Konrad-Zuse-Zentrum für Informationstechnik Berlin
Heilbronner Str. 10, D-10711 Berlin - Wilmersdorf

Rüdiger Schultz

Structure and Stability in Two-Stage Stochastic Programming

(Habilitationsschrift, Humboldt-Universität zu Berlin, Mai 1994)

Preprint SC 95-15 (May 1995)

Structure and Stability in Two-Stage Stochastic Programming

Habilitationsschrift
zur Erlangung der Lehrbefähigung
für das Fach Mathematik

vorgelegt dem Rat der Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von

Dr. Rüdiger Schultz

geb. am 28. April 1959 in Bergen/Rg.

Prof. Dr. M. Dürkop
Präsidentin
der Humboldt-Universität zu Berlin

Prof. Dr. K. Gröger
Dekan

Berlin, den 6. Mai 1994

Gutachter:

1. Prof. Dr. Z. Artstein, Rehovot
2. Prof. Dr. P. Kall, Zürich
3. Prof. Dr. B. Kummer, Berlin
4. Prof. Dr. R. J-B Wets, Davis

Acknowledgements

First of all, I want to thank Prof. Werner Römisch (Humboldt University Berlin) for many years of fruitful cooperation. Our joint work in stochastic programming has been a constant source of inspiration for the topics addressed in this thesis.
I would like to express my gratitude to my former colleagues at the Humboldt University Berlin for many stimulating discussions and continued encouragement. In particular, I wish to thank Prof. Bernd Kummer, who introduced me into scientific work as well as Profs. Jürgen Guddat and Bernd Bank for their continued interest in the progress of my research.
I want to extend my gratitude to Prof. Martin Grötschel for letting me share the excellent working conditions at the Konrad-Zuse-Zentrum für Informationstechnik Berlin since October 1993.
Further thanks are due to Prof. Peter Kall (University of Zurich) for giving me the opportunity to work at his institute during the initial stage of my activities in stochastic programming.
Moreover, this thesis has benefitted from joint research projects carried out together with Prof. Sjur Flåm (University of Bergen), Profs. Georg Pflug (University of Vienna) and Andrzej Ruszczyński (IIASA Laxenburg) and Drs. Leen Stougie (University of Amsterdam) and Maarten van der Vlerk (University of Groningen).
My stay at the University of Zurich was sponsored by a grant of the Eidgenössische Stipendienkommisssion für ausländische Studierende. For the last two years the project "Approximation und Stabilität stochastischer Optimierungsprobleme" (jointly with Werner Römisch) has been part of the Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuerung" of the Deutsche Forschungsgemeinschaft. I would like to thank the mentioned institutions for their support.
Last but not least, I wish to thank Christa Dobers (Humboldt University Berlin) for preparing parts of the manuscript and Jutta Kerger (Humboldt University Berlin) as well as Sybille Mattrisch (Konrad-Zuse-Zentrum für Informationstechnik Berlin) for linguistic support.

All chapters of this thesis are individually accessible. The first chapter is an introduction and summary. Chapter 2 has been accepted for publication in *Journal of Computational and Applied Mathematics* [105]. Chapter 3 is joint work with Werner Römisch whom I would like to thank for the permission to use it here. It is the revised version of a submission to *SIAM Journal on Optimization* [97]. Chapter 4 is based on the revised version of a submission to *Mathematical Programming* [106] and the paper [104] that appeared in *Mathematics of Operations Research*. Chapter 5 appeared as a preprint in the series of the Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuerung" of the Deutsche Forschungsgemeinschaft [107].

Berlin, May 6, 1994. Rüdiger Schultz

Contents

Abstract

This thesis is concerned with structural properties and the stability behaviour of two-stage stochastic programs.
Chapter 1 gives an introduction into stochastic programming and a summary of the main results of the thesis.
In Chapter 2 we present easily verifiable sufficient conditions for the strong convexity of the expected-recourse function in a stochastic program with linear complete recourse. Different levels of randomness in the data are considered. We start with models where only the right-hand side of the constraints is random and extend these results to the situation where also the technology matrix contains random entries. The statements on strong convexity imply new stability estimates for sets of optimal solutions when perturbing the underlying probability measure. We work out Hölder estimates (in terms of the L_1-Wasserstein distance) for optimal solution sets to linear recourse models with random technology matrix.
In Chapter 3 (*joint work with Werner Römisch, Berlin*) we are aiming at the Lipschitz stability of optimal solution sets to linear recourse models with random right-hand side. To this end , we first adapt the distance notion for the underlying probability measures to the structure of the model and derive a Lipschitz estimate for optimal solutions based on that distance. Here, the strong convexity established in Chapter 2 turns out as an essential assumption. For applications, however, a Lipschitz estimate with respect to a more accesssible probability distance is desirable. Structural properties of the expected-recourse function finally permit such an estimate in terms of the Kolmogorov-Smirnov distance of linear transforms of the underlying measures. The general analysis is specified to estimation via empirical measures. We obtain a law of iterated logarithm, a large deviation estimate and an estimate for the asymptotic distribution of optimal solution sets.
Chapters 4 and 5 deal with two-stage linear stochastic programs where integrality constraints occur in the second stage. In Chapter 4 we study basic continuity properties of the expected-recourse function for models with random right-hand side and random technology matrix. The joint continuity with respect to the decision variable and the underlying probability measure leads to qualitative statements on the stability of local optimal values and local optimal solutions.
In Chapter 5 we demonstrate that a variational distance of probability measures based on a suitable Vapnik-Červonenkis class of Borel sets leads to convergence rates of the Hölder type for the expected recourse as a function of the underlying probability measure. The rates carry over to the convergence of local optimal values. As

an application we again consider estimation via empirical measures. Beside qualitative asymptotic results for optimal values and optimal solutions we obtain a law of iterated logarithm for optimal values.

Zusammenfassung

Die vorliegende Habilitationsschrift befaßt sich mit Struktureigenschaften und dem Stabilitätsverhalten zweistufiger stochastischer Optimierungsprobleme.
Kapitel 1 bietet eine Einführung in die stochastische Optimierung und eine Übersicht über Hauptresultate der Arbeit.
In Kapitel 2 präsentieren wir leicht zu verifizierende hinreichende Bedingungen für die starke Konvexität von Kompensationsfunktionalen in stochastischen Programmen mit linearer vollständiger Kompensation. Verschiedene Grade der Zufallsabhängigkeit werden betrachtet. Wir beginnen mit Modellen, in denen nur die rechte Seite der Nebenbedingungen zufällig ist und erweitern die Resultate auf den Fall zufälliger Komponenten in der Technologiematrix. Die Aussagen zur starken Konvexität implizieren neue Stabilitätsabschätzungen für Optimalmengen bei Störung des zugrundeliegenden Wahrscheinlichkeitsmaßes. Wir erhalten Hölder-Abschätzungen (bezüglich des L_1-Wasserstein-Abstandes) für Optimalmengen von linearen Kompensationsproblemen mit zufälliger Technologiematrix.
In Kapitel 3 (*gemeinsame Arbeit mit Werner Römisch, Berlin*) streben wir Lipschitz-Stabilitätsaussagen für Optimalmengen linearer Kompensationsprobleme mit zufälliger rechter Seite an. Zu diesem Zweck passen wir den Abstandsbegriff für die zugrundeliegenden Wahrscheinlichkeitsmaße zunächst der Modellstruktur an und leiten eine Lipschitz-Abschätzung für Optimallösungen bezüglich dieses Abstandes her. Die in Kapitel 2 erzielte starke Konvexität erweist sich dabei als wesentliche Voraussetzung. Für Anwendungen ist jedoch eine Lipschitz-Abschätzung bezüglich eines leichter auswertbaren Wahrscheinlichkeitsabstandes wünschenswert. Struktureigenschaften des Kompensationsfunktionals erlauben schließlich eine solche Abschätzung bezüglich des Kolmogorov-Smirnov Abstandes linerarer Transformationen der zugrundeliegenden Maße. Die allgemeinen Untersuchungen werden dann auf die Schätzung mittels empirischer Maße spezifiziert. Wir erhalten ein Gesetz vom iterierten Logarithmus, eine Abschätzung für große Abweichungen und eine Abschätzung für die asymptotische Verteilung der Optimalmengen.
Die Kapitel 4 und 5 befassen sich mit zweistufigen linearen stochastischen Optimierungsaufgaben, bei denen Ganzzahligkeitsforderungen in der zweiten Stufe auftreten.
In Kapitel 4 untersuchen wir grundlegende Stetigkeitseigenschaften des Kompensationsfunktionals für Modelle mit zufälliger rechter Seite und zufälliger Technologiematrix. Die gemeinsame Stetigkeit bezüglich Entscheidungsvariable und zugrundeliegendem Wahrscheinlichkeitsmaß führt zu qualitativen Aussagen über die Stabilität lokaler Optimalwerte und lokaler Optimallösungen.

In Kapitel 5 zeigen wir, daß ein Variationsabstand von Wahrscheinlichkeitsmaßen basierend auf einer geeigneten Vapnik-Červonenkis Klasse von Borelmengen zu Konvergenzraten vom Höldertyp für das Kompensationsfunktional als Funktion des zugrundeliegenden Wahrscheinlichkeitsmaßes führt. Die Raten übertragen sich auf die Konvergenz lokaler Optimalwerte. Als Anwendung betrachten wir erneut die Schätzung mittels empirischer Maße. Neben qualitativen asymptotischen Resultaten für Optimalwerte und Optimallösungen erhalten wir ein Gesetz vom iterierten Logarithmus für Optimalwerte.

Chapter 1

Introduction

This thesis is concerned with structural properties and the stability behaviour of stochastic programming models. Stochastic programs are specific nonlinear programs that are derived from random optimization problems. Their objectives (and/or constraints) are typically given via parameter dependent multiple integrals. Our analysis aims at relating verifiable properties of basic ingredients of the stochastic program to its structure and stability. We focus our considerations on two-stage stochastic programs that may also contain integrality constraints.
In the introduction we collect some fundamentals of stochastic programming methodology and present the main results in condensed form.

1.1 Stochastic Programming Models

The mathematical modelling of phenomena in nature, technology and economics typically involves some level of uncertainty. Random parameters occuring among the data of an optimization problem raise the difficulty that, without knowing their outcomes, it is in general impossible to detect feasibility and/or optimality of a given decision. Depending on the modelling environment and the availability of (statistical) information on the random data stochastic programming offers models for finding optimal decisions under uncertainty. Basically, it is assumed that probability distributions can be assigned to the random data. Optimal decisions are then found in a 'here-and-now' manner, i.e. based on the a priori information alone without further observation of the random data as it is characteristic for 'wait-and-see' approaches.

Stochastic programming was pioneered in the fifties by Dantzig [18], Beale [6], Tintner [120] and Charnes/Cooper [14]. The development of the field until the seventies is reflected in Kall [46], Wets [131] and Dempster [20]. An overview of recent results can be obtained from Birge/Wets [11], Ermoliev/Wets [32] and Wets [133].

Let us start with some modelling alternatives in stochastic programming. To this end we consider the random nonlinear program

$$\min\{g_0(x,\xi) : g(x,\xi) \leq 0, x \in C\}$$

where $\xi \in I\!\!R^S$ follows a probability distribution on $I\!\!R^S$ induced by some Borel probability measure μ on $I\!\!R^S$ which does not depend on the decision variable $x \in I\!\!R^m$. The non-random set $C \subset I\!\!R^m$ is assumed to be non-empty and closed and the functions $g_0(x,.) : I\!\!R^S \longrightarrow I\!\!R, g(x,.) : I\!\!R^S \longrightarrow I\!\!R^{\tilde{m}}$ to be at least measurable for all $x \in C$.

In many practical situations the decision x has to be taken before the outcome of ξ is observed. Take planning in its widest sense as an example: production schedules, investment strategies and technological designs often have to be settled before knowing all the (random) influences of the environment.Hence, objective function and constraint set of the underlying optimization problem are not fixed when the decision x has to be made. Feasibility and optimality of the latter are, thus, meaningless notions. From mathematical viewpoint there are the alternatives to study the feasibility problem within the framework of parameter dependent inequalities and to use techniques from multiobjective programming for the optimality issue. But both these approaches neglect the "frequency-of-occurrence" information on the unknown parameters which is captured in the underlying probability distribution.

A conceptual frame for stochastic programs is given by the model

$$\min\{\int_{I\!\!R^S} f_o(x,\xi)\mu(d\xi) : \int_{I\!\!R^S} f_i(x,\xi)\mu(d\xi) \leq 0, i = 1,\cdots,N, x \in C\}$$

where the (measurable) functions $f_o, f_i (i = 1,\cdots,N)$ depend on the modelling environment. Observe that now, up to finiteness of the integrals, objective and constraints are well defined. To give an impression on the background of the above abstract model we expand a little on three instances : chance constrained, two- and multi-stage stochastic programs.

Chance constraints. For some probability level $\alpha \in [0,1]$ we define

$$f_1(x,\xi) = \begin{cases} \alpha - 1 & \text{if } g(x,\xi) \leq 0 \\ \alpha & \text{otherwise.} \end{cases}$$

Then one confirms that

$$\int_{\mathbb{R}^S} f_1(x,\xi)\mu(d\xi) \leq 0$$

is equivalent to

$$\mu(\{\xi \in \mathbb{R}^S : g(x,\xi) \leq 0\} \geq \alpha.$$

This reflects a reliability constraint: $x \in \mathbb{R}^m$ is considered feasible if the probability of the set of all those $\xi \in \mathbb{R}^S$ such that $g(x,\xi) \leq 0$ exceeds a given level α. In the literature, this type of constraint is called (joint) probabilistic or chance constraint. In particular, the situation where g is linear was studied extensively. Topics addressed in this context include sufficient conditions for the convexity of a chance constrained set ([67], [77], [78]), procedures for computing function values and gradients of the relevant constraint functions ([19], [70], [79], [119], [121]) and stability properties when perturbing the underlying probability measure ([26], [48], [94], [103], [124]).

Two-stage models. While in a chance constraint model the decision does not depend in any way on future observations of the random data, the two-stage model, though being a "here-and-now" model, includes both long-term anticipatory decisions before and short-term adaptive actions after observation of the random data. Let us introduce the variables $x_1 \in \mathbb{R}^{m_1}$ and $x_2 \in \mathbb{R}^{m_2}$ for the long- and short-term decisions, respectively. The random optimization problem then reads

$$\min\{g_0(x_1,x_2,\xi) : g(x_1,x_2,\xi) \leq O, x_1 \in C_1, x_2 \in C_2\}$$

where, for simplicity, C is assumed to be separable, i.e. $C = C_1 \times C_2$.
Given x_1 and ξ the adaptive (recourse, second-stage) action x_2 is selected as the optimal solution $\bar{x}_2 = \bar{x}_2(x_1,\xi)$ of

$$\min\{g_0(x_1,x_2,\xi) : g(x_1,x_2,\xi) \leq O, x_2 \in C_2\}.$$

The anticipatory (first-stage) decision x_1 is selected to minimize the expectation of $g_0(x_1,\bar{x}_2(x_1,.),.)$ subject to $x_1 \in C_1$:

$$\min\{\int_{\mathbb{R}^S} g_0(x_1,\bar{x}_2(x_1,\xi),\xi)\mu(d\xi) : x_1 \in C_1\}.$$

The above optimization problem is called two-stage stochastic program or stochastic program with recourse. Its objective function arises as a (multidimensional) integral with respect to an implicit integrand which is given as the optimal-value function of an optimization problem depending on the parameters x_1 and ξ. Apart from its implicity, the latter function is known to be typically non-convex and non-smooth ([4],[34]). Therefore, the research in two-stage stochastic programming has concentrated on settings where much more structure is available. The present thesis confines to two such settings. We will assume that the functions g_0 and g are basically linear and that C_2 coincides with either $\mathbb{R}^{m_2}_+$ or a mixed-integer set $\mathbb{R}^{m_{2,1}}_+ \times \mathbb{Z}^{m_{2,2}}_+, (m_{2,1} + m_{2,2} = m_2)$. Moreover, the random variable ξ will not occur in the second-stage objective g_0. In the pure linear case the above mentioned optimal-value function then turns out to be convex and can be described explicitly. In the mixed-integer case there is no convexity but still some description that is sufficiently explicit for our purposes.

To illustrate how a two-stage stochastic program arises in practical applications let us expand a little on the pure linear case. Adapting our notation to the more standard one we introduce the variables $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^{\bar{m}}$ for the first- and second-stage decisions, respectively. For simplicity, we assume that the random variable $\xi \in \mathbb{R}^S$ admits a representation $\xi = (z, A) \in \mathbb{R}^s \times \mathbb{R}^{ms}$ where z and A are a vector and a matrix, respectively, that are specified below. Furthermore, the inequality constraint involving g is changed into an equality constraint. The functions g_0 and g are then specified as follows

$$g_0(x, y, z, A) := c^T x + q^T y \qquad , \qquad g(x, y, z, A) := Ax + Wy - z.$$

The sets C_1 and C_2 are a nonempty polyhedron $C \subset \mathbb{R}^m$ and the non-negative orthant $\mathbb{R}^{\bar{m}}_+$, respectively.
The resulting linear two-stage model can now be written in a compact form:

$$\min\{c^T x + Q(x) : x \in C\},$$

where

$$Q(x) := \int_{\mathbb{R}^S} \Phi(z - Ax)\mu(d(z, A))$$

and

$$\Phi(t) := \min\{q^T y : Wy = t, y \geq 0\}.$$

Of course, it needs some further assumptions to have the above functions Q and Φ well-defined. Roughly speaking these consist of a finite first moment for the probability measure μ and primal plus dual feasibility for the linear program defining

Φ. We will return to that issue later on when inspecting the structure of two-stage stochastic programs in more detail.

The following situation gives rise to a model of the above type: Suppose that a company transforms raw materials into certain products for which there are clients whose demands have to be met. The transformation causes costs and involves certain productivities. Further constraints may concern output requirements or capacity restrictions. Production has to be planned in advance, i.e. before knowing the exact productivities of the transformation and/or demands of the clients. All that is known about these data is a probability distribution governing their outcomes. If the production does not exactly meet the demand additional (recourse) costs occur for stocking the surplus or compensating the shortfall. The company then aims at establishing a feasible production schedule such that the production costs plus the expected recourse costs become minimal.
This precisely fits into the above setting. Indeed, x corresponds to the unknown production schedule, c reflects the production costs, A the involved productivities and z the client's demand; C models possible further constraints. The second-stage (recourse) decision $y \in I\!R^{2s}$ splits into components $y^+ \in I\!R^s$ for compensation and $y^- \in I\!R^s$ for stocking. Accordingly, q splits into q^+ and q^- reflecting the respective costs such that the second stage reads

$$\min\{q^{+T}y^+ + q^{-T}y^- : y^+ - y^- = t, y^+ \geq 0, y^- \geq 0\}.$$

The function Q finally coincides with the expected recourse costs based on the (joint) probability distribution μ for the demand z and the productivity A.
Of course, there is a very simple second stage in the above example. It is easy to see that the (formally) s-dimensional minimization splits into s separate one-dimensional minimizations. If the recourse mechanism is more involved this, however, can no longer be maintained and a non-separable linear program becomes relevant. Of course, the mixed-integer case shortly addressed above fits the situation where integer (including Boolean) decisions occur in the second stage.
For a more detailed exposition of practical applications and solution techniques in two-stage stochastic programming we refer to the literature ([10], [11], [32], [37], [42], [43], [45], [47], [49], [69], [99], [100], [101], [102], [127], [128], [132]).

Multi-stage models. The philosophy that underlies a two-stage model is extended by considering a sequential process of observation and decision, i.e. given the stages

$t = 1, \cdots, T$ there are decisions $x_t \in I\!R^{m_t}$ and observations $\xi_t \in I\!R^{s_t}$. As in the two-stage case the (sequences of) decisions and observations determine a cost and the objective is to find a decision rule (which might reduce to finding an anticipatory decision x_1) such that the expectation of the costs becomes minimal. Again there are constraints linking decisions and observations. In contrast to the two-stage case, however, certain information constraints have to be taken into account which reflect the property that a decision x_t must not depend on observations $\xi_{t'}$ for which $t' > t$. The latter is referred to as non-anticipativity. It is modeled via a nest of sigma fields. More specifically, if $\mathcal{B}$ denotes the sigma field of Borel sets in $I\!R^S$ we have a sequence

$$\mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots \subset \mathcal{B}_T \subset \mathcal{B}$$

of sigma fields and the decision x_t has to be $\mathcal{B}_t -$ measurable. Of course, x_1 is anticipatory (deterministic) if we put $\mathcal{B}_1 := \{\emptyset, I\!R^S\}$, and in the two-stage case we have $\mathcal{B}_1 = \{\emptyset, I\!R^S\}, \mathcal{B}_2 = \mathcal{B}$.
We do not further expand on multi-stage stochastic programs and refer to the literature for further details ([21], [30], [33], [35], [36], [73], [91], [130], [133]).

1.2 Structure of Two-Stage Stochastic Programs

1.2.1 Linear Recourse

In this paragraph we introduce into some structural properties of the linear two-stage model set up in the previous section.
Consider

(2.1) $\min\{c^T x + Q(x) : x \in C\}$

where

(2.2) $Q(x) := \int_{I\!R^S} \Phi(z - Ax)\mu(d(z, A))$

and

(2.3) $\Phi(t) := \min\{q^T y : Wy = t, y \geq 0\}$.

Linear two-stage models where the recourse matrix W and/or the recourse costs q are random will not be considered in this thesis. For details on such models we refer to [46],[131].

Properties of the expected recourse function Q are governed by the interplay of the underlying probability measure μ and the value function Φ. A key issue in this respect is to find conditions on the problem data which are both not too restrictive for applications and imply properties of Q that are useful for the theoretical analysis and the numerical treatment, respectively.
The value function Φ is ruled by linear programming duality. In fact, if we consider the feasible regions

$$M_P(t) = \{y \in I\!R^{\bar{m}} : Wy = t, y \geq 0\}$$

and

$$M_D = \{u \in I\!R^s : W^T u \leq q\}$$

of the pair of dual linear programs associated to (2.3), then $\Phi(t)$ is finite if both $M_P(t) \neq \emptyset$ and $M_D \neq \emptyset$. Therefore, we impose the following basic assumptions

(A1) (complete recourse) $W(I\!R^{\bar{m}}_+) = I\!R^s$,

(A2) (dual feasibility) $\{u \in I\!R^s : W^T u \leq q\} \neq \emptyset$.

In terms of the modelling background (A1) says that there exists a compensation (recourse action) whatever the outcome of $z - Ax$ may be. Assumption (A2) then implies that recourse costs are bounded below.
If (A1) is violated then the model only makes sense if it is guaranteed that $z - Ax \in W(I\!R^{\bar{m}}_+)$ for $\mu-$almost all $(z, A) \in I\!R^s \times I\!R^{ms}$. The latter leads to additional constraints (called induced constraints) on the first-stage decision x and is referred to in the literature as relative complete recourse (cf. [46],[131]).

By (A1), (A2) the set M_D is non-empty, bounded and, thus, posesses vertices which we denote by $\tilde{d}_i (i = 1, \ldots, \tilde{\ell})$. Then

$$\Phi(t) = \max_{i=1,\ldots,\tilde{\ell}} \tilde{d}_i^T t.$$

The value function Φ is, hence, piecewise linear and convex. The linearity regions of Φ coincide with the (outer) normal cones $\mathcal{K}_i$ to M_D at the vertices $\tilde{d}_i (i = 1, \ldots, \tilde{\ell})$ (cf. [72],[126]).
The integrand in (2.2) now reads

(2.4) $\Phi(z - Ax) = \max_{i=1,\ldots,\tilde{\ell}} \tilde{d}_i^T (z - Ax).$

Therefore, the integral in (2.2) is finite if

(A3) $\int_{\mathbb{R}^S} (\|z\| + \|A\|)\mu(d(z,A)) < +\infty.$

Now it holds

Proposition 1.2.1 ([46], [131])
Assume (A1) - (A3), then the function Q is real-valued, convex and Lipschitz continuous on $\mathbb{R}^m$.

Proposition 1.2.2 ([46], [131])
Assume (A1) - (A3) and let $x_0 \in \mathbb{R}^m$ be such that for all $i_1, i_2 \in \{1,\ldots,\tilde{\ell}\}, i_1 \neq i_2$

(2.5) $\mu(\{(z,A) \in \mathbb{R}^S : z - Ax_0 \in \mathcal{K}_{i_1} \cap \mathcal{K}_{i_2}\}) = 0.$

Then Q is continuously differentiable at x_0 and the gradient $Q'(x_0)$ is given by

(2.6) $$Q'(x_0) = -\sum_{i=1}^{\tilde{\ell}} \int_{z - Ax_0 \in \mathcal{K}_i} A^T \tilde{d}_i \, \mu(d(z,A)).$$

Proposition 1.2.3 ([68],[129], A non-random)
Assume (A1) - (A3) and that for any non-singular matrix $B \in L(\mathbb{R}^S, \mathbb{R}^S)$ the superposition $\mu \circ B$ has a continuous density such that all lower dimensional densities are continuous too. Then Q is twice continuously differentiable.

The first-order differentiability statement essentially relies on making Lebesgue's dominated convergence theorem work for the partial difference quotients of Q; the second-order result is gained by establishing existence and continuity of the second-order partial derivatives. In [68], Proposition 1.2.3 is stated for random matrix A. Proposition 1.2.3 shows that higher-order differentiability of Q leads to quite involved conditions that are hard to check. Since the sets $\mathcal{K}_{i_1} \cap \mathcal{K}_{i_2}$ are subsets of hyperplanes in $\mathbb{R}^S$, (2.5) is verified if, for instance, μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^S$.

Given the convexity of Q, it is natural to ask for sufficient conditions guaranteeing stronger properties such as strict or strong convexity. A function is called strictly convex if the usual convexity inequality holds strictly for non-identical arguments. The function Q is called strongly convex ([75]) on a non-empty convex subset $V \subset I\!R^m$ if there exists a $\kappa > 0$ such that for all $\lambda \in [0,1]$ and all $x_1, x_2 \in V$

$$Q(\lambda x_1 + (1-\lambda)x_2) \leq \lambda Q(x_1) + (1-\lambda)Q(x_2) - \kappa\lambda(1-\lambda)\|x_1 - x_2\|^2.$$

Strict and strong convexity are useful tools in convex optimization. Strict convexity ensures uniqueness of optimal solutions; under strong convexity decent methods show improved convergence and the function obeys a certain conditioning ([3]) that is beneficial for the theoretical analysis (e.g. stability considerations, cf. Section 1.3).

