing variables provides

$$\int_{\mathbb{R}} \eta(t') \mu_n \circ g_n^{-1} \circ \kappa_a^{-1}(dt') = \int_{\kappa_a^{-1}(\mathbb{R})} \eta(\kappa_a(t)) \mu_n \circ g_n^{-1}(dt) = \int_{\mathbb{R}} \kappa_a(t) \mu_n \circ g_n^{-1}(dt).$$

Analogously,

$$\int_{\mathbb{R}} \eta(t') \mu \circ g^{-1} \circ \kappa_a^{-1}(dt') = \int_{\mathbb{R}} \kappa_a(t) \mu \circ g^{-1}(dt).$$

The above identities together with (18) confirm that the second expression on the right-hand side of (13) becomes arbitrarily small for sufficiently large n and almost all sufficiently large a .

Let us finally turn to the third expression in (13). Analogously to (14), (15), and (16) we obtain

$$\left| \int_{\mathbb{R}} (\kappa_a(t) - t) \mu \circ g^{-1}(dt) \right| \leq a^{1-p} \int_{\mathbb{R}^s} |g(h)|^p \mu(dh).$$

The integral $\int_{\mathbb{R}^s} |g(h)|^p \mu(dh)$ is finite due to Proposition 2.1(iv) and $\int_{\mathbb{R}^s} \|h\|^p \mu(dh) \leq K$. Hence, the third expression in (13) becomes less than $\varepsilon/3$ if a is large enough. \square

4 Outlook

The previous sections have shown how the mixed-integer value function serves as a point of departure for understanding the basic structure of stochastic integer programs. Let us finally have a look at some further developments whose detailed coverage is beyond the scope of the present paper.

Quantitative Statements. The continuity results of Section 3 are all qualitative by nature. Lipschitz continuity of $Q_{\mathbb{E}}$ and $Q_{\mathbb{P}}$ as functions in x has been studied in [18, 19, 23]. To quantify the continuity of $Q_{\mathbb{E}}$ and $Q_{\mathbb{P}}$ as functions of the underlying probability measure a proper metric in the space of probability measures has to be identified. Here “proper” means that the metric should allow an estimation of function value distances at all and that it should metrize important modes of convergence such as weak convergence of probability measures. For the function $Q_{\mathbb{E}}$ a first proposal along these lines was made in [20]. The mentioned quantitative studies require as input refined statements about Φ such as parts (ii) and (iii) of Proposition 2.1.

Stability. As already mentioned in Subsection 3.6, perturbation and approximation of the underlying probability measure arise quite naturally in stochastic programming. The stability analysis of stochastic programs then provides justification for replacing unknown probability measures by statistical estimates or for turning numerically intractable multivariate integrals into manageable finite sums by approximating continuous distributions via discrete ones. Typical stability results assert that optimal values and optimal solutions are (semi-)continuous (multi-)functions of the underlying probability measure, [1, 19, 20]. These results are obtained by putting general techniques from parametric optimization into perspective with stochastic programming, [21]. This leads to studying the joint dependence of relevant integral functionals on both the decision variable and the probability measure. For the latter, Proposition 3.6 provides a paradigmatic example.

Algorithms. Methods for solving the optimization problems (5) and (6), almost exclusively, rest on the assumption that the probability measures underlying the models are discrete. This does not provide a serious restriction, since, on the one hand, in many practical situations the uncertain data is available via discrete observations, only. On the other hand, the above mentioned stability results justify approximation via discrete measures should the precise model involve a continuous probability distribution. With discrete probability measures, the problems (5) and (6) can be rewritten as large-scale, block-structured, mixed-integer linear programs. Decomposition then becomes the algorithmic method of choice, but the presence of integer variables poses a number of open problems. Some first attempts were made in [9, 23], see also the survey [13].

References

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

- [2] Bank, B.; Guddat, J.; Klatte, D.; Kummer, B.; Tammer, K.: Non-linear Parametric Optimization, Akademie-Verlag, Berlin, 1982.
- [3] Bank, B.; Mandel, R.: Parametric Integer Optimization, Akademie-Verlag, Berlin 1988.
- [4] Billingsley, P.: Convergence of Probability Measures, Wiley, New York, 1968.
- [5] Billingsley, P.: Probability and Measure, Wiley, New York, 1986.
- [6] Birge, J.R.; Louveaux, F.: Introduction to Stochastic Programming, Springer, New York, 1997.
- [7] Blair, C.E.; Jeroslow, R.G.: The value function of a mixed integer program: I, Discrete Mathematics 19 (1977), 121-138.
- [8] Bonnans, J.F.; Shapiro, A.: Perturbation Analysis of Optimization Problems, Springer-Verlag, New York, 2000.
- [9] Carøe, C.C.; Schultz, R.: Dual decomposition in stochastic integer programming, Operations Research Letters 24 (1999), 37-45.
- [10] Dudley, R.M.: Real Analysis and Probability, Wadsworth & Brooks/Cole, Pacific Grove, California 1989.
- [11] Dupačová, J.: Stochastic programming with incomplete information: a survey of results on postoptimization and sensitivity analysis, Optimization 18 (1987), 507-532.
- [12] Kall, P.; Wallace, S.W.: Stochastic Programming, Wiley, Chichester, 1994.
- [13] Klein Haneveld, W.K.; van der Vlerk, M.H.: Stochastic integer programming: General models and algorithms, Annals of Operations Research 85 (1999), 39-57.
- [14] Nemhauser, G.L.; Wolsey, L.A.: Integer and Combinatorial Optimization, Wiley, New York 1988.
- [15] Ogryczak, W.; Ruszczyński, A.: From stochastic dominance to mean-risk models: Semideviations as risk measures, European Journal of Operational Research 116 (1999), 33-50.
- [16] Ogryczak, W.; Ruszczyński, A.: Dual stochastic dominance and related mean-risk models, Rutcor Research Report 10-2001, Rutgers Center for Operations Research, Piscataway, 2001.
- [17] Prékopa, A.: Stochastic Programming, Kluwer, Dordrecht, 1995.
- [18] Schultz, R.: Continuity properties of expectation functions in stochastic integer programming, Mathematics of Operations Research 18 (1993), 578-589.
- [19] Schultz, R.: On structure and stability in stochastic programs with random technology matrix and complete integer recourse, Mathematical Programming 70 (1995), 73-89.
- [20] Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse, SIAM Journal on Optimization 6 (1996), 1138-1152.
- [21] Schultz, R.: Some aspects of stability in stochastic programming, Annals of Operations Research 100 (2000), 55-84.
- [22] Schultz, R.: Probability objectives in stochastic programs with recourse, Preprint, Institute of Mathematics, Gerhard-Mercator University Duisburg, 2001.
- [23] Tiedemann, S.: Probability Functionals and Risk Aversion in Stochastic Integer Programming, Diploma Thesis, Department of Mathematics, Gerhard-Mercator University Duisburg, 2001.