In Chapter 2 we derive sufficient conditions for the strong convexity of Q. Due to the involved conditions for the twice differentiability of Q we will make some effort to avoid second derivatives. Recall that the integrand Φ in (2.2) is piecewise linear and convex. A discrete measure μ then leads to a piecewise linear function Q which can neither be strictly nor strongly convex. Therefore, we will impose certain continuity assumptions on μ.
Moreover, a separate treatment of models with random and non-random matrix A, respectively, is advisable. Indeed, if A is non-random then it can easily be seen that Q is constant on translates of the null space of A. Since this null space is typically non-trivial the function Q is "almost never" strictly (let alone strongly) convex. But apart from that, some improved convexity can be established for the function

$$(2.7)\quad \tilde{Q}(\chi) := \int_{I\!R^s} \Phi(z-\chi)\mu(dz)$$

which will be the first main result in Chapter 2:

Proposition 1.2.4 (Theorem 2.2.2 in Chapter 2)

(i) Assume (A1), (A3) and

*(A2)** *there exists a vector $\bar{u} \in I\!R^s$ such that $W^T\bar{u} < q$ componentwise,*

(A4) *the probability measure μ is absolutely continuous with respect to the Lebesgue measure on $I\!R^s$.*

Then $\tilde{Q}$ is strictly convex on any open convex subset $V \subset I\!R^s$ of the support of μ.

(ii) Assume (A1), (A2), (A3) and*

*(A4)** *there exist a convex open set $V \subset I\!\!R^s$, constants $r > 0$, $\varrho > 0$ and a density Θ_μ of μ such that*

$$\Theta_\mu(\tau) \geq r \quad \textit{for all} \quad \tau \in R^s \quad \textit{with} \quad \mathrm{dist}(\tau, V) \leq \varrho.$$

Then $\tilde{Q}$ is strongly convex on V.

As to the necessity of the above conditions we prove in Chapter 2 that, given (A1) - (A4), the condition $(A2)^*$ is also necessary for the strict convexity of $\tilde{Q}$. Further remarks concern illustrative examples and possibilities for the explicit verification of $(A2)^*$.

A basic tool for extending the above proposition to models with random matrix A is to use conditional and marginal distributions when representing the integral in (2.2):

$$(2.8)\quad Q(x) = \int_{I\!\!R^{ms}} \int_{I\!\!R^s} \Phi(z - Ax)\ \mu_1^2(A, dz)\mu_2(dA).$$

Here, $\mu_1^2(A, .)$ denotes the (regular) conditional distribution of z given A and μ_2 the marginal distribution of μ with respect to A ([24]).
To formulate the next result we split the matrix $A = (A_0, A_1)$ into a matrix A_0 whose columns are non-random and a matrix A_1 where random entries occur. We will use the symbol E for integration with respect to μ_2.

Proposition 1.2.5 (Theorem 2.3.1 in Chapter 2)
Assume (A1) - (A3), let $V \subset I\!\!R^m$ be non-empty, convex and suppose

(2.9) *for μ_2-almost all $A \in I\!\!R^{ms}$ the function*
$$\tilde{Q}_A(\chi) := \int_{I\!\!R^s} \Phi(z - \chi)\mu_1^2(A, dz)$$
is strongly convex on $A(V)$ with some modulus $\kappa(A)$, and there exists some $\kappa > 0$ such that $\kappa(A) \geq \kappa$ for μ_2-almost all A;

(2.10) $E(\|A\|^2) < +\infty$;

(2.11) *the matrix A_1 contains at least $m - s$ columns;*

(2.12) *A_o has full rank;*

(2.13) *the matrix $E(A_1^T A_1) - E(A_1)^T E(A_1)$ is positive-definite.*

Then Q is strongly convex on V.

We will comment on specific situations where the verification of the above conditions (in particular (2.9) and (2.13)) is especially simple. If, for instance, z and A are stochastically independent then $\mu_1^2(A,.)$ coincides with the marginal distribution μ_1, and Proposition 1.2.4 can be employed. If in any row of A_1 the random entries are pairwise uncorrelated then the matrix in (2.13) is diagonal with positive entries along the main diagonal.

Propositions 1.2.4 and 1.2.5 supplement known facts about the structure of Q by properties that are comparatively easy to verify and use only mild smoothness assumptions. In Chapter 3 of this thesis we will present quantitative results on the stability of optimal solutions to (2.1) - (2.3) where the strong convexity of $\tilde{Q}$ occurs as the essential assumption.

1.2.2 Mixed-Integer Linear Recourse

By introducing integrality constraints into the second stage we extend the two-stage model from the previous subsection:

$$(2.14)\ \min\{c^T x + Q(x) :\ x \in C\}$$

where

$$(2.15)\ Q(x) = \int_{\mathbb{R}^S} \Phi(z - Ax)\ \mu(d(z, A))$$

and

$$(2.16)\ \Phi(t) = \min\{q^T y + q'^T y' :\ Wy + W'y' = t,\ y' \geq 0,\ y \geq 0,\ y' \in \mathbb{R}^{m'},\ y \in \mathbb{Z}^{\bar{m}}\}.$$

The relevance of the above model was already indicated in our discussion of modelling alternatives. Integer variables in the second stage become indispensable if decisions are Boolean or restricted to multiples of basic quantities. The reason for considering integer variables in the second stage only is that, when occurring in the first stage, they can (if at all) be dealt with by means of integer and combinatorial optimization. In this context we refer to a paper by Wollmer ([135]). Integrality in the second stage will, obviously, destroy the beneficial structure of Φ met in linear recourse models. Therefore, again one has to think about the structure of the expected recourse function Q.

For the first time, a model fitting into (2.14) - (2.16) was studied by Stougie ([118], see also [83]) who gave a sufficient condition for the continuity of Q when the second stage is the following pure-integer linear program

$$\min\{q^T y : \; Wy \geq t, \; y \geq 0, \; y \in \mathbb{Z}^{\bar{m}}\}.$$

Further related contributions are due to Artstein/Wets [1], where stability issues are addressed to which we will come back in Section 1.3, and to Klein Haneveld, Louveaux, Stougie and van der Vlerk ([60],[61],[66]) who considered the case of "simple integer recourse" where the second stage reads

$$\text{(2.17)} \min\{{q^+}^T y^+ + {q^-}^T y^- : y^+ \geq t, \, y^- \geq -t, \, y^+ \geq 0, \, y^- \geq 0, \, y^+ \in \mathbb{Z}^{\bar{m}}, \, y^- \in \mathbb{Z}^{\bar{m}}\}.$$

The lack of structure in the case of a mixed-integer linear second stage is illustrated just by the very simple example where

$$Q(x) = \int_{\mathbb{R}} \Phi(z-x)\mu(dz) \quad \text{and} \quad \Phi(t) = \min\{y : y \geq t, \; y \in \mathbb{Z}\}.$$

Here, $\Phi(t) = \lceil t \rceil$ where the symbol $\lceil . \rceil$ denotes the integer round-up operation. Of course, Φ is discontinuous and this property is preserved for Q for any discrete measure μ. Hence, Q also cannot be convex for discrete μ.

Missing continuity and convexity of Q result from the comparatively poor structure of the value function Φ. Basic facts on the value function of a mixed-integer linear program are contained in [4],[5],[12]. If we assume that W and W' are rational matrices and that $\Phi(t) \in \mathbb{R}$ for all $t \in \mathbb{R}^s$, then it holds:

Proposition 1.2.6 ([5], Theorem 8.1; [12], Theorem 2.1)
There exist constants $\alpha > 0$, $\beta > 0$ such that for all $t', t'' \in \mathbb{R}^s$ we have

$$|\Phi(t') - \Phi(t'')| \leq \alpha \|t' - t''\| + \beta.$$

Proposition 1.2.7 ([12], Theorem 3.3)
There exist constants $\gamma > 0$, $\delta > 0$ and vectors $d_1, \ldots, d_\ell \in \mathbb{R}^s$, $\tilde{d}_1, \ldots, \tilde{d}_{\ell'} \in \mathbb{R}^s$ such that for all $t \in \mathbb{R}^s$

$$\Phi(t) = \min_y \{q^T y + \max_{j \in \{1,\ldots,\ell\}} d_j^T (t - Wy) : \; y \in Y(t)\}$$

where

$$Y(t) = \{y \in \mathbb{Z}^{\bar{m}} : \quad y \geq 0,\ \sum |y_i| \leq \gamma \sum |b_r| + \delta, \\ \tilde{d}_k^T(t - Wy) \geq 0,\ k = 1, \ldots, \ell'\}.$$

These statements imply (cf. Chapter 4) that Φ is lower semicontinuous on $\mathbb{R}^s$ and that the discontinuity points of Φ are contained in (countably many) translates of hyperplanes given by the facets of the polyhedral cone $W'(\mathbb{R}_+^{m'})$. Moreover, Φ is Lipschitz continuous with a uniform modulus on each of the connected components of the set of continuity points.

Together with the rationality of W and W' the following basic asssumptions ensure that Q is well defined.

$(\mathrm{A1})_{int}$ For all $t \in \mathbb{R}^s$ there exist $y \in \mathbb{Z}^{\bar{m}}$, $y' \in \mathbb{R}^{m'}$ such that $y \geq 0$, $y' \geq 0$ and $Wy + W'y' = t$.

$(\mathrm{A2})_{int}$ $\{u \in \mathbb{R}^s : W^T u \leq q, {W'}^T u \leq q'\} \neq \emptyset$

$(\mathrm{A3})_{int}$ $\int\limits_{\mathbb{R}^S} (\|z\| + \|A\|)\mu(d(z, A)) < +\infty.$

Note that the above assumptions are natural extensions of those for linear recourse. Indeed, $(\mathrm{A1})_{int}$ says that there is always a feasible second-stage action and $(\mathrm{A2})_{int}$ then guarantees the solvability of the second-stage problem. Using Proposition 1.2.6 it can be shown that a finite first moment of μ ensures finiteness of the integral defining $Q(x)$.

These prerequisites together with Fatou's lemma and Lebesgue's dominated convergence theorem will lead us in Chapter 4 to the following basic continuity results for Q. In particular, Proposition 1.2.9 extends the above mentioned continuity statement in [118].

Proposition 1.2.8 (Proposition 4.3.1 in Chapter 4)
Assume $(A1)_{int}$ – $(A3)_{int}$, then Q is a real-valued lower semicontinuous function on $\mathbb{R}^m$.

By $E(x)$ we denote the set of all those $(z, A) \in \mathbb{R}^S$ such that Φ is discontinuous at $z - Ax$.

Proposition 1.2.9 (Proposition 4.3.2 in Chapter 4)
Assume $(A1)_{int}$ – $(A3)_{int}$ *and let* $x \in I\!R^m$ *be such that* $\mu(E(x)) = 0$*, then* Q *is continuous at* x*.*

Since the discontinuities of Φ are concentrated in a set of Lebesgue measure zero the above proposition in particular says that Q is continuous on $I\!R^m$ if, for μ_2-almost all $A \in I\!R^{ms}$, the conditional distribution $\mu_1^2(A,.)$ of z given A is absolutely continuous with respect to the Lebesgue measure on $I\!R^s$. To find sufficient conditions for the Lipschitz continuity of Q we again resort to an iterated-integral representation as in (2.8) and start with the case where A is non-random. It holds

Proposition 1.2.10 (Proposition 4.3.6 in Chapter 4, A non-random)
Assume $(A1)_{int}$ – $(A3)_{int}$ *and that* μ *is absolutely continuous with respect to the Lebesgue measure on* $I\!R^s$*. Assume further that for any non-singular linear transformation* $B \in L(I\!R^s, I\!R^s)$ *the one-dimensional marginal distributions of* $\mu \circ B$ *have bounded densities which, outside some bounded interval, are monotonically decreasing with growing absolute value of the argument.*
Then Q *is Lipschitz continuous on any bounded subset of* $I\!R^m$*.*

Examples in Chapter 4 will show that the rather technical boundedness and monotonicity assumptions above are indispensable.
Via a representation like in (2.8) the result is extended to models with random A (Proposition 4.3.10 in Chapter 4). The assumptions then are very similar but in terms of $\mu_1^2(A,.)$ and they have to hold uniformly for μ_2-almost all $A \in I\!R^{ms}$.
A class of measures fulfilling the assumptions in Proposition 1.2.10 is given by r-convex measures ($r \in (-\infty, 0]$),[13], whose support is the whole of $I\!R^s$ (Proposition 4.3.9 in Chapter 4). This class includes, for instance, the nondegenerate multivariate normal distribution and the t-distribution.

Presently, statements beyond the above propositions are only available for the (essentially one-dimensional) case of simple integer recourse introduced in (2.17). Such statements include sufficient conditions for the differentiability of Q ([60],[66]) and descriptions of the convex hull of the epigraph of Q ([61]).

1.3 Stability of Two-Stage Stochastic Programs

In two-stage stochastic programming, function values and (provided they do exist) gradients of the objective are given by multidimensional integrals that often can not be computed explicitly. Moreover, the modelling is based on the *precise* knowledge of the underlying probability distribution which, of course, is quite often a very optimistic assumption. These observations lead to the question whether a two-stage (or a general) stochastic program behaves stable under perturbations of the underlying probability measure.
Indeed, if computationally intractable integrals arise one could try to approximate the given probability measure by "simpler" ones with favourable numerical properties. If one has only partial knowledge on the involved probability measure the model has to be based on an approximation of the "true" measure.

In both situations the question arises whether one could rely on the approximation, i.e. whether "small" perturbations of the measure lead to only "small" changes in the optimal value and the solution set. For abstract optimization problems this issue is addressed in parametric optimization ([4],[34]). Of course, general results from this field do also underly the stability analysis in (two-stage) stochastic programming. But what deserves particular attention is the existing specific structure:

- higher-order differentiability of the data and unique solvability of the problem are rather exceptional,

- to cover important applications it is advisable to select the parameter space as an abstract (non-linear) space of probability measures,

- continuity of model data jointly in the decision variable and the perturbation parameter, which is essential for many abstract stability results, requires some effort to be verified for stochastic programs,

- when aiming at quantitative stability one has to find probability distances that are properly adjusted to the specific problem classes.

In this thesis we present stability results for the two-stage stochastic programs introduced in Section 1.2. For linear recourse models these results extend an existing theory. The results for mixed-integer recourse seem to be among the first such findings in the literature.

1.3.1 Linear Recourse

Consider the following two-stage stochastic program where the integrating probability measure μ occurs as a perturbation parameter

(3.1) $\mathrm{P}(\mu) \quad \min\{g(x) + Q(x,\mu) : x \in C\}$

where

(3.2) $Q(x,\mu) := \int_{\mathbb{R}^S} \Phi(z - Ax)\mu(d(z,A))$

and

(3.3) $\Phi(t) := \min\{q^T y : Wy = t, y \geq 0\}$.

In (3.1) we assume that g is a convex function on $\mathbb{R}^m$ and $C \subset \mathbb{R}^m$ is a non-empty closed convex set. This differs a little from the setting in (2.1) - (2.3), but many stability results remain valid under these more general assumptions, too.
As indicated we will consider (3.1) as an optimization problem varying with the parameter μ. In the first approaches in the literature, only (Euclidean) parameters of μ instead of the entire measure μ were exposed to perturbations ([25],[40],[129]). The authors then employ results on the sensitivity of nonlinear programs (cf. [34] and the references therein). These results, however, rely on smoothness and uniqueness assumptions which considerably restrict the applicability to recourse models. Moreover, the Euclidean setting for the parameter space excludes important applications such as estimation via empirical measures. The first papers considering μ as a parameter varying in a suitable topological (or metric) space of probability measures are due to Kall [48], Robinson/Wets [88] and Römisch/Wakolbinger [98].

To specify the space of admissible perturbation parameters let us consider the set $\mathcal{P}(\mathbb{R}^S)$ of all Borel probability measures on $\mathbb{R}^S$ and equip it with weak convergence of probability measures. 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 $h : \mathbb{R}^S \to \mathbb{R}$ we have

$$\int\limits_{\mathbb{R}^S} h(\xi)\mu_n(d\xi) \to \int\limits_{\mathbb{R}^S} h(\xi)\mu(d\xi) \quad \text{as} \quad n \to \infty.$$

In the present thesis, we will employ this notion only for probability measures on a Euclidean space. The notion can be extended to probability measures on more general spaces (e.g. separable metric spaces). A basic reference for weak convergence of probability measures is the monograph by Billingsley [9].

By its topological nature, weak convergence of probability measures is the weak* convergence on $\mathcal{P}(I\!R^S)$ seen as a subset of the space of finite signed measures - the topological dual to the space of bounded continuous functions on $I\!R^S$. In finite dimension, weak convergence of probability measures coincides with the pointwise convergence of distribution functions at all continuity points of the limiting distribution function. It covers important specific modes of convergence such as almost surely converging densities (Scheffé's theorem [9]), empirical measures (Glivenko-Cantelli almost sure uniform convergence [76]) and discretizations via conditional expectations (see [10] and [49] for details). Therefore, our parameter space is, on the one hand, sufficiently comprehensible to cover relevant instances and, on the other hand, sufficiently rich in structure for substantial statements.

The stability of the model (3.1) will be studied in terms of the optimal-value function φ assigning to $\mu \in \mathcal{P}(I\!R^S)$ the global optimal value of $\mathrm{P}(\mu)$ and in terms of the solution set mapping ψ assigning to $\mu \in \mathcal{P}(I\!R^S)$ the set of global optimal solutions. For linear recourse models the inherent convexity allows us to confine our considerations to global optimal values and solutions.
The subsequent stability results are statements on qualitative and quantitative continuity of the above mappings. Differentiability is not addressed here. It is treated, for instance, in [28], [53], [96], [111], [112], [113], [115].

Fairly general qualitative stability results for two-stage stochastic programs with linear second stage were established by Kall and Robinson/Wets in [48], [88]. In these papers, the setting is slightly more general than here. We will quote from them in a form fitting the present setting.
For notational convenience we introduce the following subset of probability measures

$$\Delta_{p,K}(I\!R^S) = \{\nu \in \mathcal{P}(I\!R^S) : \int\limits_{I\!R^S} \|\xi\|^p \nu(d\xi) \leq K\}$$

where $p > 1$ and $K > 0$ are fixed real numbers.

Proposition 1.3.1 ([48], [88])
Assume (A1), (A2), let $\mu \in \Delta_{p,K}(I\!R^S)$ for some $p > 1$, $K > 0$. Suppose further that $\psi(\mu)$ is non-empty and bounded.
Then

(i) *the function φ (from $\Delta_{p,K}(I\!R^S)$ to $I\!R$) is continuous at μ;*

(ii) the multifunction ψ (from $\Delta_{p,K}(I\!R^S)$ to $I\!R^m$) is Berge upper semicontinuous at μ, i.e. for each open set $V \subset I\!R^m$ with $V \supset \psi(\mu)$ there exists a neighbourhood $\mathcal{N}$ of μ in $\Delta_{p,K}(I\!R^S)$ such that $\psi(\nu) \subset V$ for each $\nu \in \mathcal{N}$;

(iii) there exists a neighbourhood $\mathcal{N}$ of μ in $\Delta_{p,K}(I\!R^S)$ such that for all $\nu \in \mathcal{N}$ the set $\psi(\nu)$ is non-empty.

The assumption that $\mu \in \Delta_{p,K}(I\!R^S)$ for some $p > 1$, $K > 0$ and the fact that φ and ψ have the above properties as mappings acting on the subset $\Delta_{p,K}(I\!R^S)$ (instead of the whole space $\mathcal{P}(I\!R^S)$) express some uniform integrability which is indispensable according to an example in [88]. In [48], [88] this uniform integrability is formulated in a different way: the authors call a family $\mathcal{F}$ of measurable, real-valued functions on $I\!R^S$ uniformly integrable with respect to a subset $\mathcal{P}_0$ of $\mathcal{P}(I\!R^S)$ if for any $\varepsilon > 0$ there exists a compact set $C_\varepsilon \subset I\!R^S$ such that

$$\int_{I\!R^S \setminus C_\varepsilon} |f(\xi)|\nu(d\xi) < \varepsilon \quad \text{for all } \; f \in \mathcal{F} \; \text{ and all } \; \nu \in \mathcal{P}_0 \, .$$

For the stability results it is claimed that, for some bounded open set $V \supset \psi(\mu)$, the family

$$\mathcal{F} = \{\Phi(z - Ax) : x \in clV\}$$

is uniformly integrable with respect to a neighbourhood $\mathcal{N}$ of μ in $\mathcal{P}(I\!R^S)$. It can be shown that the latter is implied by the assumptions in Proposition 1.3.1.

Approaches to quantitative stability of $\mathrm{P}(\mu)$ rely on estimating differences of function values and/or gradients of Q by distances of probability measures. Beside finding a probability distance that allows an estimate at all one also wishes to "optimize" the selection, i.e. one is aimimg at weakest possible assumptions and best possible convergence rates.

The first quantitative results are due to Römisch/Wakolbinger [98], who proved a Hölder continuity result for optimal values which is based on the bounded Lipschitz metric β ([22], [81]). This result was re-established in [92] under slightly weaker moment conditions on the underlying probability measures. Moreover, a first (Hölder) estimate (in terms of β) for the Hausdorff distance of solution sets was derived in [92].
In [93] these estimates were improved by using L_p-Wasserstein distances ([81]). Moment conditions were further relaxed and Lipschitz instead of Hölder estimates were

obtained.

Given $\mu, \nu \in \mathcal{P}(I\!R^S)$, their L_p-Wasserstein distance W_p is defined as

$$W_p(\mu, \nu) = \Big[\inf \Big\{ \int\limits_{I\!R^S \times I\!R^S} \|\xi - \tilde{\xi}\|^p \eta(d\xi, d\tilde{\xi}) : \ \eta \in D(\mu, \nu) \Big\} \Big]^{1/p}$$

for all

$$\mu, \nu \in \mathfrak{M}_p(I\!R^S) = \{\mu' \in \mathcal{P}(I\!R^S) : \int\limits_{I\!R^S} \|\xi\|^p \mu'(d\xi) < +\infty\}$$

where

$$D(\mu, \nu) = \{\eta \in \mathcal{P}(I\!R^S \times I\!R^S) : \eta \circ \pi_1^{-1} = \mu, \ \eta \circ \pi_2^{-1} = \nu\}$$

and π_1, π_2 are the first and second projections, respectively.

In [80] it is shown that $(\mathfrak{M}_p(I\!R^S), W_p)$ is a metric space and the following equivalence is established: A sequence $\{\mu_n\}$ of probability measures in $\mathcal{P}(I\!R^S)$ converges in W_p to $\mu \in \mathfrak{M}_p(I\!R^S)$ if and only if $\{\mu_n\}$ converges weakly to μ and

$$\lim_{n \to \infty} \int\limits_{I\!R^S} \|\xi\|^p \mu_n(d\xi) = \int\limits_{I\!R^S} \|\xi\|^p \mu(d\xi).$$

Using W_p it is possible to obtain the following Lipschitz estimate for expectation functions whose integrands are Lipschitzian on bounded sets. Since Φ is even globally Lipschitzian (but, with random second-stage costs, in general only Lipschitzian on bounded sets) the estimate fits to what we encounter in stochastic programs with linear recourse.
Consider a real-valued function h on $I\!R^S$ which is Lipschitzian on bounded sets, i.e.

$$L_h(r) := \sup \Big\{ \frac{|h(\xi) - h(\tilde{\xi})|}{\|\xi - \tilde{\xi}\|} \ : \ \|\xi\|, \|\tilde{\xi}\| \leq r; \ \xi \neq \tilde{\xi} \Big\} \ < \ +\infty$$

for each $r > 0$.
Then it holds

Proposition 1.3.2 ([93])
Let $h : I\!R^S \to I\!R$ be Lipschitzian on bounded sets. Then for all $\mu, \nu \in \mathfrak{M}_p(I\!R^S)$

$$\Big| \int\limits_{I\!R^S} h(\xi)\mu(d\xi) - \int\limits_{I\!R^S} h(\xi)\nu(d\xi) \Big| \ \leq \ (M_q(\mu) + M_q(\nu)) \cdot W_p(\mu, \nu)$$

where

$$p > 1, \quad 1/p + 1/q = 1 \quad \textit{and} \quad M_q(\mu) := \Big(\int\limits_{I\!R^S} L_h(\|\xi\|)^q \mu(d\xi) \Big)^{1/q}.$$

Moreover, for globally Lipschitzian h, i.e. if $L_h(r) \leq L_h = const.$ for all $r > 0$, it holds for all $\mu, \nu \in \mathfrak{M}_1(I\!R^S)$

$$| \int\limits_{I\!R^S} h(\xi)\mu(d\xi) - \int\limits_{I\!R^S} h(\xi)\nu(d\xi)| \quad \leq \quad L_h \cdot W_1(\mu, \nu).$$

Using well-known facts from parametric optimization the above proposition leads to a Lipschitz estimate for optimal values.

Proposition 1.3.3 ([93])
Assume (A1) - (A3) and let $\psi(\mu)$ be non-empty and bounded. Then there exist constants $L > 0, \delta > 0$ such that

$$|\varphi(\mu) - \varphi(\nu)| \quad \leq \quad L \cdot W_1(\mu, \nu)$$

whenever $\nu \in \mathfrak{M}_1(I\!R^S), W_1(\mu, \nu) < \delta$.

Note that, compared to Proposition 1.3.1, the uniform integrability has only virtually vanished in the above statement. Recall that convergence in W_1 implies weak convergence and convergence of the first moments. Due to Theorem 5.4 in [9] this implies uniform integrability of the family $\{\|z\| + \|A\|\|x\| \; : \; x \in clV\}$ with respect to the members of any weakly convergent sequence $\mu_n \xrightarrow{w} \mu$. From this the uniform integrability of the family $\{\Phi(z - Ax) : x \in clV\}$ with respect to the same set of measures follows (cf. [48], [88]).

Proposition 1.3.3 essentially settles the quantitative continuity for optimal values of P(μ). Using the first part of Proposition 1.3.2 in [93] a similar result was derived for linear recourse models where, moreover, the second-stage costs q are random.

More complex assumptions are needed for the quantitive stability of optimal solutions. In [93] the following is shown for the Hausdorff distance of solution sets in models with non-random A.

Proposition 1.3.4 *Let in P(μ) the matrix A be non-random, g be a convex quadratic function and C be a polyhedron. Assume (A1) -(A3) and let $\psi(\mu)$ be non-empty and bounded. Suppose further that the function*

$$\tilde{Q}(\chi, \mu) := \int\limits_{I\!R^s} \Phi(z - \chi)\mu(dz) \qquad \textit{(cf. (2.7))}$$

is strongly convex on a convex open set $V \subset I\!R^s$ containing $A(\psi(\mu))$.
Then there exist constants $L > 0, \delta > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \quad \leq \quad L \cdot W_1(\mu, \nu)^{1/2}$$

whenever $\nu \in \mathfrak{M}_1(I\!R^S), W_1(\mu, \nu) < \delta$.

Examples in [93] show that the above rate 1/2 is best possible and that the result is lost for a general convex function g or a general closed convex set C.

Note that the strong convexity discussed in the previous section now enters as a crucial assumption. This is not surprising, since in parametric optimization it is well known that the quantitative continuity of optimal solutions is governed by a proper conditioning of the unperturbed objective near the (unperturbed) set of optimal solutions ([2], [3], [58]). Strong convexity inherits a quadratic conditioning. Indeed, while the epigraph of a general convex function is supported by hyperplanes, the epigraph of a strongly convex function is even supported by level sets of quadratic functions with positive-definite quadratic form. When basing the stability analysis on estimates of objective function differences (cf. Proposition 1.3.2), the stability rate is gained as the inverse of the conditioning, hence quadratic conditioning yields the stability rate 1/2 ([3], [58]).

In Section 1.2 we already pointed out that the strong convexity of Q is a much too strong assumption for models with non-random A. Note that in the above proposition it is merely the strong convexity of $\tilde{Q}$ that is needed. Proposition 1.2.4 (Theorem 2.2.2 in Chapter 2) is a handy tool for its verification.
For models $\mathrm{P}(\mu)$, however, where both z and A are random the strong convexity of Q itself can be ensured (Proposition 1.2.5 above or Theorem 2.3.1 in Chapter 2). In Chapter 2, we then obtain the following stability result.

Proposition 1.3.5 (Theorem 2.4.3 in Chapter 2, z and A random)
Consider (3.1) - (3.3). Assume (A1) - (A3) and let $\psi(\mu)$ be non-empty, bounded. Let $V \subset I\!R^m$ be some bounded, open, convex set containing $\psi(\mu)$ and suppose that Q is strongly convex on V.
Then there exist constants $L > 0$, $\delta > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \leq L \cdot W_1(\mu, \nu)^{1/2}$$

whenever $\nu \in \mathfrak{M}_1(I\!R^S)$, $W_1(\mu, \nu) < \delta$.

Although it is shown by an example that the Hölder rates in the above propositions are best possible there remains the question whether improved rates can be established for subclasses of perturbations. In Chapter 3 we pick up this question for models with non-random matrix A.

The stability results displayed up to now all rely on estimating differences of *function values* in the original and the perturbed models. In Chapter 3 we estimate differences of *(sub-)gradients* instead. In [114], [116] Shapiro has established a variational principle for optimal solutions to abstract parametric programs that is based on estimating differences of difference quotients. This led us to setting up a probability (pseudo-)distance incorporating subgradients of the function $\tilde{Q}$. Namely, given $\mu, \nu \in \mathfrak{M}_1(I\!R^s)$ and a fixed non-empty closed convex set $U \in I\!R^m$ we define

$$d(\mu, \nu; U) = \sup\{\|z^*\| : z^* \in \partial(\tilde{Q}_\nu - \tilde{Q}_\mu)(Ax),\ x \in U\}.$$

(For notational convenience, here and in Chapter 3, the dependence of $\tilde{Q}$ on the integrating measure is indicated by a subscript, thus $\tilde{Q}_\nu(.) = \tilde{Q}(., \nu)$.)
The symbol "∂" denotes Clarke's subdifferential ([15]) which is well defined since $\tilde{Q}_\nu - \tilde{Q}_\mu$ is locally Lipschitzian. Moreover, $d(\cdot, \cdot; U)$ is only a pseudo-metric on $\mathfrak{M}_1(I\!R^s)$ since $d(\mu, \nu; U) = 0$ is possible for $\mu \neq \nu$.
In Chapter 3 we first prove persistence and upper semicontinuity of optimal solutions to P(μ) when the convergence of the underlying measures is put in terms of $d(\cdot, \cdot; U)$ (Proposition 3.2.3 in Chapter 3). Then we address the key issue - a *Lipschitz* estimate for the Hausdorff distance of solution sets that is based on $d(\cdot, \cdot; U)$:

Proposition 1.3.6 (Theorem 3.2.4 in Chapter 3)
Let in P(μ) the matrix A be non-random, g be a convex quadratic function and C be a polyhedron. Assume (A1) -(A3) and let $\psi(\mu)$ be non-empty and bounded. Suppose further that the function $\tilde{Q}_\mu$ is strongly convex on a convex open set $V \subset I\!R^s$ containing $A(\psi(\mu))$. Let $U = clU_0$, where U_0 is an open, convex, bounded set such that $\psi(\mu) \subset U_0$ and $A(U) \subset V$. Then there exist constants $L > 0$, $\delta > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \leq L \cdot d(\mu, \nu; U)$$

whenever $\nu \in \mathfrak{M}_1(I\!R^s)$, $d(\mu, \nu; U) < \delta$.

When wishing to apply this result one faces the problem that, not surprisingly, in the literature there are no estimates in terms of $d(\cdot, \cdot; U)$ for specific instances of weakly converging probability measures. Therefore, we derive an upper estimate of $d(\cdot, \cdot; U)$ by the uniform (or Kolmogorov-Smirnov) distance of distribution functions which leads to the following.

Proposition 1.3.7 (Corollary 2.5 in Chapter 3)
Adopt the setting of Proposition 1.3.6.
Then there exist non-singular matrices $B_i (i = 1, \ldots, \ell)$ and a constant $L > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \leq L \sum_{i=1}^{\ell} \sup_{t \in A(U)} |F_{\mu \circ (-B_i)}(-B_i^{-1} t) - F_{\nu \circ (-B_i)}(-B_i^{-1} t)|$$

whenever $\nu \in \mathfrak{M}_1(I\!R^s)$ is chosen such that the right-hand side is sufficiently small.

The above matrices $B_i (i = 1, \ldots, \ell)$ form a (minimal) collection of basis submatrices of W such that each normal cone to a vertex of M_D is the union of suitable cones $B_i(I\!R^s_+)$ (cf. Section 1.2 and Proposition 2.2.1 in Chapter 2). $F_{\mu \circ (-B_i)}$ then denotes the distribution function of the (image) measure $\mu \circ (-B_i)$.

Propositions 1.3.4, 1.3.6, 1.3.7 work under mild smoothness assumptions and do not require the uniqueness of optimal solutions. Note that convergence of subgradients does not imply convergence of function values. Therefore it is not surprising that our assumptions do not guarantee convergence of optimal values (cf. Example 2.12 in Chapter 3).
In a very similar setting, Shapiro ([116]) has an upper Lipschitz continuity result for optimal solutions.

For the example showing sharpness of the rate in Proposition 1.3.4 both Propositions 1.3.4 and 1.3.6 finally lead to identical rates. Hence there is no general superiority of Proposition 1.3.6 over 1.3.4. However, for important specific modes of perturbation Proposition 1.3.6 yields better rates than Proposition 1.3.4. In Chapter 3 we will work out contaminated distributions and estimation via empirical measures. To give an idea, we close this subsection with a conclusion for empirical measures.

Let $\xi_1., \xi_2, \ldots, \xi_n, \ldots$ be independent $I\!R^s$-valued random variables on a probability space $(\Omega, \mathfrak{A}, \mathbf{P})$ having joint distribution μ. We consider the empirical measures

$$\mu_n(\omega) := \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i(\omega)} \qquad (\omega \in \Omega;\ n \in I\!N)$$

and we are interested in the asymptotic behaviour of the solution set $\psi(\mu_n(\cdot))$ of $P(\mu_n(\cdot))$ as n tends to infinity.
There is a rich literature on asymptotic properties of empirical measures (consult

[39], [117]). In particular, the following law of iterated logarithm for the Kolmogorov-Smirnov distance of distribution functions is known ([82])

$$(3.4)\quad \limsup_{n\to\infty} \Big(\frac{n}{2\log\log n}\Big)^{\frac{1}{2}} \sup_{t\in\mathbb{R}^s} |F_\mu(t) - F_{\mu_n(\omega)}(t)| \leq \frac{1}{2} \quad \mathbf{P}\text{-almost surely.}$$

Together with Proposition 1.3.7 this leads to the following speed-of-convergence result for optimal solutions.

Proposition 1.3.8 (Proposition 3.3.1 in Chapter 3)
Under the assumptions of Proposition 1.3.6 it holds

$$\limsup_{n\to\infty} \Big(\frac{2n}{\log\log n}\Big)^{\frac{1}{2}} \cdot d_H(\psi(\mu), \psi(\mu_n(\omega))) \leq L\ell \quad \mathbf{P} - \textit{almost surely},$$

where L and ℓ denote the Lipschitz modulus and the number of basis matrices, respectively, arising in Proposition 1.3.7.

This statement supplements the consistency results in [29], [54] by the rates of convergence for solution sets (without requiring unique solvability of $\mathrm{P}(\mu)$). Compared to considerations in [116] we do not need a linear-independence assumption imposed there.
Further results in Chapter 3 concern a large deviation estimate and an estimate for the asymptotic distribution of optimal solutions.

1.3.2 Mixed-Integer Recourse

In this subsection we review stability results for stochastic programs with mixed-integer linear recourse. The model, which will be denoted by $\mathrm{P}(\mu)_{int}$, is given as in (3.1) - (3.3) with the exception that the second stage is a mixed-integer linear program as in (2.16).

The expected recourse function Q is now typically non-convex (cf. Section 1.2) and it is appropriate to study the stability of *local* optima. This, however, causes an extra problem that has to be settled first.
Consider, for example, a real-valued function on $\mathbb{R}^m$ that is constant on some ball B_R with radius R. Of course, the ball $B_{R/2}$ (around the same point but with radius $R/2$) then is a bounded set of local minimizers. If we perturb the function by adding a non-constant linear function (with arbitrarily small norm) then none of the points

"in and near" $B_{R/2}$ remains locally optimal.

In [57], [86] a local stability analysis was proposed that excludes "pathological" sets of local minimizers such as the one above. The basic observation is that, when analyzing the stability of local solutions, one has to include all local minimizers that are , in some sense, nearby the minimizers one is interested in. In [57], [86], this leads to the concept of a complete local minimizing set (CLM set):
With some subset $V \subset I\!R^m$ we consider localized versions of the optimal-value function and the solution set mapping

$$\begin{aligned} \varphi_V(\mu) &:= \inf\{g(x) + Q(x,\mu) : x \in C \cap \operatorname{cl} V\} \\ \psi_V(\mu) &:= \{x \in C \cap \operatorname{cl} V : \ g(x) + Q(x,\mu) = \varphi_V(\mu)\}. \end{aligned}$$

Given $\mu \in \mathcal{P}(I\!R^S)$, a non-empty set $M \subset I\!R^m$ is then called a CLM set for $\mathrm{P}(\mu)_{int}$ with respect to an open set $V \subset I\!R^m$ if $M \subset V$ and $M = \psi_V(\mu)$.
Note that the set of global minimizers is always a CLM set. The subsequent local analysis, hence, readily extends to the global situation. Further examples for CLM sets are strict local minimizers. For more details we refer to [57], [86].

Recall that the convergence of expectations with weakly converging integrating probability measures was essential to obtain the continuity properties of Q needed for the analysis in the previous subsection. For linear recourse models the integrands in the mentioned expectations are globally Lipschitz continuous. Together with uniform integrability, this led to qualitative and, together with L_p-Wasserstein metrics, to quantitative stability statements.

As shown in Section 1.2, the relevant integrands for mixed-integer recourse are discontinuous. Therefore, none of the approaches to linear recourse models can be followed here. For qualitative statements a convergence theorem attributed to Rubin in the literature ([9], Theorem 5.5) will now be a proper tool. The theorem states that weak convergence of probability measures is preserved when passing to image measures, provided that the discontinuities of the involved transformations fall into a set with limit measure zero. Concerning quantitative statements we will show that a certain discrepancy (or variational distance of probability measures) allows Hölder estimates for Q. Here, a refined analysis of the value function of a mixed-integer linear program is necessary.

In the literature, qualitative continuity results for expectations with discontinuous integrands can be traced back to Langen [64]. We also mention a recent paper by

Artstein/Wets [1] tackling this problem in the context of stochastic programming. Without explicitly using Rubin's theorem the authors obtain results that are similar to ours. The quantitative statements below seem to belong to the first such statements in the literature.

To formulate our continuity results for Q we again employ the notation

$$\Delta_{p,K}(\mathbb{R}^S) = \{\nu \in \mathcal{P}(\mathbb{R}^S) : \int_{\mathbb{R}^S} \|\xi\|^p \nu(d\xi) \leq K\} \quad (p > 1,\ K > 0).$$

Proposition 1.3.9 (Proposition 4.3.11 in Chapter 4)
Assume (A1)$_{int}$, (A2)$_{int}$ and let $\mu \in \Delta_{p,K}(\mathbb{R}^S)$ for some $p > 1$, $K > 0$. If the conditional distribution $\mu_1^2(A,.)$ of z given A is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^s$ for μ_2-almost all $A \in \mathbb{R}^{ms}$, then Q, as a function from $\mathbb{R}^m \times \Delta_{p,K}(\mathbb{R}^S)$ to $\mathbb{R}$, is continuous on $\mathbb{R}^m \times \{\mu\}$.

Employing the concept of a CLM set and basic techniques from Berge's classical stability theory ([7], cf. also [4]) then yields the following qualitative stability result.

Proposition 1.3.10 (Proposition 4.4.1 in Chapter 4)
Assume (A1)$_{int}$, (A2)$_{int}$, let $\mu \in \Delta_{p,K}(\mathbb{R}^S)$ for some $p > 1$, $K > 0$ and let the conditional distribution $\mu_1^2(A,.)$ of z given A be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^s$ for μ_2-almost all $A \in \mathbb{R}^{ms}$. Suppose further that there exists a subset $M \subset \mathbb{R}^m$ which is a CLM set for $P(\mu)_{int}$ with respect to some bounded open set $V \subset \mathbb{R}^m$.
Then

- *(i) the function φ_V (from $\Delta_{p,K}(\mathbb{R}^S)$ to $\mathbb{R}$) is continuous at μ;*
- *(ii) the multifunction ψ_V (from $\Delta_{p,K}(\mathbb{R}^S)$ to $\mathbb{R}^m$) is Berge upper semicontinuous at μ,*
- *(iii) there exists a neighbourhood $\mathcal{N}$ of μ in $\Delta_{p,K}(\mathbb{R}^S)$ such that for all $\nu \in \mathcal{N}$ we have that $\psi_V(\nu)$ is a CLM set for $P(\nu)_{int}$ with respect to V.*

By Example 3.13 in Chapter 4 no Hölder continuity (at any rate) for Q can be expected when equipping $\mathcal{P}(\mathbb{R}^S)$ with the L_1-Wasserstein metric. In Chapter 5 we show that, for non-stochastic A, such a result can be established with respect to a suitable discrepancy (variational distance of probability measures).

Given a subclass $\mathcal{B}_o$ of Borel sets in $I\!R^s$, the discrepancy $\alpha_{\mathcal{B}_o}$ of $\mu, \nu \in \mathcal{P}(I\!R^s)$ is defined by

$$\alpha_{\mathcal{B}_o}(\mu,\nu) := \sup\{|\mu(B) - \nu(B)| : \; B \in \mathcal{B}_o\}.$$

(Note that the Kolmogorov-Smirnov distance is a discrepancy with $\mathcal{B}_o$ taken as the family of all lower left orthants in $I\!R^s$.)
In the present situation we take $\mathcal{B}_o$ as the class $\mathcal{B}_\mathcal{K}$ of all closed bounded polyhedra in $I\!R^s$ whose facets (i.e. $(s-1)$-dimensional faces) parallel a facet of $\mathcal{K} = W'(I\!R_+^{m'})$ or a facet of $\bigtimes_{i=1}^{s}[0,1]$.
This selection is motivated by properties of the second-stage value function Φ (cf. Section 5.2 in Chapter 5). In Chapter 5 we will also show that $\alpha_{\mathcal{B}_\mathcal{K}}$ is a metric and relate convergence in $\alpha_{\mathcal{B}_\mathcal{K}}$ to weak convergence of probability measures.
We will end up with the following Hölder estimate for the optimal-value function φ_V.

Proposition 1.3.11 (Theorem 5.4.1 in Chapter 5) *Assume (A1)$_{int}$, (A2)$_{int}$, let $\mu \in \mathcal{P}(I\!R^s)$ be absolutely continuous with respect to the Lebesgue measure on $I\!R^s$ and let there exist constants $p > 1$ and $K > 0$ such that $\mu \in \Delta_{p,K}(I\!R^s)$. Suppose further that there exists a subset $M \subset I\!R^m$ which is a CLM set for P(μ)$_{int}$ with respect to some bounded open set $V \subset I\!R^m$.*
Then there exist constants $L > 0$ and $\delta > 0$ such that

$$|\varphi_V(\mu) - \varphi_V(\nu)| \le L \cdot \alpha_{\mathcal{B}_\mathcal{K}}(\mu,\nu)^{\frac{p-1}{p(s+1)}}$$

whenever $\nu \in \Delta_{p,K}(I\!R^s)$, $\alpha_{\mathcal{B}_\mathcal{K}}(\mu,\nu) < \delta$.

The convergence rate in the above statement improves if the second stage in P$(\mu)_{int}$ is a pure integer linear program.

Since, up to location in parallel hyperplanes, there are only finitely many facets occuring in the elements of $\mathcal{B}_\mathcal{K}$, the family $\mathcal{B}_\mathcal{K}$ forms a Vapnik-Červonenkis class (VČ class) of Borel sets in $I\!R^s$ ([122], [76], [117]). The law of iterated logarithm in (3.4) extends to discrepancies $\alpha_{\mathcal{B}_o}$ where $\mathcal{B}_o$ is a VČ class ([62]). Therefore, Proposition 1.3.11 leads to the following speed-of-asymptotic-convergence result for optimal values when estimating μ in P$(\mu)_{int}$ by empirical measures.

Proposition 1.3.12 (Proposition 5.5.5 in Chapter 5)
Assume (A1)$_{int}$, (A2)$_{int}$, let $\mu \in \Delta_{p,K}(I\!R^s)$ for some $p > 1$, $K > 0$ and suppose that

there exists a subset $M \subset I\!\!R^m$ which is a CLM set for $P(\mu)_{int}$ with respect to some bounded open set $V \in I\!\!R^m$.
Then there exists a constant $c > 0$ such that for P-almost all $\omega \in \Omega$

$$\limsup_{n\to\infty} \Big(\frac{n}{2\log\log n}\Big)^{\frac{1}{2}\cdot\frac{p-1}{p(s+1)}} |\varphi_V(\mu_n(\omega)) - \varphi_V(\mu)| \le c.$$

Further asymptotic-convergence results in Chapter 5 concern the continuity of optimal values and the upper semicontinuity of optimal solutions.

Chapter 2

Strong Convexity in Stochastic Programs with Complete Recourse

Abstract
For stochastic programs with complete (linear) recourse we present easily verifiable sufficient conditions for the strong convexity of the expected-recourse function. Both, programs with random right-hand side and with random technology matrix are considered. Among the implications of strong convexity those with respect to the stability of stochastic programs are worked out in detail. In this way, former results on the quantitative stability of optimal solutions are extended.

2.1 Introduction

We are interested in deriving strengthened versions of convexity (mainly strong convexity) for the objective function in a stochastic program with complete (linear) recourse. The latter is given by

(1.1) $\min\{g(x) + Q(x) : \; x \in C\}$,

where

(1.2) $Q(x) = \int\limits_{\Omega} \Phi(z(\omega) - A(\omega)x) P(d\omega)$

and

(1.3) $\Phi(t) = \min\{q^T y : Wy = t, \; y \geq 0\}$.

Here, $g : I\!R^m \to I\!R$ is a convex function and $C \subset I\!R^m$ is a non-empty closed convex set. Furthermore, we have some probability space $(\Omega, \mathfrak{A}, P)$ and measurable mappings $z : \Omega \to I\!R^s$, $A : \Omega \to I\!R^{ms}$. The images of A are understood as $s \times m$ matrices. Often it will be convenient to consider the probability measure $\mu := P \circ (z, A)^{-1}$, which is induced on $I\!R^{(m+1)s}$ by the mapping $(z, A) : \Omega \to I\!R^s \times R^{ms}$. Finally, $q \in I\!R^{\bar{m}}$ and $W \in L(I\!R^{\bar{m}}, I\!R^s)$ are a fixed vector and matrix, respectively.
The following hypotheses ensure that the function Q in (1.2) is well defined (for details see [46], [131]):

(A1) (complete recourse) for each $t \in I\!R^s$ there exists some $y \in I\!R^{\bar{m}}_+$ such that $Wy = t$,

(A2) (dual feasibility) there exists some $u \in I\!R^s$ such that $W^T u \leq q$,

(A3) (finite first moment) $\int\limits_{I\!R^{(m+1)s}} (\|z\| + \|A\|)\mu(d(z, A)) < +\infty.$

In connection with a vector, $\|.\|$ denotes the Euclidean norm. In connection with a matrix, it denotes the induced matrix norm.
Stochastic programs with (complete) recourse arise in the modelling of two-stage optimization processes, where infeasibilities caused by the uncertainty of data in the first stage can be compensated in a second stage after realizing the random data.
In the present chapter we place the main accent on studying analytical properties of the expected-recourse function Q in (1.2). From the literature ([46], [131]) it is well-known that Q is a convex function on $I\!R^m$, provided (A1) - (A3) hold. Furthermore, Q is continuously differentiable if μ has some continuity properties. Higher-order differentiability of Q has been investigated, too ([68], [129]).
Our aim is to develop sufficient conditions for the strong convexity of Q, i.e. given some convex subset $V \subset I\!R^m$ there exists some $\kappa > 0$ such that for all $x_1, x_2 \in V$ and all $\lambda \in [0, 1]$

$$(1.4) \qquad Q(\lambda x_1 + (1-\lambda)x_2) \leq \lambda Q(x_1) + (1-\lambda)Q(x_2) - \kappa\lambda(1-\lambda)\|x_1 - x_2\|^2.$$

This notion plays an important role for quantitative investigations in convex analysis. Provided that Q is sufficiently smooth, strong convexity is equivalent to strong monotonicity of the gradient and positive definiteness of the Hessian, respectively ([74], [75], [123]). In our analysis the former equivalence will be used to verify strong convexity. The following motivates our interest in strong convexity:

(i) Strongly convex functions obey nice properties when being minimized with standard methods of non-linear programming (e.g. if Q has a Lipschitzian gradient, then the steepest-descent method is linearly convergent). Of course, difficulties in computing the integral in (1.2) prevent an application of such standard methods for the stochastic program (1.1). However, there exist numerical approaches to stochastic programming which "imitate" descent techniques by using estimated (sub)gradients instead of the true ones [43]. In this context, it is reasonable to presume a more rapid convergence when the original function Q is strongly convex.

(ii) If Q is strongly convex near the solution set of (1.1), then there is a unique minimizer x_* and for any feasible $x \in I\!R^m$ that is sufficiently close to x_* we have $Q(x) \geq Q(x_*) + \frac{1}{2}\kappa\|x - x_*\|^2$. The latter can be considered as a specific form of a well-conditioned (local) minimizer as studied for instance in [3]. Implications of such a conditioning for asymptotic properties of estimations in stochastic programming have recently been studied in [50]. For the uniqueness of the minimizer, of course, already strict convexity is sufficient. As a by-product of our analysis we will also obtain a sufficient condition for strict convexity.

(iii) The stability analysis of stochastic programs is another field where the strong convexity of Q (or suitable relaxations) turn out beneficial ([27], [93], [116]). A number of quantitative stability results for optimal solutions are based on the strong convexity of the objective function in the unperturbed problem.

The present chapter contributes to the above lines of research by establishing easily verifiable sufficient conditions for strong convexity in terms of the data in (1.1) - (1.3) (Sections 2.2 and 2.3). Hence, it is possible to check right from the model (1.1) - (1.3) whether it has favourable numerical properties, is well-conditioned in some sense or behaves stable under perturbations. In Section 2.4 we present implications of the structural results on Q for the stability of the model (1.1) - (1.3) when subjecting the underlying probability measure μ to perturbations. The latter is motivated both by numerical reasons and by often having only partial information on the "true" measure μ (cf. [27], [48], [88], [116], [124]).

2.2 Strong Convexity - Non-Stochastic Technology Matrix

Quite often, the model (1.1) - (1.3) is studied in the specific situation where the matrix A is non-random, or formally, where $A : \Omega \to I\!R^{ms}$ is constant P-almost surely. Then, of course, the function Q from (1.2) is constant on translates of the null space of A. If the latter is non-trivial, one easily confirms that Q cannot be strongly convex. This leads to analyzing the function

$$(2.1)\quad \tilde{Q}(\chi) := \int_{\mathbb{R}^s} \Phi(z-\chi)\mu(dz),$$

where $\chi \in \mathbb{R}^s$ and μ denotes the probability measure on $\mathbb{R}^s$ induced by $z : \Omega \to \mathbb{R}^s$. Strong convexity of $\tilde{Q}$ arises as an essential assumption in the stability analysis for stochastic programs ([93], [95] and Chapter 3) and, as it will turn out in Section 2.3, it is crucial for the strong convexity of Q if A is random.
We impose the basic hypotheses (A1) - (A3) and note that, here, (A3) reads

$$\text{(A3)} \qquad \int_{\mathbb{R}^s} \|z\|\mu(dz) < +\infty.$$

Furthermore, we assume

(A4) the probability measure μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^s$.

It is well-known ([46], [131]) that, given (A1) - (A4), the function $\tilde{Q}$ is continuously differentiable on $\mathbb{R}^s$. Then, $\tilde{Q}$ is strongly convex (with constant $\kappa > 0$) on some convex subset $V \subset \mathbb{R}^s$ if and only if

$$(\tilde{Q}'(\chi_1) - \tilde{Q}'(\chi_2))^T(\chi_1 - \chi_2) \geq 2\kappa\|\chi_1 - \chi_2\|^2$$

for all $\chi_1, \chi_2 \in V$ ([74], [123]).
Before presenting the main result of this section we collect a few prerequisites from linear parametric programming about the value function

$$\Phi(t) = \min\{q^T y : Wy = t,\, y \geq 0\}.$$

The following pair of dual linear programs is associated to Φ:

$$(2.2)\qquad \min\{q^T y :\ Wy = t,\ y \geq 0\},$$
$$(2.3)\qquad \max\{t^T u :\ W^T u \leq q\} \qquad .$$

By (A1) and (A2), both programs are solvable for any $t \in \mathbb{R}^s$. Furthermore, the feasible region M_D of (2.3) is compact and, therefore, it coincides with the convex hull of its vertices $\tilde{d}_1, \ldots, \tilde{d}_{\tilde{\ell}} \in \mathbb{R}^s$. According to Satz 6.7 in ([72], p. 156) and the Basis Decomposition Theorem in [126] the following holds

Proposition 2.2.1 *Assume (A1), (A2), then*

(i) $\Phi(t) = \max_{i=1,\ldots,\tilde{\ell}} \tilde{d}_i^T t$ *for all* $t \in I\!R^s$*;*

(ii) $\Phi(t) = \tilde{d}_i^T t$ *for all* $t \in \mathcal{K}_i$*, where* $\mathcal{K}_i$ *denotes the normal cone to* M_D *at* $\tilde{d}_i$*, i.e.* $\mathcal{K}_i = \{v \in I\!R^s : v^T(u - \tilde{d}_i) \leq 0$ *for all* $u \in M_D\}$ $(i = 1, \ldots, \tilde{\ell})$*;*

(iii) it holds that $\bigcup_{i=1}^{\tilde{\ell}} \mathcal{K}_i = I\!R^s$*, and for* $i_1 \neq i_2$ *the intersection* $\mathcal{K}_{i_1} \cap \mathcal{K}_{i_2}$ *coincides with a common closed face of dimension less than* s*;*

(iv) $\mathcal{K}_{i_1} \cap \mathcal{K}_{i_2}$ *has dimension* $s-1$ *if and only if the vertices* $\tilde{d}_{i_1}$ *and* $\tilde{d}_{i_2}$ *are adjacent;*

(v) each of the cones $\mathcal{K}_i$ *is a finite union of simplicial cones which can be represented as* $B(I\!R^s_+)$*, i.e. as the image of* $I\!R^s_+$ *under a linear transformation* $B \in L(I\!R^s, I\!R^s)$ *induced by a basis submatrix* B *of* W*.*

In our notation we now have the following representation for the gradient of $\tilde{Q}$ (cf. [46], [131]):

$$\tilde{Q}'(\chi) = \sum_{i=1}^{\tilde{\ell}} (-\tilde{d}_i)\mu(\chi + \mathcal{K}_i),$$

where $\chi + \mathcal{K}_i$ stands for the Minkowski sum $\{\chi\} + \mathcal{K}_i$. Introducing the notations

$$d_i := -\tilde{d}_i,\ f_i(\chi) := \mu(\chi + \mathcal{K}_i) \ \text{ and } \ I := \{1, \ldots, \tilde{\ell}\}$$

we obtain

$$\tilde{Q}'(\chi) = \sum_{i \in I} d_i f_i(\chi). \tag{2.4}$$

Furthermore, supp μ denotes the smallest closed set $C \subset I\!R^s$ with $\mu(C) = 1$.
The following theorem is the main result of the present section.

Theorem 2.2.2

(i) Assume (A1), (A3), (A4) and

*(A2)** *there exists a vector* $\bar{u} \in I\!R^s$ *such that* $W^T\bar{u} < q$ *componentwise.*

Then $\tilde{Q}$ *is strictly convex on any open convex subset* $V \subset I\!R^s$ *of* supp μ*.*

(ii) Assume (A1), (A2), (A3) and*

*(A4)** *there exist a convex open set $V \subset \mathbb{R}^s$, constants $r > 0$, $\varrho > 0$ and a density Θ_μ of μ such that*

$$\Theta_\mu(t') \geq r \quad \textit{for all} \quad t' \in R^s \quad \textit{with} \quad \mathrm{dist}(t', V) \leq \varrho.$$

Then $\tilde{Q}$ is strongly convex on V.

Proof: The proofs for the parts (i) and (ii) differ only in their final steps. The subsequent joint considerations are based on (A1), (A2)*, (A3) and (A4).
Let $V \subset \mathbb{R}^s$ be some convex open subset of supp μ, let $\chi \in V$ and $v \in \mathbb{R}^s$ such that $\chi + v \in V$. The proof will finally be given by monotonicity arguments, i.e. by (uniform) lower estimates for

$$(\tilde{Q}'(\chi + v) - \tilde{Q}'(\chi))^T v.$$

In view of (2.4) it holds

$$(2.5) \quad (\tilde{Q}'(\chi + v) - \tilde{Q}'(\chi))^T v = \sum_{i \in I} f_i(\chi + v) d_i^T v - \sum_{i \in I} f_i(\chi) d_i^T v.$$

By (A4) and Proposition 2.2.1 (iii),

$$\sum_{i \in I} f_i(\chi + v) = \sum_{i \in I} f_i(\chi) = 1$$

and, of course,

$$f_i(\chi + v) \geq 0, \quad f_i(\chi) \geq 0 \qquad \text{for all} \quad i \in I.$$

Therefore we have two well-defined probability distributions on $\mathbb{R}$ with mass points $d_i^T v$ $(i \in I)$ and masses $f_i(\chi + v)$ $(i \in I)$ and $f_i(\chi)$ $(i \in I)$, respectively. Denoting the corresponding distribution functions by F_v and $F_{o,v}$, respectively, the identity (2.5) is continued using Riemann-Stieltjes integrals and integration by parts:

$$(2.6) \quad \begin{aligned}(\tilde{Q}'(\chi + v) - \tilde{Q}'(\chi))^T v &= \int_{\mathbb{R}} \tau dF_v(\tau) - \int_{\mathbb{R}} \tau dF_{o,v}(\tau) \\ &= \int_{\mathbb{R}} (F_{o,v}(\tau) - F_v(\tau)) d\tau.\end{aligned}$$

Now let us confirm that

$$(2.7) \quad F_{o,v}(\tau) - F_v(\tau) \geq 0 \qquad \text{for all} \quad \tau \in \mathbb{R}.$$

Denote $I_v(\tau) = \{i \in I : d_i^T v \leq \tau\}$. By (A4) it holds

$$\begin{aligned} F_{o,v}(\tau) &= \mu\Big(\bigcup_{i \in I_v(\tau)} \{\chi + \mathcal{K}_i\}\Big) \qquad \text{and} \\ F_v(\tau) &= \mu\Big(\bigcup_{i \in I_v(\tau)} \{\chi + v + \mathcal{K}_i\}) \end{aligned}$$

We establish (2.7) by showing that for arbitrary $\tau \in I\!R$

$$\bigcup_{i \in I_v(\tau)} \{\chi + v + \mathcal{K}_i\} \subset \bigcup_{i \in I_v(\tau)} \{\chi + \mathcal{K}_i\}. \tag{2.8}$$

Let $I_v(\tau) \neq \emptyset$ and $I_v(\tau) \neq I$, otherwise (2.8) is trivial. Assume there were $i_o \in I_v(\tau)$ and

$$w_{i_o} \in \{\chi + v + \mathcal{K}_{i_o}\} \cap \{I\!R^s \setminus \bigcup_{i \in I_v(\tau)} \{\chi + \mathcal{K}_i\}\}.$$

Then there would exist $i_1 \in I \setminus I_v(\tau)$ such that

$$w_{i_o} \in \chi + \mathcal{K}_{i_1}$$

(recall that $\bigcup\limits_{i \in I} \mathcal{K}_i = I\!R^s$).
Since $i_1 \notin I_v(\tau)$ and $i_o \in I_v(\tau)$, we have

$$d_{i_o}^T v \leq \tau < d_{i_1}^T v. \tag{2.9}$$

Denote $\mathcal{D} := \operatorname{conv}\{d_i : i \in I\}$ and recall that $d_i = -\tilde{d}_i$ $(i \in I)$. The definition of $\mathcal{K}_i$ (cf. Proposition 2.2.1 (ii)) yields

$$d_i \in \arg\min\{t^T d : d \in \mathcal{D}\} \quad \text{for all} \quad t \in \mathcal{K}_i \quad \text{and all} \quad i \in I. \tag{2.10}$$

In view of $w_{i_o} - \chi - v \in \mathcal{K}_{i_o}$ we have by (2.10)

$$d_{i_o} \in \arg\min\{(w_{i_o} - \chi - v)^T d : d \in \mathcal{D}\}. \tag{2.11}$$

Furthermore, $w_{i_o} - \chi \in \mathcal{K}_{i_1}$ and (2.10) yield

$$d_{i_1} \in \arg\min\{(w_{i_o} - \chi)^T d : d \in \mathcal{D}\}. \tag{2.12}$$

In particular, (2.11) implies

$$(w_{i_o} - \chi - v)^T d_{i_o} \leq (w_{i_o} - \chi - v)^T d_{i_1}$$

and therefore

$$(w_{i_o} - \chi)^T d_{i_o} \leq (w_{i_o} - \chi)^T d_{i_1} - v^T d_{i_1} + v^T d_{i_o} \underset{(2.9)}{<} (w_{i_o} - \chi)^T d_{i_1},$$

which contradicts (2.12), and (2.7) is shown.
In the next step we show that

$$(2.13)\quad \alpha^* := \min_{j\in I} \inf_{\substack{\tilde{v}\in\mathcal{K}_j \\ \|\tilde{v}\|=1}} \max_{\substack{i\in I \\ d_i,d_j \text{ adjacent}}} (d_i - d_j)^T \tilde{v} > 0.$$

It is easy to see that (2.10) implies $\alpha^* \geq 0$. So let us assume that $\alpha^* = 0$. Then, for some $j \in I$,

$$\inf_{\substack{\tilde{v}\in\mathcal{K}_j \\ \|\tilde{v}\|=1}} \max_{\substack{i\in I \\ d_i,d_j \text{ adjacent}}} (d_i - d_j)^T \tilde{v} = 0.$$

Hence, for any $n \in I\!N \setminus \{0\}$ there exists a $v_n \in \mathcal{K}_j$, $\|v_n\| = 1$ such that

$$(d_i - d_j)^T v_n \leq \frac{1}{n}$$

for all $i \in I$ such that d_i and d_j are adjacent.
By compactness, the sequence $\{v_n\}$ has an accumulation point $\bar{v} \in \mathcal{K}_j$, $\|\bar{v}\| = 1$. Passing to the limit in the above inequality yields

$$(d_i - d_j)^T \bar{v} = 0$$

for all $i \in I$ such that d_i and d_j are adjacent. Since $\bar{v} \neq 0$, the latter implies $\operatorname{int} \mathcal{D} = \emptyset$, which is impossible due to (A2)*. (Recall that $-d_i$ $(i \in I)$ are the vertices of the feasible set in (2.3).)
For the remainder let us fix some constant $\alpha > 0$ such that $\alpha < \alpha^*$. For $v \in I\!R^s$ introduced at the beginning of the proof there exists some $j = j(v) \in I$ such that $v \in \mathcal{K}_j$. By (2.10) we obtain $d_j^T v \leq d_i^T v$ for all $i \in I$. Using (2.7) we may estimate below the final expression in (2.6) and obtain

$$(2.14)\quad (\tilde{Q}'(\chi + v) - \tilde{Q}'(\chi))^T v \geq \int_{d_j^T v}^{d_j^T v + \alpha\|v\|} (F_{o,v}(\tau) - F_v(\tau)) d\tau.$$

By (2.13), there exists an $i_* = i_*(v) \in I$ such that d_{i_*} and d_j are adjacent and

$$d_j^T v + \alpha\|v\| < d_{i_*}^T v.$$

Consider $\mathcal{F}_{i_*j} := \mathcal{K}_{i_*} \cap \mathcal{K}_j$. Since d_{i_*} and d_j are adjacent, Proposition 2.1 (iv) implies that $\mathcal{F}_{i_*j}$ is a joint facet (closed $(s-1)$-dimensional face) of $\mathcal{K}_{i_*}$ and $\mathcal{K}_j$.
Then, for all $\tau \in I\!R$ with $d_j^T v \leq \tau \leq d_j^T v + \alpha\|v\|$ the following inclusion is valid

$$(2.15)\quad \bigcup_{0\leq\lambda<1} \{\chi + \lambda v + \mathcal{F}_{i_*j}\} \subset \{\chi + \bigcup_{i\in I_v(\tau)} \mathcal{K}_i\} \setminus \{\chi + v + \bigcup_{i\in I_v(\tau)} \mathcal{K}_i\}.$$

Let us first verify

$$(2.16)\quad \bigcup_{0\le\lambda<1}\{\chi+\lambda v+\mathcal{F}_{i_*j}\}\subset\{\chi+\bigcup_{i\in I_v(\tau)}\mathcal{K}_i\}.$$

Of course, $j\in I_v(\tau)$ for each of the above $\tau\in I\!\!R$. Consider arbitrary $\lambda\in[0,1)$ and $v_{i_*j}\in\mathcal{F}_{i_*j}\subset\mathcal{K}_j$. Since $\mathcal{K}_j$ is a convex cone and $v\in\mathcal{K}_j$, this implies $\lambda v+v_{i_*j}\in\mathcal{K}_j$, verifying (2.16).
Now we show that

$$\bigcup_{0\le\lambda<1}\{\chi+\lambda v+\mathcal{F}_{i_*j}\}\cap\{\chi+v+\bigcup_{i\in I_v(\tau)}\mathcal{K}_i\}=\emptyset\quad\text{for all }\tau\text{ in question.}$$

Assume on the contrary that for some τ in question there were $\lambda\in[0,1)$, $v_{i_*j}\in\mathcal{F}_{i_*j}$, $u_i\in\mathcal{K}_i$ $(i\in I_v(\tau))$ such that

$$\chi+\lambda v+v_{i_*j}=\chi+v+u_i.$$

This would yield

$$u_i=-(1-\lambda)v+v_{i_*j}\in\mathcal{K}_i$$

and, by (2.10),

$$(v_{i_*j}-(1-\lambda)v)^Td_i\le(v_{i_*j}-(1-\lambda)v)^Td\quad\text{for all}\quad d\in\mathcal{D}.$$

In particular

$$(v_{i_*j}-(1-\lambda)v)^Td_i\le(v_{i_*j}-(1-\lambda)v)^Td_{i_*},$$

implying

$$v_{i_*j}^Td_i-v_{i_*j}^Td_{i_*}\le(1-\lambda)v^T(d_i-d_{i_*}).$$

By the selection of i_* and by $i\in I_v(\tau)$ we have

$$v^Td_{i_*}>v^Td_j+\alpha\|v\|\ge\tau\ge v^Td_i.$$

Hence

$$v_{i_*j}^Td_i-v_{i_*j}^Td_{i_*}<0$$

in contradiction to $v_{i_*j} \in \mathcal{F}_{i_*j} \subset \mathcal{K}_{i_*}$ and (2.10). This verifies (2.15).
Let $\tau \in I\!R$ such that $d_j^T v \leq \tau \leq d_j^T v + \alpha\|v\|$. Then, the following holds due to (A4), (2.8) and (2.15):

$$\begin{aligned} F_{o,v}(\tau) - F_v(\tau) &= \mu\Big(\bigcup_{i \in I_v(\tau)} \{\chi + \mathcal{K}_i\}\Big) - \mu\Big(\bigcup_{i \in I_v(\tau)} \{\chi + v + \mathcal{K}_i\}\Big) \\ &= \mu\Big(\Big\{\chi + \bigcup_{i \in I_v(\tau)} \mathcal{K}_i\Big\} \setminus \Big\{\chi + v + \bigcup_{i \in I_v(\tau)} \mathcal{K}_i\Big\}\Big) \\ &\geq \mu\Big(\bigcup_{0 \leq \lambda \leq 1} \{\chi + \lambda v + \mathcal{F}_{i_*j}\}\Big). \end{aligned}$$

The last expression does not depend on τ. Hence, (2.14) implies

$$(2.17) \quad (\tilde{Q}'(\chi + v) - \tilde{Q}'(\chi))^T v \geq \alpha \cdot \|v\| \cdot \mu\Big(\bigcup_{0 \leq \lambda \leq 1} \{\chi + \lambda v + \mathcal{F}_{i_*j}\}\Big).$$

The set $\bigcup_{0 \leq \lambda \leq 1} \{\chi + \lambda v + \mathcal{F}_{i_*j}\}$ is cylindric and located between the two parallel affine hyperplanes

$$\chi + \text{span}\, \mathcal{F}_{i_*j} \qquad \text{and} \qquad \chi + v + \text{span}\, \mathcal{F}_{i_*j}.$$

We start the (lower) estimation of $\mu(\bigcup_{0 \leq \lambda \leq 1} \{\chi + \lambda v + \mathcal{F}_{i_*j}\})$ by deriving a lower bound for the Hausdorff distance of the above hyperplanes.
Recall that $\mathcal{F}_{i_*j} = \mathcal{K}_{i_*} \cap \mathcal{K}_j$ and therefore ((2.10))

$$v_{i_*j}^T d_{i_*} = v_{i_*j}^T d_j = \min\{v_{i_*j}^T d : d \in \mathcal{D}\} \quad \text{for all } v_{i_*j} \in \mathcal{F}_{i_*j}.$$

Since $\mathcal{F}_{i_*j}$ is $(s-1)$-dimensional, the orthogonal complement $(\text{span}\, \mathcal{F}_{i_*j})^\perp$ is thus spanned by the vector $d_{i_*} - d_j$.
Hence, the Hausdorff distance of $\chi + \text{span}\, \mathcal{F}_{i_*j}$ and $\chi + v + \text{span}\, \mathcal{F}_{i_*j}$ equals

$$(2.18) \quad \frac{(d_{i_*} - d_j)^T v}{\|d_{i_*} - d_j\|}.$$

Introducing the positive constant

$$\delta := \max\{\|d_{i_1} - d_{i_2}\| : i_1, i_2 \in I,\ d_{i_1}, d_{i_2} \text{ adjacent}\}$$

we obtain

$$(2.19) \quad \frac{(d_{i_*} - d_j)^T v}{\|d_{i_*} - d_j\|} \geq \frac{\alpha}{\delta}\|v\|.$$

Let us now verify the assertions (i) and (ii) of our theorem. Since V is an open subset of supp μ and $\chi \in V$, it holds

$$\mu\Big(\bigcup_{0\le\lambda\le1} \{\chi + \lambda v + \mathcal{F}_{i_*(v)j(v)}\}\Big) > 0 \quad \text{for all} \quad v \in I\!R^s \setminus \{0\}.$$

Together with (2.17) this yields

$$(\tilde{Q}'(\chi+v) - \tilde{Q}'(\chi))^T v > 0$$

for all $\chi \in V$ and all $v \in I\!R^s \setminus \{0\}$ such that $\chi + v \in V$.
Hence, the gradient of $\tilde{Q}$ is strictly monotone and, therefore, $\tilde{Q}$ is strictly convex on V.
Now impose (A4)* and denote $\mathcal{F}^{\varrho}_{i_*j} := \{\tilde{v} \in \mathcal{F}_{i_*j} : \|\tilde{v}\| \le \varrho\}$. Then it holds

$$(2.20)\quad \mu\Big(\bigcup_{0\le\lambda\le1} \{\chi + \lambda v + \mathcal{F}_{i_*j}\}\Big) \;\ge\; \mu\Big(\bigcup_{0\le\lambda\le1} \{\chi + \lambda v + \mathcal{F}^{\varrho}_{i_*j}\}\Big)$$
$$\ge\; r \cdot \ell_s\Big(\bigcup_{0\le\lambda\le1} \{\lambda v + \mathcal{F}^{\varrho}_{i*j}\}\Big)$$

where ℓ_s denotes the s-dimensional Lebesgue measure.
By (2.18) amd (2.19) one obtains

$$(2.21)\quad \ell_s\Big(\bigcup_{0\le\lambda\le1} \{\lambda v + \mathcal{F}^{\varrho}_{i_*j}\}\Big) \ge \frac{\alpha}{\delta}\|v\| \cdot \ell_{s-1}(\mathcal{F}^{\varrho}_{i_*j}\Big).$$

We introduce the constant

$$\ell_{\min} = \min\{\ell_{s-1}(\mathcal{F}^{\varrho}_{i_1 i_2}) : i_1, i_2 \in I,\ \mathcal{F}_{i_1,i_2} = \mathcal{K}_{i_1} \cap \mathcal{K}_{i_2} \ \text{ is a facet}\}$$

Altogether, only finitely many facets $\mathcal{F}_{i_1 i_2}$ may occur. Therefore, $\ell_{\min} > 0$. Using (2.17), (2.20) and (2.21) we finally obtain

$$(\tilde{Q}'(\chi+v) - \tilde{Q}'(\chi))^T v \ge \alpha \cdot \|v\| \cdot r \cdot \frac{\alpha}{\delta}\|v\| \cdot \ell_{\min} = \alpha^2 \ell_{\min} \delta^{-1} r \|v\|^2.$$

By construction, the (positive) constants α, δ and $\ell_{\min}$ do neither depend on χ nor on v. Hence $\tilde{Q}$ is strongly convex on V with the constant $\tilde{\kappa} = \frac{1}{2}\alpha^2 \ell_{\min} \delta^{-1} r$. ■

In what follows, we discuss the assumptions of the above theorem, present special cases and comment on relations to elder results.

Proposition 2.2.3 *Assume (A1), (A2), (A3), (A4). Then (A2)* is necessary for the strict convexity of* $\tilde{Q}$.

Proof: Assume that $\tilde{Q}$ is strictly convex and that (A2)* does not hold. Then there exist $\bar{v}_o \in I\!R^s$, $v_o \in I\!R^s \setminus \{0\}$ such that

$$\mathrm{conv}\{d_i : i \in I\} \subset \bar{v}_o + (\mathrm{span}\{v_o\})^{\perp}.$$

With arbitrary $\chi \in I\!R^s$ we have

$$(\tilde{Q}'(\chi + v_o) - \tilde{Q}'(\chi))^T v_o = \Big(\sum_{i \in I} f_i(\chi + v_o) d_i - \sum_{i \in I} f_i(\chi) d_i \Big)^T v_o = 0$$

in contradiction to the strict monotonicity of $\tilde{Q}'$. ■

Remark 2.2.4 *If μ can be represented as a convex combination $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$ (with $0 < \lambda < 1$), then Theorem 2.2.2 already works if (A4) (or (A4)*) holds for one of the measures μ_1, μ_2, only. Indeed, $\tilde{Q}$ is then strictly (strongly) convex as the sum of a convex and a strictly (strongly) convex function.*

Remark 2.2.5 *If (A4)* fails, i.e. if there does not exist a positive uniform lower bound for the density Θ_μ, then the estimate (2.17) in the proof of Theorem 2.2.2 can be used to check whether some modified lower bound*

$$(\tilde{Q}'(\chi + v) - \tilde{Q}'(\chi))^T v \geq \kappa \|v\| \cdot \phi(\|v\|)$$

is available, where $\phi : I\!R_+ \to I\!R_+$ is some strictly increasing function with $\phi(0) = 0$. In this way one obtains another type of monotonicity of $\tilde{Q}'$ resulting in another type of uniform convexity for $\tilde{Q}$. More specifically, the convexity module $\|\chi_1 - \chi_2\|^2$ in the definition of strong convexity is replaced by $\tilde{\phi}(\|\chi_1 - \chi_2\|)$, where $\tilde{\phi} : I\!R_+ \to I\!R_+$ is a strictly increasing function with $\tilde{\phi}(0) = 0$. For a detailed discussion of this type of uniform convexity see [123].

Remark 2.2.6 *The proof of Theorem 2.2.2 also gives some indication how to proceed when wishing to estimate the strong-convexity constant $\tilde{\kappa}$. Then, of course, much depends on how explicit the polyhedron M_D (cf. (2.3)), its vertices and normal cones are available. In Section 2.3 it will be important that (for fixed $\varrho > 0$) the constant $\tilde{\kappa}$ is of the form $\tilde{\kappa} = \tilde{\kappa}_o(q, W) \cdot \tilde{\kappa}_1(\mu)$ where $\tilde{\kappa}_o = \frac{1}{2}\alpha^2\delta^{-1}\ell_{\min}$ depends only on q, W and $\tilde{\kappa}_1 = r$ depends only on μ.*

Remark 2.2.7 *If (A2) is fulfilled but not (A2)*, the latter may be guaranteed by a proper increase of the second stage costs q_k $(k = 1, \ldots, \bar{m})$.*

To illustrate the impact of assumption (A2)* we consider the function $\tilde{Q}$ for

$$W = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 1 & 1 & 1 & -1 \end{pmatrix}$$

and different instances of the cost vector $q \in I\!R^4$. It is easy to see that W fulfils (A1), and we will always assume that $\mu \in \mathcal{P}(I\!R^2)$ is selected in such a way that (A3) and (A4) are met.

Let $q = (-1,-1,-1,1)^T$. Then the feasible region M_D in (2.3) degenerates to a singleton. Of course, (A2)* is violated and, moreover, there is only one vertex of M_D. Hence, the gradient of $\tilde{Q}$ is constant, and $\tilde{Q}$ is linear. If we increase the costs to $q = (1,-1,1,1)^T$, then M_D turns into the convex hull of the points $(-2,-1)^T$ and $(2,-1)^T$. Therefore, the second component of $\tilde{Q}'$ is always equal to 1, implying that $\tilde{Q}'$ cannot be strictly monotone. For $q = (1,0,1,1)^T$ we have $M_D = \operatorname{conv}\{(-1,0)^T, (1,0)^T, (2,-1)^T, (-2,-1)^T\}$ which has a non-empty interior, and Theorem 2.2(i) works.

In the following proposition we present necessary and sufficient conditions for (A2)* for two specific classes of second-stage problems.

Proposition 2.2.8

(i) (Extended simple recourse) Suppose that $W \in L(I\!R^{2s}, I\!R^s)$, $W = (H, -H)$ with some non-singular matrix $H = L(I\!R^s, I\!R^s)$. Split q into q^+, q^- such that $q^T = (q^{+T}, q^{-T})$, $q^+, q^- \in I\!R^s$. Then (A1) is always fulfilled and (A2) holds if and only if $q^+ + q^- > 0$ componentwise.*

(ii) Suppose that $W \in L(I\!R^{s+1}, I\!R^s)$ fulfils (A1) and that (A2) holds. Then (A2) is fulfilled if and only if $q \notin \operatorname{im} W^T$.*

Proof:

(i) It is well-known that our assumptions imply (A1). Furthermore, it holds

$$\begin{array}{lll} & W^T u < q & \text{for some} \quad u \in I\!R^s \\ \text{iff} & H^T u < q^+ \text{ and } -H^T u < q^- & \text{for some} \quad u \in I\!R^s \\ \text{iff} & -q^- < H^T u < q^+ & \text{for some} \quad u \in I\!R^s \\ \text{iff} & 0 < q^+ + q^- & \text{(since } H \text{ is non-singular).} \end{array}$$

(ii) **only if:** Suppose there were $\tilde{u}, \bar{u} \in I\!R^s$ such that $W^T\tilde{u} = q$ and $W^T\bar{u} < q$. Consider $u(\tau) = \tilde{u} + \tau(\bar{u} - \tilde{u})$ for arbitrary $\tau > 0$. It holds

$$W^T u(\tau) = W^T\tilde{u} + \tau W^T(\bar{u} - \tilde{u}) < q \quad \text{for all} \quad \tau > 0.$$

Hence, the set $\{u \in I\!R^s : W^T u \leq q\}$ is unbounded in contradiction to (A1), (A2) and the duality theorem of linear programming.
if: (A1) and $W \in L(I\!R^{s+1}, I\!R^s)$ imply that each collection of s different columns of W forms a nonsingular matrix (see e.g. [46]). Furthermore, the matrices $B_1, \ldots, B_{s+1}$ formed in this way are just the basis matrices mentioned in Proposition 2.2.1 (v). Moreover, each of the cones $\mathcal{K}_i$ here coincides with $B_i(I\!R^s_+)$ for some $i \in \{1, \ldots, s+1\}$.
For any basis matrix B_i ($i \in \{1, \ldots, s+1\}$) we denote by N_i the non-basic part of W (here, of course, N_i always consists of one column) and we denote by q_{B_i}, q_{N_i} the sub-vectors of q formed by the components corresponding to the columns in B_i and N_i, respectively.
Then (cf. e.g. [46]) $\tilde{d}_i = (B_i^{-1})^T q_{B_i}$ and $q_{N_i} - N_i^T (B_i^{-1})^T q_{B_i} \geq 0$ for all $i \in \{1, \ldots, s+1\}$.
Consider $W^T \tilde{d}_i$. The components of $W^T \tilde{d}_i$ belonging to columns in B_i obviously coincide with the corresponding components in q_{B_i}. For the non-basic component we must have

$$N_i^T \tilde{d}_i = N_i^T (B_i^{-1})^T q_{B_i} < q_{N_i}$$

since otherwise $W^T \tilde{d}_i = q$ in contradiction to $q \notin \operatorname{im} W^T$.
Now it is easy to see that $\bar{u} := \frac{1}{s+1} \sum_{i=1}^{s+1} \tilde{d}_i$ fulfils $W^T \bar{u} < q$ componentwise.

■

Remark 2.2.9 *Using completely different techniques, another sufficient condition for the strong convexity of $\tilde{Q}$ was derived in Theorem 3.1 in [95]. Compared to Theorem 2.2.2 the analysis in [95] needs the additional assumption that $\tilde{Q}$ has a locally Lipschitzian gradient. Moreover, the verification of the sufficient condition in Theorem 3.1 in [95] is more technical since the kernel of the matrix whose columns are d_i ($i \in I$) and certain generalized directional derivatives of the functions f_i ($i \in I$) have to be studied.*

Altogether, Theorem 2.2.2 provides quite handy tools to verify strict and strong convexity of $\tilde{Q}$, respectively. Furthermore, it yields information about the structure of the modulus of strong convexity, and it shows how to perturb recourse models without additional convexity properties to arrive at strictly or strongly convex functions $\tilde{Q}$.

2.3 Strong Convexity - Stochastic Technology Matrix

Let us first recall some concepts from probability theory and introduce a few notations.

Let $\pi_{\mathbb{R}^s}$ and $\pi_{\mathbb{R}^{ms}}$ denote the projections from $\mathbb{R}^s \times \mathbb{R}^{ms}$ to $\mathbb{R}^s$ and $\mathbb{R}^{ms}$, respectively. The induced measures $\mu_1 := \mu \circ \pi_{\mathbb{R}^s}^{-1}$, $\mu_2 := \mu \circ \pi_{\mathbb{R}^{ms}}^{-1}$ are then called the marginal distributions of μ with respect to z and A, respectively. By $\mu_1^2(A,.)$ we denote the regular conditional distribution of z given A. It has the following properties

(3.1) $\mu_1^2(A,.)$ is a probability measure on $\mathbb{R}^s$ for any $A \in \mathbb{R}^{ms}$;

(3.2) the function $\mu_1^2(.,B_1) : \mathbb{R}^{ms} \to [0,1]$ is measurable for any Borel set B_1 in $\mathbb{R}^s$;

(3.3) for any Borel set B in $\mathbb{R}^{(m+1)s}$ it holds

$$\mu(B) = \int\limits_{\mathbb{R}^{ms}} \int\limits_{\mathbb{R}^s} \mathbb{1}_B(z,A)\mu_1^2(A,dz)\mu_2(dA),$$

where $\mathbb{1}_B$ denotes the indicator function of B.

Since μ acts on a complete, separable metric space, the regular conditional distribution $\mu_1^2(A,.)$ exists, indeed (cf. [24], [38]).

Let us now return to our model (1.1) - (1.3). We will write the images $A(\omega)$ of A in the form $A(\omega) = (A_o, A_1(\omega))$ where $A_o \in L(\mathbb{R}^k, \mathbb{R}^s)$, $A_1(\omega) \in L(\mathbb{R}^{m-k}, \mathbb{R}^s)$, $0 \le k \le m$, and A_o is formed by those column vectors of $A(\omega)$ whose entries are all constant P-almost surely, i.e. in $A(\omega)$ we separate the random from the non-random part. Of course, μ_2 is then concentrated on the range space of $A(.)$, i.e. on a subspace of dimension (at most) $(m-k)s$.

To sketch the central idea for the subsequent approach we remark that (3.3) implies for all $x \in \mathbb{R}^m$

$$Q(x) = \int\limits_{\mathbb{R}^{ms}} \int\limits_{\mathbb{R}^s} \Phi(z - Ax)\mu_1^2(A,dz)\mu_2(dA).$$

Now assume that the inner integral obeys (with a certain uniformity in A) a strong convexity as in Section 2.2 and then study the impact of the outer integral. To simplify the notation we will use the symbol E for integration with respect to μ_2.

Theorem 2.3.1 *Assume (A1) - (A3), let $V \subset I\!R^m$ be non-empty, convex and suppose*

(3.4) *for μ_2-almost all $A \in I\!R^{ms}$ the function*
$$\tilde{Q}_A(\chi) := \int_{I\!R^s} \Phi(z-\chi)\mu_1^2(A, dz)$$
is strongly convex on $\tilde{V} := A(V)$ with some modulus $\kappa(A)$, and there exists some $\kappa > 0$ such that $\kappa(A) \geq \kappa$ for μ_2-almost all A;

(3.5) $E(\|A\|^2) < +\infty$*;*

(3.6) *$k \leq s$, i.e. in $A_1(\omega)$ there are at least $m-s$ columns;*

(3.7) *A_o has full rank;*

(3.8) *the matrix $E(A_1^T A_1) - E(A_1)^T E(A_1)$ is positive-definite.*

Then Q is strongly convex on V.

Proof: Let $x_1, x_2 \in V$ and $\lambda \in [0,1]$. We have

$$
\begin{aligned}
(3.9)\quad & Q(\lambda x_1 + (1-\lambda)x_2) \\
&= \int_{I\!R^{(m+1)s}} \Phi(z - A(\lambda x_1 + (1-\lambda)x_2))\mu(d(z,A)) \\
&\underset{(3.3)}{=} \int_{I\!R^{ms}} \int_{I\!R^s} \Phi(z - \lambda A x_1 - (1-\lambda)Ax_2)\mu_1^2(A,dz)\mu_2(dA) \\
&= \int_{I\!R^{ms}} \tilde{Q}_A(\lambda A x_1 + (1-\lambda)Ax_2)\mu_2(dA) \\
&\underset{(3.4)}{\leq} \int_{I\!R^{ms}} (\lambda \tilde{Q}_A(Ax_1) + (1-\lambda)\tilde{Q}_A(Ax_2) - \kappa(A)\lambda(1-\lambda)\|Ax_1 - Ax_2\|^2)\mu_2(dA) \\
&\underset{(3.3)}{=} \lambda Q(x_1) + (1-\lambda)Q(x_2) - \lambda(1-\lambda)\int_{I\!R^{ms}} \kappa(A)\|Ax_1 - Ax_2\|^2\mu_2(dA) \\
&\underset{(3.4)}{\leq} \lambda Q(x_1) + (1-\lambda)Q(x_2) - \kappa\lambda(1-\lambda)\int_{I\!R^{ms}} \|Ax_1 - Ax_2\|^2\mu_2(dA) \\
&= \lambda Q(x_1) + (1-\lambda)Q(x_2) - \kappa\lambda(1-\lambda)(x_1-x_2)^T E(A^T A)(x_1 - x_2).
\end{aligned}
$$

It remains to show that the positive-semidefinite matrix $E(A^T A)$ is even positive-definite. Recalling that $A = (A_o, A_1)$, where we have omitted the ω in A_1 for convenience, we obtain

$$(3.10)\quad E(A^T A) = E\begin{pmatrix} A_o^T A_o & A_o^T A_1 \\ A_1^T A_o & A_1^T A_1 \end{pmatrix} = \begin{pmatrix} A_o^T A_o & A_o^T E(A_1) \\ E(A_1)^T A_o & E(A_1^T A_1) \end{pmatrix}.$$

Due to the assumptions (3.6) and (3.7), the matrix $A_o^T A_o$ is invertible and we can multiply (3.10) from the left by

$$\begin{pmatrix} I_k & 0 \\ -E(A_1)^T A_o (A_o^T A_o)^{-1} & I_{m-k} \end{pmatrix}$$

where I_k, I_{m-k} denote the identities of dimensions k and $m-k$, respectively. The multiplication does not change the determinant of (3.10), and yields

$$\begin{pmatrix} A_o^T A_o & A_o^T E(A_1) \\ 0 & E(A_1^T A_1) - E(A_1)^T A_o (A_o^T A_o)^{-1} A_o^T E(A_1) \end{pmatrix}.$$

Of course, the above matrix is non-singular if and only if the matrix

$$(3.11) E(A_1^T A_1) - E(A_1)^T A_o (A_o^T A_o)^{-1} A_o^T E(A_1),$$

usually called the Schur complement in the literature, is non-singular. Obviously, (3.11) coincides with

$$E(A_1^T A_1) - E(A_1)^T E(A_1) + E(A_1)^T (I_s - A_o (A_o^T A_o)^{-1} A_o^T) E(A_1).$$

By assumption (3.8) we are done when having verified that

$$(3.12) E(A_1)^T (I_s - A_o (A_o^T A_o)^{-1} A_o^T) E(A_1)$$

is positive-semidefinite.
To this end, observe that

$$I_s - A_o (A_o^T A_o)^{-1} A_o^T$$

is the matrix of the orthogonal projection from $I\!R^s$ onto the linear subspace

$$\{u \in I\!R^s : \ A_o^T u = 0\}$$

and, hence, is positive-semidefinite. Therefore, also (3.12) is positive-semidefinite.

■

At the beginning of Section 2.2 we have mentioned the missing strong convexity of Q for a non-stochastic matrix A. The above theorem shows how randomness in A improves the convexity properties of Q.

Corollary 2.3.2 *Assume (A1) - (A3), let $V \subset I\!R^m$ be non-empty, convex and suppose that (3.4), (3.5) hold. Then, for all $x_1, x_2 \in V$ and all $\lambda \in [0,1]$, we have*

$$Q(\lambda x_1 + (1-\lambda) x_2) \le \lambda Q(x_1) + (1-\lambda) Q(x_2) - \kappa \lambda (1-\lambda) \|E(A) x_1 - E(A) x_2\|^2.$$

Proof: The proof immediately follows from (3.9) and Jensen's inequality. ■

Corollary 2.3.2 says that, given (3.4) and the mild integrability assumption (3.5), one arrives at a strong convexity property which is quite comparable to that for non-stochastic A with $E(A)$ playing the role of A. "Investing a bit more", namely (3.6) - (3.8), we obtain the "real" strong convexity for Q (Theorem 2.3.1). Together with Theorem 2.2.2, Theorem 2.3.1 provides the essential tools for verifying the strong convexity of Q in (1.2) right from the model (1.1) - (1.3). Of course, the assumptions in Theorem 2.3.1 need some discussion, which is carried out below. In particular, assumption (3.4) seems hard to be verified. However, there exist specific situations where one is able to say a bit more about $\mu_1^2(A,.)$ and in such cases also (3.4) can be checked quite easily.

Remark 2.3.3 *If the random vectors z and A are stochastically independent, then $\mu_1^2(A,.) = \mu_1$ holds for μ_2-almost all $A \in I\!R^{ms}$. Hence, (3.4) is satisfied if the function $\tilde{Q}(\chi) := \int\limits_{I\!R^s} \Phi(z-\chi)\mu_1(dz)$ is strongly convex on $\bigcup\limits_{\omega\in\Omega} A(\omega)(V)$.*

Remark 2.3.4 *Assume (A1), (A2)*, (A3) and suppose that there exist a convex set $V \in I\!R^m$, constants $r > 0$, $\varrho > 0$ and density functions $\Theta_{1,A}^2$ of $\mu_1^2(A,.)$ such that $\Theta_{1,A}^2(t') \geq r$ for all $t' \in I\!R^s$ with $\mathrm{dist}(t', \bigcup\limits_{\omega\in\Omega} A(\omega)(V)) \leq \varrho$ and μ_2-almost all $A \in I\!R^{ms}$. Then Theorem 2.2.2 (ii) and Remark 2.6 imply the validity of assumption (3.4). Of course, the above implies necessarily that $\bigcup\limits_{\omega\in\Omega} A(\omega)(V)$ is bounded. Hence, it is reasonable to assume that V is bounded and that μ_2 has bounded support.*

Remark 2.3.5 *The regular conditional distribution $\mu_1^2(A;.)$ has a density for μ_2-almost all $A \in I\!R^{ms}$ if <u>either</u> z and A are independent and μ_1 has a density <u>or</u> if there exists a joint density of z and the random components in A, <u>or</u> if z and A are dependent, μ_2 is discrete (with countably many mass points) and μ_1 has a density. In the second case, the density for $\mu_1^2(A,.)$ computes as the quotient of the joint density and the marginal density for z.*

Remark 2.3.6 *Assumption (3.8) is fulfilled if in any row of $A_1(\omega)$ the random entries are pairwise uncorrelated (i.e. their covariance is zero). Indeed, one then easily computes that the matrix $E(A_1^T A_1) - E(A_1)^T E(A_1)$ is diagonal with positive entries along the main diagonal.*

2.4 Applications to Stability

Let us now present a few consequences of the preceding results for the stability of stochastic programs. We will see how the Theorems 2.2.2 and 2.3.1 provide sufficient stability conditions.

Consider

$$P(\mu) \qquad \min\{g(x) + Q(x,\mu) : x \in C\}$$

in the same way as in (1.1) - (1.3), but understanding the underlying probability measure μ as varying in a suitable space of parameters. Stability of $P(\mu)$ is now studied in terms of the (extend real-valued) function φ assigning to μ the optimal value of $P(\mu)$ and, in terms of the set-function ψ, assigning to μ the set of optimal solutions to $P(\mu)$. Due to convexity, the optimality is always a global one.

The stability analysis of stochastic programs with respect to perturbations of the underlying probability measure allows a unified approach to questions arising from approximating complicated measures by simpler ones or to problems in connection with incomplete information about the measure. For details see e.g. [27], [48], [88] - [116], [124].

We specify the parameter space for $P(\mu)$ as

$$\mathfrak{M}_1(I\!R^S) = \{\mu' \in \mathcal{P}(I\!R^S) : \int\limits_{I\!R^S} \|\xi\| \mu'(d\xi) < +\infty\}$$

where $\mathcal{P}(I\!R^S)$ denotes the set of all Borel probability measures on $I\!R^S$ $(S := (m+1)s)$. For $\mu, \nu \in \mathfrak{M}_1(I\!R^S)$ then the L_1-Wasserstein distance is defined as follows

$$W_1(\mu,\nu) = \inf\Big\{ \int\limits_{I\!R^S \times I\!R^S} \|\xi - \tilde{\xi}\| \eta(d\xi, d\tilde{\xi}) : \; \eta \in D(\mu,\nu) \Big\}$$

where

$$D(\mu,\nu) = \{\eta \in \mathcal{P}(I\!R^S \times I\!R^S) : \eta \circ \pi_1^{-1} = \mu, \; \eta \circ \pi_2^{-1} = \nu\}.$$

Now $(\mathfrak{M}_1(I\!R^S), W_1)$ is a metric space, and in [80] it is shown that a sequence $\{\mu_n\}$ of probability measures $\mu_n \in \mathcal{P}(I\!R^S)$ converges in W_1 to $\mu \in \mathfrak{M}_1(I\!R^S)$ if and only if $\{\mu_n\}$ converges weakly to μ and $\lim\limits_{n\to\infty} \int\limits_{I\!R^S} \|\xi\| \mu_n(d\xi) = \int\limits_{I\!R^S} \|\xi\| \mu(d\xi)$.

The sequence $\{\mu_n\}$ is said to converge weakly to μ if $\int h(\xi)\mu_n(d\xi) \to \int h(\xi)\mu(d\xi)$ for any bounded continuous function $h : I\!R^S \to I\!R$ ([9]).

In [93], [95] analytical properties of the mappings φ and ψ were derived when equipping $\mathfrak{M}_1(\mathbb{R}^S)$ with weak convergence and the Wasserstein metric, respectively. Theorem 2.4 in [93] and Proposition 2.1 in [95] represent quite comprehensive results on the (Lipschitz) continuity of φ. Much less is known about (mainly quantitative) continuity properties of ψ. It has already been observed in [93], [95] that, for non-stochastic A, the strong convexity of the function $\tilde{Q}$ (cf. (2.1)) is crucial in this respect. In what follows, we extend the quantitative analysis of solution stability to the case of a non-stochastic technology matrix A. Simultaneously, our sufficient conditions for strong convexity provide tools for checking assumptions which are much easier to handle than those in [93], [95].

Theorem 2.4.1 *Consider (1.1) - (1.3), assume (A1) - (A3) and suppose that $\psi(\mu)$ is non-empty and bounded. Let $V \subset \mathbb{R}^m$ be some open, bounded, convex set containing $\psi(\mu)$ and suppose that (3.4), (3.5) hold.*
Then there exist constants $L > 0$, $\delta > 0$ such that $\psi(\nu) \neq \emptyset$ and

$$d_H(E(A)(\psi(\nu)), E(A)(\psi(\mu))) \leq L \cdot W_1(\mu, \nu)^{1/2}$$

whenever $\nu \in \mathfrak{M}_1(\mathbb{R}^S)$, $W_1(\mu, \nu) < \delta$.

Here, $E(A)$ has the same meaning as in Section 2.3 and d_H denotes the usual Hausdorff distance of sets.

Proof: By Theorem 2.4 and Remark 2.5 in [93] there exist constants $L_o > 0$, $\delta_o > 0$ such that $\emptyset \neq \psi(\nu) \subset V$ and

$$|\varphi(\mu) - \varphi(\nu)| \leq L_o W_1(\mu, \nu) \quad \text{for all} \quad \nu \in \mathfrak{M}_1(\mathbb{R}^S), W_1(\mu, \nu) < \delta_o.$$

Denote $G(x, \nu) := g(x) + Q(x, \nu)$ for all $x \in \mathbb{R}^m$, $\nu \in \mathfrak{M}_1(\mathbb{R}^S)$.
Let $\bar{x} \in \psi(\mu)$. The optimality of $\bar{x}$ and Corollary 2.3.2 imply for all $x \in C \cap V$:

$$\begin{aligned} G(\bar{x}, \mu) &\leq G(\frac{1}{2}(x + \bar{x}), \mu) \\ &\leq \frac{1}{2}G(x, \mu) + \frac{1}{2}G(\bar{x}, \mu) - \frac{\kappa}{4}\|E(A)x - E(A)\bar{x}\|^2. \end{aligned}$$

This yields

$$G(x, \mu) \geq G(\bar{x}, \mu) + \frac{\kappa}{2}\|E(A)x - E(A)\bar{x}\|^2. \tag{4.1}$$

By (4.1), the set $E(A)(\psi(\mu))$ is a singleton.
Put $\delta := \delta_o$, consider some $\nu \in \mathfrak{M}_1(I\!R^S)$ with $W_1(\mu,\nu) < \delta$ and some $x \in \psi(\nu)$. It holds

$$\begin{aligned} \|E(A)x - E(A)\bar{x}\|^2 &\leq \frac{2}{\kappa}(G(x,\mu) - G(\bar{x},\mu)) \\ &\leq \frac{2}{\kappa}(|\varphi(\mu) - \varphi(\nu)| + |G(x,\mu) - G(x,\nu)|) \\ &\leq \frac{2}{\kappa}(L_o W_1(\mu,\nu) + |Q(x,\mu) - Q(x,\nu)|). \end{aligned}$$

Recall that $x \in \psi(\nu) \subset V$ and that V is bounded. In view of the discussion in [93] (page 247 and Remarks 2.2, 2.5) there exists some constant $L_1 > 0$ such that

$$|Q(x,\mu) - Q(x,\nu)| \leq L_1 W_1(\mu,\nu) \qquad \text{for all} \quad x \in V.$$

Thus

$$\|E(A)x - E(A)\bar{x}\|^2 \leq \frac{2}{\kappa}(L_o + L_1)W_1(\mu,\nu) \quad \text{for all} \quad x \in \psi(\nu).$$

■

Theorem 2.4.1 contains Theorem 2.2 in [95] (where $A(\omega) \equiv A$) as a special case. Therefore, Example 2.3 in [95] can again be used to show that the rate 1/2 on the right-hand side of the above estimate is best possible. Further, the examples in Remark 2.9 in [93] show that, for general convex g and C, the above estimate does not extend to the Hausdorff distance of the solution sets. For a non-stochastic technology matrix A, the following proposition gives sufficient conditions on g and C such that the estimate extends.

Proposition 2.4.2 *([93], Theorem 2.7)*
Consider (1.1) - (1.3) with non-stochastic A. Assume (A1) - (A3), let $\psi(\mu)$ be non-empty, bounded, g be convex quadratic and $C \subset I\!R^m$ be a polyhedron. Suppose further that the function $\tilde{Q}$ (cf. (2.1)) is strongly convex on a convex open set V containing $A(\psi(\mu))$.
Then there exist constants $L > 0$, $\delta > 0$ such that $\psi(\nu) \neq \emptyset$ and

$$d_H(\psi(\mu), \psi(\nu)) \leq L \cdot W_1(\mu,\nu)^{1/2}$$

whenever $\nu \in \mathfrak{M}_1(I\!R^s)$, $W_1(\mu,\nu) < \delta$.

For recourse models with random technology matrix there is another possibility to obtain a Hölder estimate for the Hausdorff distance of solution sets. One simply has to combine Theorem 2.3.1 with the essential ideas from Theorem 2.4.1.

Theorem 2.4.3 *Consider (1.1) - (1.3). Assume (A1) - (A3) and let $\psi(\mu)$ be non-empty, bounded. Let $V \subset I\!R^m$ be some bounded, open, convex set containing $\psi(\mu)$ and suppose that (3.4) - (3.8) hold.*
Then there exist constants $L > 0$, $\delta > 0$ such that $\psi(\nu) \neq \emptyset$ and

$$d_H(\psi(\mu), \psi(\nu)) \leq L \cdot W_1(\mu, \nu)^{1/2}$$

whenever $\nu \in \mathfrak{M}_1(I\!R^S)$, $W_1(\mu, \nu) < \delta$.

Proof: Use the strong convexity of Q instead of the inequality from Corollary 2.3.2 and repeat the proof of Theorem 2.4.1. ■

Also in the above theorem the exponent 1/2 on the right-hand side of the estimate is best possible (cf. Example 2.3 in [95], which fits the above setting, too). If the functions $\tilde{Q}$ and Q are not strongly convex but fulfil some uniform convexity as sketched for $\tilde{Q}$ in Remark 2.2.5, then the technique used in the proof of Theorem 2.4.1 leads to a quantification of solution convergence with the inverse of the convexity module on the right-hand side of the estimate. This fits into the general framework for quantitative stability presented in [3].
The strong-convexity issue is also relevant in connection with stability results obtained by other authors: Variants of the second-order sufficient condition (SOSC) have a central place in Dupačová's investigations on the stability of recourse problems (cf. [25], [27]). Obviously, strong convexity is closely related to the SOSC (cf. also [87]). Replacing the SOSC by the strong convexity of Q on suitable subspaces enables us to perform a stability analysis in the sense of [25], [27] under less restrictive differentiability assumptions on Q. Shapiro [116] (cf. Chapter 3) developed a quantitative version of the upper semicontinuity of the mapping ψ for recourse models. His analysis is based on an at least quadratic growth of the objective in the unperturbed problem along feasible directions near the (possibly multivalued) solution set. The connection to (4.1) is evident and so strong convexity of Q (or $\tilde{Q}$) contributes to verifying Shapiro's growth condition.
The Theorems 2.2.2 and 2.3.1 also provide some guidelines to build the model (1.1) - (1.3) such that Q becomes strongly convex. For instance, if the second-stage costs q fulfil (A2) but not (A2)*, then a slight raise of q as mentioned in Remark 2.2.7 can be a remedy. The material in [88] gives the necessary argument that a perturbation of q is possible from the viewpoint of stability. Another possiblity to achieve a better conditioned model via the strong convexity of the objective is to implant random elements into A such that the assumptions (3.4) - (3.8) in Theorem 2.3.1 are satisfied. If, for instance, A is originally non-random and the distribution μ_1 of

z fulfils (A4)*, then one could try to randomize suitable components of A (e.g. by discrete random variables) such that z and A are independent and (3.5) - (3.8) hold. By Theorem 2.3.1 the resulting Q is strongly convex. The corresponding recourse model now differs from the original one in its underlying probability distribution. However, the randomization can be organized in connection with the weak convergence of probability measures, and, again, it is possible to benefit from stability results when justifying the model change.

Acknowledgement

This chapter has benefitted from discussions with Werner Römisch, Matthias Gelbrich and Matthias Nowak (all Humboldt-Universität Berlin).

Chapter 3

Lipschitz Stability for Stochastic Programs with Complete Recourse

Abstract
This chapter investigates the stability of optimal-solution sets to stochastic programs with complete recourse, where the underlying probability measure is understood as a parameter varying in some space of probability measures. In [116] Shapiro has proved Lipschitz upper semicontinuity of the solution set mapping. Inspired by this result we introduce a subgradient distance for probability distributions and establish the persistence of optimal solutions. For a subclass of recourse models we show that the solution set mapping is (Hausdorff) Lipschitz continuous with respect to the subgradient distance. Moreover, the subgradient distance is estimated above by the Kolmogorov-Smirnov distance of certain distribution functions related to the recourse model. The Lipschitz continuity result is illustrated by verifiable sufficient conditions for stochastic programs to belong to the mentioned subclass and by examples showing its validity and limitations. Finally, the Lipschitz continuity result is used to derive some new results on the asymptotic behaviour of optimal solutions when the probability measure underlying the recourse model is estimated via empirical measures (law of iterated logarithm, large deviation estimate, estimate for asymptotic distribution).
This chapter is joint work with Werner Römisch (Humboldt University Berlin).

3.1 Introduction

We study quantitative stability and asymptotic properties of optimal solutions to stochastic programs with complete recourse. The latter are given by

$$(1.1) \qquad P(\mu) \qquad \min\{g(x) + \tilde{Q}_\mu(Ax) : \; x \in C\},$$

where

$$(1.2) \qquad \tilde{Q}_\mu(\chi) = \int\limits_{I\!R^s} \Phi(z-\chi)\mu(\mathrm{d}z)$$

and

$$(1.3) \qquad \Phi(t) = \min\{q^T y : Wy = t, \; y \geq 0\}.$$

For the data we assume that $g : I\!R^m \to I\!R$ is a convex function, $C \subset I\!R^m$ is a non-empty closed convex set, $q \in I\!R^{\bar{m}}$ and A, W are matrices of proper dimensions. As indicated in (1.1), the integrating probability measure μ is understood as a parameter for which we assume to vary in $\mathfrak{M}_1(I\!R^s)$ - the space of all Borel probability measures on $I\!R^s$ with finite first moment, i.e. $\int\limits_{I\!R^s} \|z\|\mu(\mathrm{d}z) < +\infty$ for all $\mu \in \mathfrak{M}_1(I\!R^s)$. Further assumptions that ensure (1.1) – (1.3) to be well-defined will be given in Section 3.2.
It is well-known that (1.1) – (1.3) models a two-stage decision process under uncertainty with first-stage decision x, random entry z and second-stage (or recourse) decision y. For a more detailed introduction into this class of models, including a basic analysis of the function $\tilde{Q}_\mu$ in (1.2), we refer to [46], [131]. Here we only mention that $\tilde{Q}_\mu$ is convex whenever it is well-defined.
In the present chapter, the accent is on studying the impact of changes in the underlying probability measure μ on the problem (1.1). To this end, we assign to $\mu \in \mathfrak{M}_1(I\!R^s)$ the (global) optimal value $\varphi(\mu)$ and the set of (global) optimal solutions $\psi(\mu)$. The mappings φ and ψ are common objects of study in the stability analysis of optimization problems. In the context of stochastic programming the above set up (i.e. understanding the underlying measure as the quantity subjected to perturbations) has two principal origins: the numerical intractability of the integral in (1.2) and the incomplete information on μ that one is faced with in general. In the first case, approximations of a complicated measure μ by simpler ones give rise to a perturbation analysis. In the second case, perturbations come in via attempts to construct some "reasonable" measure μ based on the (statistical) information that is available on the random parameter z. For more details on the stability of

stochastic programs we refer to [28], [29], [48], [54], [88], [93], [111], [112], [113], [124], [133], [134].

The subsequent analysis is entirely concerned with quantitative continuity properties of the optimal-set mapping ψ. As in [93], [95] we dispense with the assumption that the solution set of the unperturbed problem is a singleton.

For the model (1.1) – (1.3), uniqueness of optimal solutions is rather exceptional as is seen by the following example. Let us already mention that the example does fit the setting of our central stability estimate, in particular, the function $\tilde{Q}_\mu$ is here strongly convex on a suitable subset (cf. Theorem 3.2.4 below).

Example 3.1.1 *Let in (1.1) – (1.3) $m = 2$, $s = 1$, $g(x) \equiv 0$, $A = (1,0)$, $(0,0)^T \in C$, $\bar{m} = 2$, $q = (1,1)^T$, $W = (1,-1)$ and μ be the uniform distribution on the closed interval $[-1/2, 1/2]$. Then it is straightforward to see that $\psi(\mu) = \ker A \cap C = \{(0,\xi)^T \in C,\ \xi \in I\!R\}$.*

One observes that $\tilde{Q}_\mu$ in (1.2) is always constant on translates of the null space $\ker A$ of A. Hence, uniqueness of optimal solutions is only guaranteed if the constraint set C picks just one element out of the relevant level set of $\tilde{Q}_\mu$.

The present investigations have been stimulated by recent results of Shapiro. In [116] the author proves an upper Lipschitz continuity estimate for ψ under the assumption that, for the unperturbed problem $P(\mu)$, the objective function grows at least quadratically for feasible points near the set of optimal solutions. The right-hand side of the estimate essentially consists of the maximal norm of elements arising in the Clarke subdifferential ([15]) of the function $\tilde{Q}_\nu - \tilde{Q}_\mu$ (cf. (1.2)) at points belonging to a suitable neighbourhood. Here, $\tilde{Q}_\nu$ corresponds to the perturbed problem $P(\nu)$, $\nu \in \mathfrak{M}_1(I\!R^s)$.

In the present chapter we introduce a "subgradient distance" for $\mu, \nu \in \mathfrak{M}_1(I\!R^s)$ based on the above maximal norm (cf. (2.1) below). We focus on the stability of models which fit into (1.1) – (1.3) and obey the additional properties that g is convex quadratic, C is a non-empty polyhedron and $\tilde{Q}_\mu$ is strongly convex on a suitable neighbourhood of $A(\psi(\mu))$. Then, the Lipschitz upper semicontinuity extends to Lipschitz continuity of the Hausdorff distance of solution sets (Theorem 3.2.4). Since the subgradient distance of $\mu, \nu \in \mathfrak{M}_1(I\!R^s)$ can always be estimated above by the Kolmogorov-Smirnov distance of certain distribution functions related to μ, ν and the algebra in (1.3) (Corollary 3.2.5, Remark 3.2.6), this leads to a powerful tool for quantitative statements on the stability of optimal solutions. In Section 3.3, one such application is worked out in detail – in the presence of empirical measures we derive some new results on the asymptotic behaviour of solution sets (law of

iterated logarithm, large deviation estimate, estimate for asymptotic distribution). In particular, former results are extended to the case where solution sets are not necessarily singletons. For the large deviation result no additional assumptions on the underlying probability measure μ are required.
Some further propositions and examples serve to supplement and illustrate the main issue of the chapter: In Proposition 3.2.3 the persistence of optimal solutions under perturbations in the "subgradient distance" is addressed. Corollary 3.2.13 displays some handy conclusions for the special case of "simple recourse". Examples in Section 3.2 show that Shapiro's assumptions in [116] do not guarantee the lower semicontinuity of ψ (Example 3.2.6), that Theorem 3.2.4 is lost for general convex g and C (Example 3.2.7) and that the setting of Theorem 3.2.4 does not guarantee stability of the optimal value (Example 3.2.9).
Compared to [93, 95], where the stability analysis is based on the L_1-Wasserstein distance and where Hölder continuity (with exponent 1/2) was obtained, the present chapter leads to Lipschitz continuity. To outweigh improvements over former results let us first mention that there exist sequences of measures where both the results from [93, 95] and the present one lead to the same convergence rates (cf. the below discussion after Proposition 3.2.13). On the other hand, there are important specific modes of perturbation (contaminated distributions, empirical measures) where the Hölder result in [93, 95] yields the rate 1/2, whereas the present approach leads to the rate 1 (Proposition 3.2.14, Section 3.3).

3.2 Stability

The following basic assumptions are well-known to ensure that the function $\tilde{Q}_\mu$ in (1.2) is well-defined and convex on $I\!R^m$ (cf. [46], [131]):

(A1) $\{Wy:\ y \in I\!R^{\bar{m}}_+\} = I\!R^s$ (complete recourse),

(A2) $\{u \in I\!R^s:\ W^T u \leq q\} \neq \emptyset$ (dual feasibility),

(A3) $\mu \in \mathfrak{M}_1(I\!R^s)$.

For arbitrary $\mu, \nu \in \mathfrak{M}_1(I\!R^s)$ and some fixed non-empty, closed, convex set $U \subset I\!R^m$ we define the following "subgradient distance" d of μ and ν:

$$d(\mu, \nu; U) = \sup\{\|z^*\| : z^* \in \partial(\tilde{Q}_\nu - \tilde{Q}_\mu)(Ax),\ x \in U\}. \tag{2.1}$$

Here, "∂" denotes Clarke's subdifferential ([15]). Since both $\tilde{Q}_\nu$ and $\tilde{Q}_\mu$ are convex, their difference is locally Lipschitzian and, hence, the Clarke subdifferential in

(2.1) is well-defined. Provided that U is bounded one uses simple properties of the Clarke subdifferential to show that $d(\cdot,\cdot;U)$ is a pseudo-metric on $\mathfrak{M}_1(I\!R^s)$; note that $d(\mu,\nu;U)=0$ is possible for $\mu\neq\nu$. Subsequent considerations involving d will use the fact that, in finite dimension, the Clarke subdifferential may be represented as the convex hull of limits of sequences of gradients collected at differentiability points and possibly avoiding arguments in a set of Lebesgue measure zero (Theorem 2.5.1 in [15]). The following lemma provides some more insight into d.

Lemma 3.2.1 *Let $h_1,h_2 : I\!R^s \to I\!R$ be locally Lipschitzian. Then it holds for arbitrary $\chi\in I\!R^s$ that $d_H(\partial h_1(\chi),\partial h_2(\chi)) \le \sup\{\|z^*\| : z^*\in\partial(h_1-h_2)(\chi)\}$, where d_H denotes the Hausdorff distance of sets.*

Proof: For $\partial h_i(\chi)$, $i=1,2$, we have the following representation (Theorem 2.5.1 in [15]):

$$\partial h_i(\chi) = \mathrm{conv}\ \mathcal{L}_{h_i}(\chi)$$

where

$$\begin{aligned}\mathcal{L}_{h_i}(\chi) \;=\; \{z:\ &\text{there exist }\ \chi_n\in \mathrm{Diff}(h_1)\cap\mathrm{Diff}(h_2)\ \text{ such that}\\ &\chi_n\to\chi\ \text{ and }\ h_i'(\chi_n)\to z\ \text{ as }\ n\to\infty\}.\end{aligned}$$

Here $\mathrm{Diff}(h_i)$ denotes the set of differentiability points of h_i. Clearly, $\mathrm{Diff}(h_1)\cap\mathrm{Diff}(h_2)\subset\mathrm{Diff}(h_i)$ and, by Rademacher's Theorem $I\!R^s\setminus(\mathrm{Diff}(h_1)\cap\mathrm{Diff}(h_2))$ has Lebesgue measure zero.
Assume that

$$d_H(\partial h_1(\chi),\partial h_2(\chi)) > \sup\{\|z^*\| : z^*\in\partial(h_1-h_2)(\chi)\}$$

for some $\chi\in I\!R^s$.
This implies

$$d_H(\mathcal{L}_{h_1}(\chi),\mathcal{L}_{h_2}(\chi)) > \sup\{\|z^*\| : z^*\in\partial(h_1-h_2)(\chi)\},$$

and, hence, by the definition of the Hausdorff distance, there exists a $z_{1,0}^*\in\mathcal{L}_{h_1}(\chi)$ (without loss of generality) such that

$$\|z_{1,0}^*-z_2^*\| > \sup\{\|z^*\| : z^*\in\partial(h_1-h_2)(\chi)\} \tag{2.2}$$

for all $z_2^*\in\mathcal{L}_{h_2}(\chi)$.
Since $z_{1,0}^*\in\mathcal{L}_{h_1}(\chi)$, there exists a sequence of points $\chi_n\in\mathrm{Diff}(h_1)\cap\mathrm{Diff}(h_2)$ such that $h_1'(\chi_n)\to z_{1,0}^*$ as $n\to\infty$.

Now consider the sequence $\{h_2'(\chi_n)\}$. By the local Lipschitz property of h_2 it has an accumulation point $z_{2,0}^*$ that obviously belongs to $\mathcal{L}_{h_2}(\chi)$. In view of (2.2)

$$\|z_{1,0}^* - z_{2,0}^*\| > \sup\{\|z^*\| : z^* \in \partial(h_1 - h_2)(\chi)\},$$

but, on the other hand, $z_{1,0}^* - z_{2,0}^* \in \mathcal{L}_{h_1-h_2}(\chi) \subset \partial(h_1 - h_2)(\chi)$, which is an obvious contradiction. ■

In our quantitative stability analysis for optimal solutions of perturbed stochastic programs d will be the distance to measure "how far" a perturbation $P(\nu)$ is away from the original program $P(\mu)$. In the context of stochastic programming also Shapiro [116] has used information contained in the definition of d to derive quantitative stability properties. Kummer [63] has results on the quantitative stability of general convex programs based on the Hausdorff distance of subgradients, which appears in Lemma 3.2.1. Our considerations start with the following result by Shapiro [116]:

Theorem 3.2.2 *Suppose* (A1) - (A3), $\psi(\mu) \neq \emptyset$ *and that*

(2.3) *there exists a convex open set* U_o *containing* $\psi(\mu)$ *and a constant* $\alpha > 0$ *such that*
$g(x) + \tilde{Q}_\mu(Ax) \geq \varphi(\mu) + \alpha \cdot \mathrm{dist}(x, \psi(\mu))^2$ *for all* $x \in C \cap U_o$,
where dist *denotes the usual point-to-set distance.*

Then the following estimate is valid for all $\nu \in \mathfrak{M}_1(I\!R^s)$

$$\sup_{x \in \psi(\nu) \cap U_o} \mathrm{dist}(x, \psi(\mu)) \leq \alpha^{-1} \cdot d(\mu, \nu; \mathrm{cl}\, U_o),$$

where the left-hand side is defined to be zero if $\psi(\nu) \cap U_o = \emptyset$.

The above theorem asserts the upper Lipschitz continuity of the solution set mapping ψ with respect to the pseudo-metric d. It does not contain the persistence of optimal solutions, i.e. it is not clear whether the perturbed program $P(\nu)$ has a non-empty set of optimal solutions if $d(\mu, \nu; \mathrm{cl}\, U_o)$ is sufficiently small. The next proposition answers this question.

Proposition 3.2.3 *Suppose* (A1) - (A3) *and that* $\psi(\mu)$ *is non-empty and bounded. Let* $U_o \subset I\!R^m$ *be an open, convex, bounded set containing* $\psi(\mu)$.
Then there exists a constant $\delta > 0$ *such that*

$$\emptyset \neq \psi(\nu) \subset U_o$$

for all $\nu \in \mathfrak{M}_1(I\!R^s)$ *such that* $d(\mu, \nu; \mathrm{cl}\, U_o) < \delta$.

Proof: We introduce the following notations:

$$\begin{aligned} G(x,\nu) &:= g(x) + \tilde{Q}_\nu(Ax),\ \nu \in \mathfrak{M}_1(I\!R^s), \\ \psi_d(\mu) &:= \operatorname{argmin}\{G(x,\mu) + d^T x : x \in C\},\ d \in I\!R^m,\ \text{and} \\ \mathcal{N}_r(M) &:= \{x \in I\!R^m : \operatorname{dist}(x,M) < r\},\ M \subset I\!R^m,\ r > 0. \end{aligned}$$

Select some $r > 0$ such that $\mathcal{N}_r(\psi(\mu)) \subset U_o$. Since $\psi(\mu)$ is bounded and $G(\cdot,\mu)$ is convex, well-known results on the stability of convex programs apply. In particular, Theorem 4.3.3 and Corollary 4.3.3.2 from [4] imply that there exists a constant $\delta' > 0$ such that

$$\emptyset \neq \psi_d(\mu) \subset \mathcal{N}_r(\psi(\mu)) \quad \text{for all } d \in I\!R^m \text{ with } \|d\| < \delta'. \tag{2.4}$$

In order to apply results on the stability of certain generalized equations ([63]) we introduce the set-valued mappings $\Gamma_\nu : \operatorname{cl} U_o \to I\!R^m$, $\nu \in \mathfrak{M}_1(I\!R^s)$, given by $\Gamma_\nu(x) = \partial_x G(x,\nu) + N_C(x)$. Here, ∂_x denotes the subdifferential of $G(\cdot,\nu)$ and $N_C(x)$ the normal cone to C at x, both in the sense of convex analysis ([89]).
Of course, $x \in \psi_d(\mu)$ is equivalent to $-d \in \Gamma_\mu(x)$.
The compactness of $\operatorname{cl} U_o$, elementary properties of the convex subdifferential and the normal-cone operator $N_c(\cdot)$ together with relation (2.4) now imply that the assumptions of Proposition 6 in [63] are fulfilled. Proposition 3 in [63] then says that Γ_μ is a regular multifunction, i.e. there exists a constant $\tilde{\delta} > 0$ such that the generalized equation

$$0 \in \tilde{\Gamma}(x), \quad x \in \operatorname{cl} U_o$$

is solvable for any admissible multifunction $\tilde{\Gamma}$ satisfying

$$\Gamma_\mu(x) \subset \tilde{\Gamma}(x) + \tilde{\delta} B_m \quad \text{for all} \quad x \in \operatorname{cl} U_o,$$

where $B_m \subset I\!R^m$ denotes the closed unit ball.
For the definition of admissibility we refer to [63]. For our purposes it is sufficient to know that upper semicontinuous multifunctions with nonempty, closed, convex image sets (hence, all the mappings Γ_ν) are admissible.
Let $\nu \in \mathfrak{M}_1(I\!R^s)$ such that $d(\mu,\nu;\operatorname{cl} U_o) < \tilde{\delta}$. Lemma 3.2.1 now implies that

$$\Gamma_\mu(x) \subset \Gamma_\nu(x) + \tilde{\delta} B_m \qquad \text{for all} \quad x \in \operatorname{cl} U_o.$$

By the regularity of Γ_μ this yields that $\psi(\nu)$ is non-empty whenever $d(\mu,\nu;\operatorname{cl} U_o) < \tilde{\delta}$. Now select $\delta > 0$ such that $\delta < \min\{\delta', \tilde{\delta}\}$. Let $\nu \in \mathfrak{M}_1(I\!R^s)$ such that

$d(\mu, \nu; \mathrm{cl}\, U_o) < \delta$. Then it holds $\psi(\nu) \cap \mathrm{cl}\, U_o \neq \emptyset$. Let us assume that $\psi(\nu) \not\subset U_o$. By convexity, this yields $\psi(\nu) \cap \mathrm{bd}\, U_o \neq \emptyset$. Let $\tilde{x} \in \psi(\nu) \cap \mathrm{bd}\, U_o$. It holds

$$\Gamma_\nu(\tilde{x}) \subset \Gamma_\mu(\tilde{x}) + \delta B_m \quad \text{and} \quad 0 \in \Gamma_\nu(\tilde{x}).$$

Hence, there exists a $\tilde{d} \in I\!\!R^m$, $\|\tilde{d}\| < \delta'$ such that $\tilde{d} \in \Gamma_\mu(\tilde{x})$.
By (2.4) this implies $\tilde{x} \in \mathcal{N}_r(\psi(\mu))$, contradicting $\mathcal{N}_r(\psi(\mu)) \cap \mathrm{bd}\, U_o = \emptyset$, and the proof is complete ■

Now we direct our attention to stochastic programs for which Theorem 3.2.2 extends to the Lipschitz continuity of ψ with respect to the Hausdorff distance of sets and the pseudo-metric d of probability measures.

Theorem 3.2.4 *Suppose* (A1) - (A3) *and that $\psi(\mu)$ is non-empty and bounded. Let g be a convex quadratic function and C be a polyhedron. Assume that there exists a convex open subset V of $I\!\!R^s$ such that $A(\psi(\mu)) \subset V$ and the function $\tilde{Q}_\mu$ is strongly convex on V. Let $U = \mathrm{cl}\, U_o$, where U_o is an open, convex, bounded set such that $\psi(\mu) \subset U_o$ and $A(U) \subset V$.*
Then there exist constants $L > 0$, $\delta > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \leq L \cdot d(\mu, \nu; U)$$

whenever $\nu \in \mathfrak{M}_1(I\!\!R^s)$, $d(\mu, \nu; U) < \delta$.

Recall that $\tilde{Q}_\mu$ is said to be strongly convex on V if there exists a constant $\kappa > 0$ such that for all $\chi, \tilde{\chi} \in V$ and $\lambda \in [0, 1]$

$$\tilde{Q}_\mu(\lambda\chi + (1-\lambda)\tilde{\chi}) \leq \lambda \tilde{Q}_\mu(\chi) + (1-\lambda)\tilde{Q}_\mu(\tilde{\chi}) - \kappa\lambda(1-\lambda)\|\chi - \tilde{\chi}\|^2.$$

Proof: Given an open ball B_∞ (with respect to the norm $\|\cdot\|_\infty$ and around zero) such that $\psi(\mu) \subset B_\infty$, we select a $\delta > 0$ such that $\emptyset \neq \psi(\nu) \subset B_\infty$ for all $\nu \in \mathfrak{M}_1(I\!\!R^s)$, $d(\mu, \nu; U) < \delta$ (Proposition 3.2.3). We denote $C_o := C \cap \mathrm{cl}\, B_\infty$. Note that the compact set C_o is again a polyhedron. Then it holds for all $\nu \in \mathfrak{M}_1(I\!\!R^s)$, $d(\mu, \nu; U) < \delta$,

$$\psi(\nu) = \mathrm{argmin}\{g(x) + \tilde{Q}_\nu(Ax) : \ x \in C_o\}.$$

Furthermore, the compactness of C_o guarantees

$$\begin{aligned}\min_x\{g(x) + \tilde{Q}_\nu(Ax) : x \in C_o\} &= \min_{x,v}\{g(x) + \tilde{Q}_\nu(v) : Ax = v, \ x \in C_o\} \\ &= \min_v\{\tilde{Q}_\nu(v) + \min_x\{g(x) : Ax = v, \ x \in C_o\} : v \in A(C_o)\}.\end{aligned}$$

Introducing $\pi(v) := \min_x\{g(x) : Ax = v, x \in C_o\}$, $\bar{X}(v) := \operatorname{argmin}\{g(x) : Ax = v, x \in C_o\}$ and $\bar{Y}(\nu) := \operatorname{argmin}\{\tilde{Q}_\nu(v) + \pi(v) : v \in A(C_o)\}$, we obtain by verification of the respective inclusions

$$\psi(\nu) = \bar{X}(\bar{Y}(\nu))$$

for all $\nu \in \mathfrak{M}_1(I\!R^s)$, $d(\mu, \nu; U) < \delta$.

The multifunction $\bar{X}(\cdot)$ is Lipschitzian on its effective domain $\operatorname{dom} \bar{X} = \{v \in I\!R^s : \bar{X}(v) \neq \emptyset\}$. (Remark at p. 220 in [59], Satz 4.3.3 in [55], cf. the Appendix for a display of the proof.) Therefore, there exists a constant $L_o > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) = d_H(\bar{X}(\bar{Y}(\nu)), \bar{X}(\bar{Y}(\mu))) \leq L_o \cdot \sup_{v \in \bar{Y}(\nu)} \|v - v_\mu\| \tag{2.5}$$

whenever $\nu \in \mathfrak{M}_1(I\!R^s)$, $d(\mu, \nu; U) < \delta$.

Since $\pi(\cdot)$ is convex on $A(C_o)$ and $\tilde{Q}_\mu$ is strongly convex on $V \supset A(\psi(\mu))$, the set $\bar{Y}(\mu)$ reduces to a singleton $\{v_\mu\}$. Moreover, the function $G(v, \mu) := \tilde{Q}_\mu(v) + \pi(v)$ is strongly convex on V with modulus $\kappa > 0$.

Decrease, if necessary, $\delta > 0$ such that $\psi(\nu) \subset U_o$ whenever $\nu \in \mathfrak{M}_1(I\!R^s)$, $d(\mu, \nu; U) < \delta$ (Proposition 3.2.3). Let $v \in \bar{Y}(\nu)$. Then $\bar{X}(v) \subset \psi(\nu)$. Since $\psi(\nu) \subset U_o$ and $\{v\} = A\bar{X}(v)$ it follows $v \in A(U_o) \subset V$. Consider the point $\frac{1}{2}(v + v_\mu)$ belonging to $A(C_o) \cap V$.

Then

$$\begin{aligned} G(v_\mu, \mu) &\leq G(\frac{1}{2}(v + v_\mu), \mu) \\ &\leq \frac{1}{2}G(v, \mu) + \frac{1}{2}G(v_\mu, \mu) - \frac{\kappa}{4}\|v - v_\mu\|^2, \end{aligned}$$

by the strong convexity of $G(\cdot, \mu)$ on V.

This implies (if $\|v - v_\mu\| \neq 0$)

$$\begin{aligned} \frac{\kappa}{2}\|v - v_\mu\| &\leq \frac{G(v, \mu) - G(v_\mu, \mu)}{\|v - v_\mu\|} \\ &= \frac{(G(v_\mu, \nu) - G(v_\mu, \mu)) - (G(v_\mu, \nu) - G(v, \mu))}{\|v - v_\mu\|} \\ &\leq \frac{(G(v_\mu, \nu) - G(v_\mu, \mu)) - (G(v, \nu) - G(v, \mu))}{\|v - v_\mu\|} \\ &\leq \frac{|(\tilde{Q}_\nu - \tilde{Q}_\mu)(v_\mu) - (\tilde{Q}_\nu - \tilde{Q}_\mu)(v)|}{\|v - v_\mu\|} \end{aligned}$$

By Lebourg's mean value theorem ([15]) there exists a point v^* on the line segment $[v_\mu, v]$ (entirely belonging to $A(U)$) such that the above estimate continues

$$\leq \sup \Big| \Big\langle \partial(\tilde{Q}_\nu - \tilde{Q}_\mu)(v^*), \frac{v_\mu - v}{\|v_\mu - v\|} \Big\rangle \Big|.$$

Hence,

$$\begin{aligned} \sup_{v \in \bar{Y}(\nu)} \|v - v_\mu\| &\leq \frac{2}{\kappa} \sup\{\|z^*\| : z^* \in \partial(\tilde{Q}_\nu - \tilde{Q}_\mu)(v) : v \in A(U)\} \\ &= \frac{2}{\kappa} d(\mu, \nu; U). \end{aligned}$$

Together with (2.5) this completes the proof. ■

Corollary 3.2.5 *Adopt the setting of Theorem* 3.2.4.
Then there exist non-singular matrices $B_i (i = 1, \ldots, \ell)$ and a constant $L > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \leq L \sum_{i=1}^{\ell} \sup_{t \in A(U)} |F_{\mu \circ (-B_i)}(-B_i^{-1} t) - F_{\nu \circ (-B_i)}(-B_i^{-1} t)|$$

whenever $\nu \in \mathfrak{M}_1(\mathbb{R}^s)$ is chosen such that the right-hand side is sufficiently small. The notation F refers to the distribution function of the probability measure in the subscript.

Proof. Using Theorem 2.5.1 in [15] we obtain the following representation of $d(\mu, \nu; U)$

$$d(\mu, \nu; U) = \sup\{\|\nabla(\tilde{Q}_\nu - \tilde{Q}_\mu)(Ax)\| : \; x \in U \setminus E\}, \tag{2.6}$$

where E contains those $x \in \mathbb{R}^m$ such that $\tilde{Q}_\nu - \tilde{Q}_\mu$ is not differentiable at Ax and $A(E)$ has Lebesgue measure zero.
Recall that the integrand Φ in (1.2) is a piecewise linear convex function on $\mathbb{R}^s$ and that there exist basis submatrices $B_1, \ldots, B_\ell$ of W such that the simplicial cones $B_1(\mathbb{R}^s_+), \ldots, B_\ell(\mathbb{R}^s_+)$ are linearity regions of Φ (in general not the maximal ones) (cf. [72], [126] or Proposition 2.2.1 in Chapter 2). Of course, $\bigcup\limits_{i=1}^{\ell} B_i(\mathbb{R}^s_+) = \mathbb{R}^s$, and B_i $(i = 1, \ldots, \ell)$ can be chosen in such a way that $B_i(\mathbb{R}^s_+) \cap B_j(\mathbb{R}^s_+)$, $i \neq j$, is always contained in a hyperplane in $\mathbb{R}^s$. Thus, $\mathcal{H} := \mathbb{R}^s \setminus \bigcup\limits_{i=1}^{\ell} \operatorname{int} B_i(\mathbb{R}^s_+)$ is contained in a finite union of hyperplanes in $\mathbb{R}^s$.

Let us now confirm that, for some single hyperplane $\mathcal{H}_o \subset I\!R^s$, the set

$$\mathcal{Z}_{\nu,o} := \{\chi \in I\!R^s : \nu(\chi + \mathcal{H}_o) > 0\}$$

has Lebesgue measure zero: It holds

$$\begin{aligned} \chi + \mathcal{H}_o &= \chi + \{t \in I\!R^s : a^T t = 0\} \\ &= \{t \in I\!R^s : a^T(t - \chi) = 0\} = a^{-1}(\{a^T\chi\}), \end{aligned}$$

where $a : I\!R^m \to I\!R$ denotes the linear transformation induced by a^T and a^{-1} is the pre-image.
Hence, $\mathcal{Z}_{\nu,o} = \{\chi \in I\!R^s : \nu \circ a^{-1}(\{a^T\chi\}) > 0\}$. Now $\nu \circ a^{-1}$ is a probability measure on $I\!R$ and $a^T\chi$ is an atom of $\nu \circ a^{-1}$. Since $\nu \circ a^{-1}$ has at most countably many atoms, $\mathcal{Z}_{\nu,o}$ is contained in a countable union of hyperplanes and has Lebesgue measure zero.
Therefore the sets

$$\mathcal{Z}_\mu := \{\chi \in I\!R^s : \mu(\chi + \mathcal{H}) > 0\} \ \text{ and } \ \mathcal{Z}_\nu := \{\chi \in I\!R^s : \nu(\chi + \mathcal{H}) > 0\}$$

have Lebesgue measure zero.
Now select E in (2.6) as the pre-image $A^{-1}(\mathcal{Z}_\mu \cup \mathcal{Z}_\nu)$.
Then $\tilde{Q}_\nu - \tilde{Q}_\mu$ is differentiable at Ax for all $x \in U \setminus E$ and we have for those x

$$\begin{aligned} \nabla(\tilde{Q}_\nu - \tilde{Q}_\mu)(Ax) &= \int_{I\!R^s} \nabla\Phi(z - Ax)(\nu - \mu)(\mathrm{d}z) \\ &= \int_{\bigcup_{i=1}^{\ell} \operatorname{int} B_i(I\!R^s_+)} \nabla\Phi(z - Ax)(\nu - \mu)(\mathrm{d}z) \\ &= \sum_{i=1}^{\ell} d_i \cdot (\nu - \mu)(Ax + B_i(I\!R^s_+)) \\ &= \sum_{i=1}^{\ell} d_i(F_{\nu\circ(-B_i)}(-B_i^{-1}Ax) - F_{\mu\circ(-B_i)}(-B_i^{-1}Ax)), \end{aligned}$$

where $-d_i$ is the gradient of Φ on $\operatorname{int} B_i(I\!R^s_+)$, $i = 1, \ldots, \ell$.
The assertion now immediately follows from Theorem 3.2.4, (2.6) and the above identity. Since $A(U \setminus E)$ is less explicitly known than $A(U)$, we finally take the supremum over the larger set $A(U)$. ∎

Remark 3.2.6 *The above estimate is closely related to Theorem 2.1 in [116] where the author uses the normal cones $\bar{C}_j$ $(j = 1, \ldots, \tilde{\ell})$ to the set $\{u \in I\!R^s : W^T u \leq q\}$ at its vertices $\tilde{d}_j$ $(j = 1, \ldots, \tilde{\ell})$. From linear parametric programming it is known*

([72], [126]) that each of the cones $\bar{C}_j$ $(j = 1, \dots, \tilde{\ell})$ is the union of certain cones $B_i(\mathbb{R}^s_+)$ $(i \in \{1, \dots, \ell\})$ arising in the above corollary. We have preferred to use the cones $B_i(\mathbb{R}^s_+)$ since these are simplicial cones, which allows a direct relation to distribution functions.

Remark 3.2.7 *Consider the right-hand side of the estimate in Corollary 3.2.5 and take the suprema with respect to $t \in \mathbb{R}^s$ instead of $t \in A(U)$. In this way we obtain a Lipschitz estimate with respect to the uniform (or Kolmogorov-Smirnov) distance of the distribution functions $F_{\mu\circ(-B_i)}$ and $F_{\nu\circ(-B_i)}$ $(i = 1, \dots, \ell)$.*

Remark 3.2.8 *Theorem 3.2.4 remains valid under any hypotheses on g and C leading to Lipschitz continuity of the multifunction $\bar{X}(v) := \mathrm{argmin}\{g(x) : Ax = v, x \in C\}$.*

The next example shows that (already for contaminated distributions) Theorem 3.2.4 is lost for a general closed convex set $C \subset \mathbb{R}^m$. Another counterexample involving the function g can be constructed following the guidelines of Remark 2.9 in [93].

Example 3.2.9 *Let in (1.1) - (1.3) $m = 2$, $s = 1$, $g(x) \equiv 0$, $A = (1, 0)$, $C = \{x \in \mathbb{R}^2 : (x_2)^2 \leq x_1\}$, $q = (1, 1)^T$, $W = (1, -1)$ and μ be the uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$. Let δ_1 denote the probability measure on $\mathbb{R}$ having unit mass at 1 and construct perturbations μ_t of μ by setting $\mu_t = (1-t)\mu + t\delta_1$, $t \in [0, 1]$. Then, $\psi(\mu) = \{0\}$ and the strong-convexity assumption for $\tilde{Q}_\mu$ holds for $V = (-\frac{1}{2}, \frac{1}{2})$. Furthermore, one computes that $(x_{1,t}, \sqrt{x_{1,t}})^T \in \psi(\mu_t)$ for $0 < t < 2/3$, where $x_{1,t} = t/2(1-t)$. Hence, $x_{1,t} > \frac{1}{2}t$ and $d_H(\psi(\mu), \psi(\mu_t)) \geq \frac{1}{\sqrt{2}}\sqrt{t}$ for all $t \in (0, 2/3)$. With $U \subset \mathbb{R}^2$ taken as the closed ball around zero with radius $1/2$ (for instance) one confirms that $d(\mu, \mu_t; U) = \mathrm{const} \cdot t$, i.e. the assertion of Theorem 3.2.4 does not hold.*

Note that in the above example there is even no upper Lipschitz continuity of ψ. This indicates that, in general, one cannot hope to obtain the second-order growth condition (2.3) in Theorem 3.2.2 without adding assumptions on g and C.

Remark 3.2.10 *Using similar techniques as in the proof of Theorem 2.7 in [93] it can be shown that the assumptions from Theorem 3.2.4 imply the second-order growth condition (2.3) in Theorem 3.2.2 to hold.*

The comparison of the Theorems 3.2.2 and 3.2.4 is completed by the following example, which demonstrates that the setting in Theorem 3.2.2 is the more general one. Indeed, Theorem 3.2.2 does not guarantee the lower semicontinuity of the mapping ψ which, of course, is a special implication of the Hausdorff-Lipschitz result in Theorem 3.2.4.

Example 3.2.11 *Let in (1.1) - (1.3)* $m = s = 1$, $g(x) \equiv 0$, $A = 1$, $C = [-1, 1]$, $q = (1, 1)^T$, $W := (1, -1)$ *("simple recourse") and* μ *be the uniform distribution on* $[-1, -\frac{1}{2}] \cup [\frac{1}{2}, 1]$. *Let* $\bar{\mu}$ *be the uniform distribution on* $[-\frac{1}{2}, \frac{1}{2}]$ *and construct perturbations* μ_t *of* μ *by setting* $\mu_t = \mu + t(\bar{\mu} - \mu)$, $t \in [0, 1]$ *("contaminated distributions"). Then one computes that* $\psi(\mu) = [-\frac{1}{2}, \frac{1}{2}]$ *and that*

$$\tilde{Q}_\mu(x) \geq \tilde{Q}_\mu(x_\mu) + [\text{dist}(x, \psi(\mu))]^2$$

for all $x \in (-1, 1)$ *and all* $x_\mu \in \psi(\mu)$. *Hence (2.3) is fulfilled and Theorem 3.2.2 applies. On the other hand,* $\psi(\mu_t) = \{0\}$ *for all* $t \in (0, 1]$ *and* $d(\mu, \mu_t; C) = \text{const} \cdot t$. *Thus,* ψ *does not share the Lipschitz property from Theorem 3.2.4, moreover,* ψ *is not lower semicontinuous at* μ.

The next example is interesting since it shows that Theorem 3.2.4 does not ensure the stability of the optimal value.

Example 3.2.12 *Let in (1.1) - (1.3)* $m = s = 1$, $g(x) \equiv 0$, $C = [-1, 1]$, $A = 1$, $q = (1, 1)$, $W = (1, -1)$, $\mu = \delta_o$ *and construct pertrubations* μ_n *of* μ *by setting* $\mu_n = (1 - \frac{1}{n})\delta_o + \frac{1}{n}\delta_{n^2}$ $(n \in I\!N)$. *Then we have* $\tilde{Q}_\mu(x) = |x|$ *and, thus,* $\varphi(\mu) = 0$, $\psi(\mu) = \{0\}$. *Furthermore* $\tilde{Q}_{\mu_n}(x) = (1 - \frac{1}{n})|x| + \frac{1}{n}(n^2 - x)$. *Therefore,* $\psi(\mu_n) = \{0\}$, $\varphi(\mu_n) = n$. *The assumptions of Proposition 2.3 and Theorem 3.2.4 are fulfilled, but* $\varphi(\mu_n) \to \infty$ *as* $n \to \infty$.

The following result (established in Chapter 2) provides a handy tool to check the strong convexity of $\tilde{Q}_\mu$ needed in Theorem 3.2.4.

Proposition 3.2.13 (Theorem 2.2.2 in Chapter 2)
Let $V \subset I\!R^s$ *be open and convex. Assume*

(a) (A1),

(b) *there exists a* $\bar{u} \in I\!R^s$ *such that* $W^T\bar{u} < q$ *componentwise,*

(c) (A3),

(d) μ *has a density* Θ_μ *on* $I\!R^s$,

(e) *there exist constants* $r > 0$, $\varrho > 0$ *such that* $\Theta_\mu(t) \geq r$ *for all* $t \in I\!R^s$ *such that* $\operatorname{dist}(t, V) \leq \varrho$.

Then $\tilde{Q}_\mu$ *is strongly convex on* V.

In [93], [95] the quantitative continuity of the mapping ψ is studied with respect to the L_1-Wasserstein distance $W_1(\mu, \nu)$ for measures μ, ν in $\mathfrak{M}_1(I\!R^s)$ [81]. In fact, Theorem 2.7 in [93] states the Hölder continuity (with exponent 1/2) of $d_H(\psi(\mu), \psi(\nu))$ with respect to $W_1(\mu, \nu)$ under precisely the same assumptions as in Theorem 3.2.4 in the present chapter. Furthermore, [93] contains an example (Remark 2.9 in [93]) showing the optimality of the convergence rate 1/2. We now analyze this example with respect to the pseudo-metric $d(\mu, \nu; U)$: The setting is as in the Examples 3.2.11 and 3.2.12, but μ is taken as the uniform distribution on $[-1/2, 1/2]$ and the perturbations μ_n are given by the distribution functions

$$F_{\mu_n}(t) = \begin{cases} 1/2 & t \in [-\varepsilon_n, \varepsilon_n) \\ F_\mu(t) & \text{otherwise,} \end{cases}$$

where (ε_n) is arbitrary and tending to zero from above $(n \to \infty)$. It holds that $\psi(\mu) = \{0\}$ and $\psi(\mu_n) = [-\varepsilon_n, \varepsilon_n]$. The assumptions of Theorem 3.2.4 are fulfilled. In [93] we computed $W_1(\mu, \mu_n) = \varepsilon_n^2$, which showed optimality of the Hölder exponent 1/2. Here, we obtain that, with $U = [-1/4, 1/4]$, $d(\mu, \mu_n; U) = \text{const} \cdot \varepsilon_n$. Hence, in the worst case Theorem 3.2.4 does not outperform Theorem 2.7 in [93]. However, for certain specific modes of perturbation, Theorem 3.2.4 yields stronger estimates than Theorem 2.7 in [93] (contaminated distributions, asymptotic properties of non-parametric estimators, see the analysis below).
To explain the essence of Theorem 3.2.4 a bit more, let us mention that Theorem 2.7 in [93] also yields the convergence rate 1/2 when replacing the Wasserstein distance W_1 by

$$d^*(\mu, \nu; U) = \sup\{|(\tilde{Q}_\nu - \tilde{Q}_\mu)(Ax)| : x \in U\},$$

where $U \subset I\!R^m$ is a suitable non-empty, convex, compact set. Therefore, the approach taken here differs from former ones by measuring the distance of the objectives in the original and the perturbed programs rather in terms of their subgradients than in terms of their function values. Of course $d(\mu, \nu; U)$ may tend to zero while $d^*(\mu, \nu; U)$ does not, which explains the collapse of optimal-value convergence observed in Example 3.2.12. Hence, when aiming at the stability of the optimal value one should resort to a distance like d^*. In [63], Proposition 8, Kummer showed that convergence to zero of d^* does imply the same for d, provided that the original function $\tilde{Q}_\mu$ is differentiable.

Proposition 3.2.14 *(contaminated distributions)*
Suppose (A1) - (A3) *and that $\psi(\mu)$ is non-empty and bounded. Let g be convex quadratic and C be polyhedral. Assume that there exists a convex open subset V of $I\!R^s$ such that $A(\psi(\mu)) \subset V$ and that the function $\tilde{Q}_\mu$ is strongly convex on V. Let $\bar{\mu} \in \mathfrak{M}_1(I\!R^s)$ be arbitrarily fixed and define $\mu_t = (1-t)\mu + t\bar{\mu}$, $t \in [0,1]$.*
Then there exist constants $L > 0$ and $t_o > 0$ such that

$$d_H(\psi(\mu), \psi(\mu_t)) \leq Lt$$

for all $t \in [0, t_o]$.

Proof: Note that $\tilde{Q}_{\mu_t} - \tilde{Q}_\mu = t(\tilde{Q}_{\bar{\mu}} - \tilde{Q}_\mu)$. Calculus rules for the Clarke subdifferential thus imply that $d(\mu; \mu_t; U) = t \cdot d(\mu, \bar{\mu}; U)$, where, of course, $d(\mu, \bar{\mu}; U)$ is a constant. The result now immediately follows from Theorem 3.2.4. ■

The following result refers to the special case of simple recourse, i.e. Φ in (1.3) is given by

$$\Phi(t) := \min\{{q^+}^T y^+ + {q^-}^T y^- : y^+ - y^- = t,\ y^+ \geq 0,\ y^- \geq 0\} \tag{2.7}$$

where $\bar{m} = 2s$ and $q^+, q^- \in I\!R^s$. By a direct estimate from Theorem 3.2.2 Shapiro has derived a similar result (Theorem 3.1 in [116]).

Proposition 3.2.15 *Let $P(\mu)$ be a simple-recourse model, $\psi(\mu)$ be non-empty and bounded, g be convex quadratic and C be a non-empty polyhedron. Assume that $q^+ + q^- > 0$ (componentwise) and that all the one-dimensional marginal distributions μ_j of μ $(j = 1, \ldots, s)$ have finite first moments and densities that are positively bounded below on some open neighbourhoods of the orthogonal projections of $\psi(\mu)$ to the coordinate axes.*
Then there exists a constant $L > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \leq L \cdot \sum_{j=1}^{s} \sup_{t_j \in \mathrm{proj}_j(A(U))} |F_{\mu_j}(t_j) - F_{\nu_j}(t_j)|$$

whenever $\nu \in \mathfrak{M}_1(I\!R^s)$ is chosen such that the right-hand side is sufficiently small.

Proof: First note that the function Φ in (2.7) is separable with respect to the components of t. Therefore, the functions $\tilde{Q}_\nu$ and $\tilde{Q}_\mu$ here only depend on the one-dimensional marginal distributions of ν and μ (cf. also [46], [131]), and we can assume without loss of generality that ν, μ are probability measures with independent one-dimensional marginals. Then our assumptions and Proposition 3.2.13 yield

that $\tilde{Q}_\mu$ is strongly convex on some convex open set $V \supset A(\psi(\mu))$.
Here, the cones $B_i(I\!R^s_+)$ $(i = 1, \ldots, 2^s)$ are orthants.
To estimate $(\nu - \mu)(Ax + B_i(I\!R^s_+))$ let us fix some $B_i(I\!R^s_+)$ for which we assume without loss of generality that

$$B_i(I\!R^s_+) = \bigtimes_{j=1}^{s_i}(-\infty, 0] \times \bigtimes_{j=s_i+1}^{s}[0, +\infty).$$

Our independence assumption then yields

$$\begin{aligned}(\nu - \mu)(Ax + B_i(I\!R^s_+)) &= \prod_{j=1}^{s_i}\nu_j((-\infty, (Ax)_j]) \cdot \prod_{j=s_i+1}^{s}\nu_j([(Ax)_j, +\infty)) \\ &\quad - \prod_{j=1}^{s_i}\mu_j((-\infty, (Ax)_j]) \cdot \prod_{j=s_i+1}^{s}\mu_j([(Ax)_j, +\infty)).\end{aligned}$$

Using the inequality

$$\Big|\prod_{j=1}^{s}\alpha_j - \prod_{j=1}^{s}\beta_j\Big| \le \sum_{j=1}^{s}|\alpha_j - \beta_j| \quad \text{for } \ 0 \le \alpha_j, \beta_j \le 1, \ j = 1, \ldots, s$$

(which can be shown by induction) we obtain

$$\begin{aligned}&|(\nu - \mu)(Ax + B_i(I\!R^s_+))| \\ &\quad\le \sum_{j=1}^{s_i}|F_{\nu_j}((Ax)_j) - F_{\mu_j}((Ax)_j)| + \sum_{j=s_i+1}^{s}|F^-_{\nu_j}((Ax)_j) - F^-_{\mu_j}((Ax)_j)|,\end{aligned}$$

where the superscripts in the last term indicate limits from the left.
For $x \in U \setminus E$ (with E as in the proof of Corollary 3.2.5) the superscripts can be dropped and we obtain

$$|(\nu - \mu)(Ax + B_i(I\!R^s_+))| \le \sum_{j=1}^{s}|F_{\nu_j}((Ax)_j) - F_{\mu_j}((Ax)_j)|$$

and the proof is completed as with Corollary 3.2.5. ■

3.3 Applications to Asymptotic Analysis

In the present section, we show how to employ the Lipschitz stability result of Section 3.2 to derive asymptotic properties of optimal solutions when estimating μ in $P(\mu)$ by empirical measures. We obtain a law of iterated logarithm, a large-deviation estimate and an estimate for the asymptotic distribution of the optimal-solution sets without imposing that $\psi(\mu)$ must be a singleton. The basic tools are known limit theorems for the Kolmogorov-Smirnov distance of the empirical distribution function. Let $\xi_1., \xi_2, \ldots, \xi_n, \ldots$ be independent $I\!R^s$-valued random variables on a probability space $(\Omega, \mathfrak{A}, \mathbf{P})$ having joint distribution μ. We consider the empirical measures

$$\mu_n(\omega) := \frac{1}{n}\sum_{i=1}^{n}\delta_{\xi_i(\omega)} \qquad (\omega \in \Omega;\ n \in I\!N)$$

and we are interested in the asymptotic behaviour of the solution set $\psi(\mu_n(\cdot))$ of $P(\mu_n(\cdot))$ as n tends to infinity. Our results are put in terms of the Hausdorff distance $d_H(\psi(\mu), \psi(\mu_n(\cdot)))$, which is an $\mathfrak{A}$-measurable mapping due to Theorem 2K. in [90].

Proposition 3.3.1 *Under the assumptions of Theorem* 3.2.4 *it holds*

$$\limsup_{n\to\infty}\Big(\frac{2n}{\log\log n}\Big)^{\frac{1}{2}} \cdot d_H(\psi(\mu), \psi(\mu_n(\omega))) \leq L\ell \quad \mathbf{P} - \textit{almost surely},$$

where L and ℓ denote the Lipschitz modulus and the number of basis matrices, respectively, arising in Corollary 3.2.5.

Proof: Let B_j, $j = 1, \ldots, \ell$, denote the relevant basis submatrices of W.
Then $\mu_n(\omega) \circ (-B_j)$ coincides with the empirical measure of $\mu \circ (-B_j)$ and the following law of iterated logarithm holds ([82], p. 302, [110])

$$\limsup_{n\to\infty}\Big(\frac{n}{2\log\log n}\Big)^{\frac{1}{2}} \sup_{t\in I\!R^s} |F_{\mu\circ(-B_j)}(t) - F_{\mu_n(\omega)\circ(-B_j)}(t)| \leq \frac{1}{2}$$

$\mathbf{P}$-almost surely for all $j = 1, \ldots, \ell$.
Hence, the estimate from Corollary 3.2.5 is valid for $\mathbf{P}$-almost all $\omega \in \Omega$ with $\nu := \mu_n(\omega)$, provided that $n = n(\omega) \in I\!N$ is sufficiently large.
Thus we have for $\mathbf{P}$-almost all $\omega \in \Omega$

$$\limsup_{n\to\infty}\Big(\frac{n}{2\log\log n}\Big)^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) \leq \frac{L\ell}{2}$$

■

In [29], [54] the authors obtain consistency results under weak hypotheses on the optimization problems involved (based on the theory of epi-convergence). The above proposition supplements these results by giving, under stronger assumptions, the (optimal) rate of convergence for the solution sets. Compared to considerations in [116] we can dispense with a linear-independence assumptions imposed there. This became possible, since we used simplical cones instead of more general ones (cf. Remark 3.2.6). Compared to [29], [134] we do not need the unique solvability of $P(\mu)$.

Proposition 3.3.2 *Under the assumptions of Theorem* 3.2.4 *there exists a constant* $\varepsilon_o > 0$ *such that it holds for all* $\varepsilon \in (0, \varepsilon_o]$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}(\{\omega : d_H(\psi(\mu), \psi(\mu_n(\omega))) \geq \varepsilon\}) \leq -2\Big(\frac{\varepsilon}{L\ell}\Big)^2.$$

Proof: For brevity, we introduce the following notation

$$\begin{aligned} \Upsilon_n(\omega) &:= \max_{j=1,\ldots,\ell} \eta_{n,j}(\omega), \\ \eta_{n,j}(\omega) &:= \sup_{t\in I\!\!R^s} |F_{\mu\circ(-B_j)}(t) - F_{\mu_n(\omega)\circ(-B_j)}(t)| \quad (\omega \in \Omega). \end{aligned}$$

Now select $\varepsilon_o > 0$ in such a way that $L\ell\Upsilon_n(\omega) < \varepsilon_o$ and Corollary 3.2.5 imply

$$d_H(\psi(\mu), \psi(\mu_n(\omega))) \leq L\ell\Upsilon_n(\omega).$$

Then we have for each $\varepsilon \in (0, \varepsilon_o]$ and all $n \in I\!\!N$

$$\begin{aligned} \{\omega : d_H(\psi(\mu), \psi(\mu_n(\omega))) \geq \varepsilon\} &\subseteq \Big\{\omega : \Upsilon_n(\omega) \geq \frac{\varepsilon_o}{L\ell}\Big\} \cup \Big\{\omega : \Upsilon_n(\omega) \geq \frac{\varepsilon}{L\ell}\Big\} \\ &= \Big\{\omega : \Upsilon_n(\omega) \geq \frac{\varepsilon}{L\ell}\Big\} = \bigcup_{j=1}^{\ell} \Big\{\omega : \eta_{n,j}(\omega) \geq \frac{\varepsilon}{L\ell}\Big\} \end{aligned}$$

and, hence,

$$\mathbf{P}(\{\omega : d_H(\psi(\mu), \psi(\mu_n(\omega))) \geq \varepsilon\}) \leq \sum_{j=1}^{\ell} \mathbf{P}\Big(\Big\{\omega : \eta_{n,j}(\omega) \geq \frac{\varepsilon}{L\ell}\Big\}\Big).$$

The multivariate version of the Dvoretzky-Kiefer-Wolfowitz inequality (cf. [51], [110]) then implies that, for each $\delta > 0$ and $j \in \{1, \ldots, \ell\}$, there exist constants $C_j > 0$ such that

$$\mathbf{P}\Big(\Big\{\omega : \eta_{n,j}(\omega) \geq \frac{\varepsilon}{L\ell}\Big\}\Big) \leq C_j \exp\Big(-(2-\delta)n\Big(\frac{\varepsilon}{L\ell}\Big)^2\Big) \quad \text{for all } \; n \in I\!\!N.$$

Hence, we obtain for any $n \in I\!\!N$ and $\delta > 0$,

$$\frac{1}{n} \log \mathbf{P}(\{\omega : d_H(\psi(\mu); \psi(\mu_n(\omega)) \geq \varepsilon\}) \leq \frac{1}{n} \log\Big(\sum_{j=1}^{\ell} C_j\Big) - (2-\delta)\Big(\frac{\varepsilon}{L\ell}\Big)^2$$

and the proof is complete. ∎

Compared to Theorem 4.6 in [50] which represents a large deviation result for more general stochastic programs, the above proposition only requires the weak moment condition (A3) and yields an explicit estimate instead of an implicit one involving a

conditioning function that is often hard to quantify. We also refer to the exponential bound in Theorem 2 [124] which, in the context of two-stage stochastic programming, is working for non-unique solutions but only applies to measures μ with bounded support.
Another substantial step in the asymptotic analysis of optimal solutions would consist of obtaining asymptotic distributions of the sequence of closed random sets

$$(n^{1/2}(\psi(\mu_n(\cdot)) - x))_{n\in\mathbb{N}} \qquad (\text{for each } \ x \in \psi(\mu))$$

on the hyperspace of closed subsets of $\mathbb{R}^m$. In [53], [112] this problem was tackled for stochastic programs involving expectation functions with smooth integrands. Moreover, it was assumed that the unperturbed problem has a unique optimal solution. For stochastic programs with complete recourse the relevant integrands are typically non-smooth (cf. (1.2), (1.3)) and uniqueness of optimal solutions is rather exceptional (cf. Example 3.1.1) such that the results from [53], [112] do not apply. From Theorem 3.2.4, however, a lower estimate for the asymptotic distribution of

$$(n^{1/2}d_H(\psi(\mu), \psi(\mu_n(\cdot))))_{n\in\mathbb{N}}$$

can be derived. This is done next. The result is inspired by the concept of normalized convergence and the corresponding techniques in [31]. For simple-recourse models the lower estimate becomes more detailed (Remark 3.3.4).

Proposition 3.3.3 *Under the assumptions of Theorem* 3.2.4 *there exist probability distribution functions G_j, $j = 1, \dots, \ell$, on $\mathbb{R}$ such that it holds*

$$\liminf_{n\to\infty} \mathbf{P}(\{\omega : n^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) < t\}) \geq 1 + \sum_{j=1}^{\ell} \Big(G_j\Big(\frac{t}{L\ell}\Big) - 1\Big)$$

for all $t \geq 0$, where L and ℓ denote the Lipschitz modulus and the number of basis matrices, respectively, arising in Corollary 3.2.5.

Proof: Let $\eta_{n,j}(\omega)$ be given as in the proof of Proposition 3.3.2.
From the asymptotic distribution theory for the Kolmogorov-Smirnov distance it is known that for each $j = 1, \dots, \ell$ the sequence

$$(n^{\frac{1}{2}}\eta_{n,j}(\cdot))_{n\in\mathbb{N}}$$

converges in distribution to some real random variable ζ_j (Theorem 2 in [52]; chapt. 2.1.5 in [110]).

Let $t \geq 0$, $n \in I\!N$ and consider the following events in $\mathfrak{A}$:

$$\begin{aligned} A &:= \Big\{\omega : Ln^{\frac{1}{2}} \sum_{j=1}^{\ell} \eta_{n,j}(\omega) < t\Big\}, \\ A_j &:= \Big\{\omega : n^{\frac{1}{2}} \eta_{n,j}(\omega) < \frac{t}{L\ell}\Big\} \qquad (j = 1, \ldots, \ell) \\ B_\delta &:= \Big\{\omega : L \sum_{j=1}^{\ell} \eta_{n,j}(\omega) < \delta\Big\}, \end{aligned}$$

where $\delta > 0$ is selected according to Corollary 3.2.5 such that, for all $\omega \in B_\delta$, $d_H(\psi(\mu), \psi(\mu_n(\omega)))$ can be estimated by the expression defining B_δ.
Corollary 3.2.5 then yields the following chain of inequalities

$$\begin{aligned} &\mathbf{P}(\{\omega : n^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) < t\}) \\ &\quad \geq \mathbf{P}(\{\omega : n^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) < t\} \cap B_\delta) \\ &\quad \geq \mathbf{P}(A \cap B_\delta) \geq \mathbf{P}\Big(\bigcap_{j=1}^{\ell} A_j \cap B_\delta\Big) \\ &\quad = \mathbf{P}\Big(\bigcap_{j=1}^{\ell} A_j\Big) - \mathbf{P}\Big(\bigcap_{j=1}^{\ell} A_j \cap \bar{B}_\delta\Big) \\ &\quad \geq \mathbf{P}\Big(\bigcap_{j=1}^{\ell} A_j\Big) - \mathbf{P}(\bar{B}_\delta) \\ &\quad = 1 - \mathbf{P}\Big(\bigcup_{j=1}^{\ell} \bar{A}_j\Big) - \mathbf{P}(\bar{B}_\delta) \\ &\quad \geq 1 - \sum_{j=1}^{\ell} \mathbf{P}(\bar{A}_j) - \mathbf{P}(\bar{B}_\delta) = 1 + \sum_{j=1}^{\ell} (\mathbf{P}(A_j) - 1) - \mathbf{P}(\bar{B}_\delta) \end{aligned}$$

Hence we obtain the following estimate for all $t \geq 0$ and $n \in I\!N$:

$$\begin{aligned} &\mathbf{P}(\{\omega : n^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) < t\}) \\ &\quad \geq 1 + \sum_{j=1}^{\ell} \Big(\mathbf{P}\Big(\Big\{\omega : n^{\frac{1}{2}} \eta_{n,j}(\omega) < \frac{t}{L\ell}\Big\}\Big) - 1\Big) - \mathbf{P}\Big(\Big\{\omega : L \sum_{j=1}^{\ell} \eta_{n,j}(\omega) \geq \delta\Big\}\Big) \end{aligned}$$

The latter probability tends to zero as $n \to \infty$ because of the Glivenko-Cantelli theorem and we finally obtain via the Portmanteau theorem for each $t \geq 0$:

$$\begin{aligned} &\liminf_{n\to\infty} \mathbf{P}(\{\omega : n^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) < t\}) \\ &\quad \geq 1 + \sum_{j=1}^{\ell} \Big(\liminf_{n\to\infty} \mathbf{P}\Big(\Big\{\omega : n^{\frac{1}{2}} \eta_{n,j}(\omega) < \frac{t}{L\ell}\Big\} - 1\Big)\Big) \end{aligned}$$

$$\geq 1 + \sum_{j=1}^{\ell} \Big(\mathbf{P}\Big(\Big\{\omega : \zeta_j(\omega) < \frac{t}{L\ell}\Big\}\Big) - 1\Big) = 1 + \sum_{j=1}^{\ell} \Big(G_j\Big(\frac{t}{L\ell}\Big) - 1\Big)$$

where $G_j(u) := \mathbf{P}(\{\omega : \zeta_j(\omega) < u\})$ for all $u \in I\!R$. ■

Remark 3.3.4 *Unfortunately, the limit distributions G_j $(j = 1, \ldots, \ell)$ cannot be characterized in general for multidimensional distributions $F_{\mu\circ(-B_j)}$ (see [52]). However, in case of simple recourse we obtain the following estimate by using Corollary 3.2.15 instead of 3.2.5 and under the assumption that all one-dimensional marginal distributions of μ are continuous:*

$$\liminf_{n\to\infty} \mathbf{P}(\{\omega : n^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) < t\}) \geq 1 + s\Big(H\Big(\frac{t}{Ls}\Big) - 1\Big)$$

for all $t \geq 0$, where $H(u) := 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1} e^{-2k^2u^2}$ $(u \geq 0)$ is the asymptotic distribution in the Kolmogorov limit theorem.

Appendix

In this appendix we are concerned with the set-valued mapping

$$\psi(d) := \operatorname{argmin}\{x^T Hx + c^T x : Dx \leq d\}$$

where H is a symmetric positive semidefinite (m, m)-matrix, $c \in I\!R^m$, D is some $(\bar{m}, m)$-matrix and $d \in I\!R^{\bar{m}}$. For the comfort of the reader we display how the Lipschitz continuity of ψ on its effective domain $\operatorname{dom}\psi := \{d \in I\!R^{\bar{m}} : \psi(d) \neq \emptyset\}$ is derived in [55].

Proposition A.1 *There exists a constant $L > 0$ such that for all $d_1, d_2 \in \operatorname{dom}\psi$*

$$d_H(\psi(d_1), \psi(d_2)) \leq L\|d_1 - d_2\|.$$

Proof: Let $d \in \operatorname{dom}\psi$ and $\bar{x} \in \psi(d)$. According to Theorem 3, §12.2., in [16] we have the representation

$$\psi(d) = M(d, \bar{x}) := \{x \in I\!R^m : Dx \leq d,\ Hx = H\bar{x},\ c^T x = c^T\bar{x}\}.$$

By Hoffman's theorem on perturbed linear inequalities ([44],[125]) there exists a constant $L_o > 0$ such that for all $(d_1, \bar{x}_1), (d_2, \bar{x}_2) \in \operatorname{dom} M$

$$d_H(M(d_1, \bar{x}_1), M(d_2, \bar{x}_2)) \leq L_o(\|d_1 - d_2\| + \|\bar{x}_1 - \bar{x}_2\|). \tag{A.1}$$

Consider some fixed $d_o \in \operatorname{dom} \psi$. Then there exist a constant $L_1 > 0$ (not depending on d_o) and a constant $\delta_1 = \delta_1(d_o) > 0$ such that for all $d \in \operatorname{dom} \psi$ with $\|d - d_o\| \leq \delta_1$ it holds

$$\text{(A.2)} \qquad \psi(d) \subset \psi(d_o) + L_1 \|d - d_o\| B.$$

([56], [85]).
Now for each $d \in \operatorname{dom} \psi$ with $\|d - d_o\| \leq \delta_1$ and each $x(d) \in \psi(d)$ there exists an $x_o = x_o(x(d)) \in \psi(d_o)$ such that

$$\|x_o - x(d)\| = \operatorname{dist}(x(d), \psi(d_o)).$$

In particular, it holds

$$\psi(d) = M(d, x(d)) \quad \text{and} \quad \psi(d_o) = M(d_o, x_o).$$

Hence, by (A.1) and (A.2),

$$\begin{aligned} d_H(\psi(d), \psi(d_o)) &= d_H(M(d, x(d)), M(d_o, x_o)) \\ &\leq L_o(\|d - d_o\| + \|x(d) - x_o\|) \\ &\leq L_o(1 + L_1)\|d - d_o\|. \end{aligned}$$

Thus, for any $d_o \in \operatorname{dom} \psi$, there exist a constant $L > 0$ (not depending on d_o) and a constant $\delta = \delta(d_o) > 0$ such that for all $d \in \operatorname{dom} \psi$, $\|d - d_o\| \leq \delta$

$$\text{(A.3)} \qquad d_H(\psi(d), \psi(d_o)) \leq L\|d - d_o\|.$$

Using a covering argument the proof is completed:
Provided that $\operatorname{dom} \psi \neq \emptyset$ it holds $\operatorname{dom} \psi = \operatorname{dom} M_o$ where $M_o(d) = \{x \in \mathbb{R}^m : Dx \leq d\}$ ([41]). Therefore, $\operatorname{dom} \psi$ is convex.
Now let $d_1, d_2 \in \operatorname{dom} \psi$ and consider the line segment $[d_1, d_2] \subset \operatorname{dom} \psi$. By compactness, there exist a constant $\delta_o > 0$ and finitely many points $d(0) = d_1, \ldots, d(N) = d_2 \in [d_1, d_2]$ such that

- $d(i) \neq d(j)$ for all $i \neq j$, $d(i+1) \in [d(i), d(i+2)]$ $(i = 0, \ldots, N-2)$,
- the union of the (open) δ_o-neighbourhoods $\mathcal{N}_{\delta_o}(d(i))$ covers $[d_1, d_2]$,
- for each $d(i)$ $(i = 0, \ldots, N)$ the estimate (A.3) holds with $d_o := d(i)$, $\delta := \delta_o$ and L as above.

For $i = 0, \dots, N-1$ it holds

$$\mathcal{N}_{\delta_o}(d(i)) \cap \mathcal{N}_{\delta_o}(d(i+1)) \cap [d(i), d(i+1)] \neq \emptyset.$$

Select points $\tilde{d}(i)$ $(i = 0, \dots, N-1)$ from these intersections. Then it holds

$$\begin{aligned} d_H(\psi(d_1), \psi(d_2)) &\leq d_H(\psi(d(0)), \psi(\tilde{d}(0))) + d_H(\psi(\tilde{d}(0)), \psi(d(1))) + \ldots + \\ &\quad + d_H(\psi(\tilde{d}(N-1)), \psi(d(N))) \\ &\leq L(\|d(0) - \tilde{d}(0)\| + \|\tilde{d}(0) - d(1)\| + \ldots + \|\tilde{d}(N-1) - d(N)\|) \\ &= L\|d_1 - d_2\|, \end{aligned}$$

and the proof is complete. ■

Acknowledgement
This chapter has benefitted from several discussions with Alexander Shapiro (Georgia Institute of Technology, Atlanta) and Diethard Klatte (University of Zurich). Suggestions by Alexander Shapiro led to a more transparent proof of Theorem 3.2.4.

Chapter 4

Structure and Stability in Stochastic Programs with Complete Integer Recourse

Abstract
For two-stage stochastic programs with integrality constraints in the second stage we study continuity properties of the expected recourse as a function both of the first-stage policy and the integrating probability measure.
Sufficient conditions for lower semicontinuity, continuity and Lipschitz continuity with respect to the first-stage policy are presented. Furthermore, joint continuity in the policy and the probability measure is established. This leads to conclusions on the stability of optimal values and optimal solutions to the two-stage stochastic program when subjecting the underlying probability measure to perturbations.

4.1 Introduction

Consider a probability space $(\Omega, \mathfrak{A}, P)$ and measurable mappings $z : \Omega \to I\!R^s$, $A : \Omega \to I\!R^{ms}$ where the images of A are understood as $s \times m$ matrices. A two-stage stochastic integer program with random technology matrix is then given by

$$\min\{g(x) + Q(x) : \; x \in C\}$$

where

$$Q(x) = \int\limits_{\Omega} \Phi(z(\omega) - A(\omega)x) P(d\omega)$$

and

$$\text{(1.1)}\quad \Phi(t) = \min\{q^T y + q'^T y' : Wy + W'y' = t,\ y' \geq 0,\ y \geq 0,\ y' \in I\!R^{m'},\ y \in Z\!\!Z^{\bar{m}}\}.$$

Basically, we assume that $g : I\!R^m \to I\!R$ is continuous, $C \subset I\!R^m$ nonempty, closed, $q \in I\!R^{\bar{m}}$, $q' \in I\!R^{m'}$ and that the matrices $W \in L(I\!R^{\bar{m}}, I\!R^s)$, $W' \in L(I\!R^{m'}, I\!R^s)$ have rational entries. By $Z\!\!Z^{\bar{m}}$ we denote the subset of integer vectors in $I\!R^{\bar{m}}$. Further assumptions ensuring that the above expressions are well-defined are presented below. The measurable mappings z and A induce a probability measure $\mu := P \circ (z, A)^{-1}$ on $I\!R^S$ where S=(m+1)s. Our model then reads

$$P(\mu) \qquad\qquad \min\{g(x) + Q(x, \mu) : x \in C\}$$

where

$$\text{(1.2)}\qquad Q(x, \mu) = \int\limits_{I\!R^S} \Phi(z - Ax)\mu(d(z, A)).$$

The stochastic program $P(\mu)$ is an appropriate model for an optimization process where, in a first stage, a decision x must be taken under uncertainty on the problem data z and A, and, in a second stage, a decision (y, y') is made after realization of (z, A). The second stage is formalized via the optimization problem behind Φ (see [46]). The latter, for instance, may model an optimal compensation of the surplus (shortfall) $z - Ax$ or may deliver detailed scheduling decisions (based on $z - Ax$) in a hierarchical system. The integral Q given by (1.2) models the expected additional costs due to the second-stage action (y, y'). The peculiarities of $P(\mu)$ are two-fold: in the first stage we allow for randomness not only in the vector z but also in the technology matrix A and in the second (or recourse) stage we restrict some decisions y to be integral.

Restricting decisions in the second stage of a stochastic program with recourse to be integers is interesting from theoretical viewpoint since the implications of the "smoothing effect" of the integral in (1.2) are not obvious. Moreover, in applications a proper modelling often requires integer variables which, obviously, is especially relevant when the second-stage program is a combinatorial optimization problem ([65], [83], [118]).

For these reasons we keep the second stage a general linear mixed-integer program. We study the "smoothing effect" of the integral in (1.2), i.e. we derive continuity properties of Q both with respect to x and μ. The joint continuity of Q in x and μ leads to conclusions on the stability of $P(\mu)$ when μ varies in a certain set of probability measures. Like in other branches of stochastic programming stability considerations of the model under perturbations of the integrating probability measure are important prerequisites when justifying approximation schemes ([10], [49]) or when replacing incompletely known distributions by suitable estimates ([27]).

To obtain a first impression on the difficulties that appear when implanting integrality constraints into the second stage of $P(\mu)$ consider the simple example

$$
\begin{aligned}
Q(x,\mu) &= \int_{I\!R} \Phi(z-x)\mu(dz), \\
\Phi(t) &= \min\{y : y \geq t,\ y \in Z\!\!\!Z\},
\end{aligned} \tag{1.3}
$$

where μ is the uniform distribution on the interval $[0, 1/4]$.
Of course,

$$
Q(x,\mu) = \int_{I\!R} \lceil z - x \rceil \mu(dz)
$$

where $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. Simple calculations show that $Q(.,\mu)$ is neither convex nor differentiable. Moreover, for a discrete distribution μ the function $Q(.,\mu)$ becomes discontinuous. So, in contrast to recourse problems without integer variables, neither convexity nor differentiability of $Q(.,\mu)$ can be expected to hold under reasonably comprehensible assumptions.
Compared to the rich literature on structure and stability for stochastic programs without integer requirements (as a collection take, for instance, [46], [131] and [27], [48], [88], [93], [116]) there are only a few contributions to the problems addressed in the present chapter. The first one seems to be due to Stougie [118] who established that $Q(.,\mu)$ is continuous provided that μ has a uniformly continuous density and that Φ fulfils some boundedness requirement (cf. also [83]).
The more recent papers [60], [61], [66] focus on simple integer recourse where Φ specifies to

$$
\Phi(t) = \min\{{q^+}^T y^+ + {q^-}^T y^- : y^+ \geq t,\ y^- \geq -t,\ y^+ \geq 0,\ y^- \geq 0,\ y^+ \in Z\!\!\!Z^s,\ y^- \in Z\!\!\!Z^s\}.
$$

Exploiting the inherent separability the authors derive explicit formulae, (sharp) convex lower bounds and sufficient conditions for continuity, convexity and differentiability of $Q(.,\mu)$. Whereas [61], [66] treat the case of random right-hand side, problems containing also a random technology matrix are considered in [60].
Our results on the joint continuity of Q in x and μ are related to continuity results for expectations with discontinuous integrands obtained in [1] and [64].
The chapter is organized as follows: In Section 4.2 we collect a few prerequisites from probability theory and basic results on the value function of a mixed-integer

linear program with parameters in the right-hand side of the constraints ([5], [12]). In Section 4.3 we present sufficient conditions for lower semicontinuity, continuity and Lipschitz continuity of $Q(.,\mu)$ as well as for the joint continuity of $Q(.,.)$. In Section 4.4, we derive continuity of the optimal-value function and Berge upper semicontinuity of the solution set mapping when understanding $P(\mu)$ as a parametric program with respect to μ.

4.2 Prerequisites

Recall that the integrand Φ in (1.2) is the value-function of a linear mixed-integer program with parameters in the right-hand side of the equality constraints. Properties of such value functions are derived, for instance, in the monograph [5] and in the article [12] from where we quote the propositions below. Basically, we assume that for each $t \in I\!R^s$ the constraint set of the program defining $\Phi(t)$ is non-empty and that $\Phi(0) = 0$. Then $\Phi(t) \in I\!R$ for all $t \in I\!R^s$ (cf. e.g. [71], Proposition I.6.7.) and the following proximity result holds:

Proposition 4.2.1 *([5], Theorem 8.1; [12], Theorem 2.1)*
There exist constants $\alpha > 0$, $\beta > 0$ such that for all $t', t'' \in I\!R^s$ we have

$$|\Phi(t') - \Phi(t'')| \leq \alpha \|t' - t''\| + \beta.$$

Moreover, the value function Φ admits the following representation:

Proposition 4.2.2 *([12], Theorem 3.3)*
There exist constants $\gamma > 0$, $\delta > 0$ and vectors $d_1, \ldots, d_\ell \in I\!R^s$, $\tilde{d}_1, \ldots, \tilde{d}_{\tilde{\ell}} \in I\!R^s$ such that for all $t \in I\!R^s$

$$\Phi(t) = \min_y \{q^T y + \max_{j \in \{1,\ldots,\ell\}} d_j^T (t - Wy) : \ y \in Y(t)\}$$

where

$$\begin{aligned} Y(t) = \{y \in Z\!\!\!Z^{\bar{m}} : \quad & y \geq 0, \ \sum |y_i| \leq \gamma \sum |t_r| + \delta, \\ & \tilde{d}_k^T (t - Wy) \geq 0, \ k = 1, \ldots, \tilde{\ell}\}. \end{aligned}$$

Straightforward duality considerations and a proximity argument for optimal solutions led to the above result. In fact, the vectors $d_1, \ldots, d_\ell$ come up as the vertices of the polyhedron $\{u \in I\!R^s : {W'}^T u \leq q'\}$, and the vectors $\tilde{d}_1, \ldots, \tilde{d}_{\tilde{\ell}}$ stem from an inequality description of the polyhedral cone $W'(I\!R_+^{m'})$.

Continuity properties of Φ can be derived from Proposition 4.2.2. Namely, if $Y(.)$ remains constant on an open neighbourhood of some point $\bar{t} \in \mathbb{R}^s$, then, on this neighbourhood, Φ is the pointwise minimum of finitely many continuous (piecewise linear) functions and, hence, continuous at $\bar{t}$. If $\tilde{t} \in \mathbb{R}^s$ is such that $Y(.)$ does not remain constant on any open neighbourhood of $\tilde{t}$, then there must exist $\tilde{y} \in \mathbb{Z}^{\bar{m}}$, $\tilde{y} \geq 0$ such that at least one of the inequalities

$$\sum |\tilde{y}_i| \leq \gamma \sum |\tilde{t}_r| + \delta$$

and

$$\tilde{d}_k^T(\tilde{t} - W\tilde{y}) \geq 0, \quad k = 1, \ldots, \tilde{\ell}$$

holds as an equation.
In fact, only the second group of inequalities is relevant, as we will explain now: Using only duality arguments we obtain

$$\Phi(t) = \min_y \{q^T y + \max_j d_j^T(t - Wy) : \ y \geq 0, \ y \in \mathbb{Z}^{\bar{m}}, \ \tilde{d}_k^T(t - Wy) \geq 0, k = 1, \ldots, \tilde{\ell}\}.$$

The merrit of Theorem 3.3 in [12] (Proposition 4.2.2) is to restrict the above minimization(over an infinite set) to the finite set $Y(t)$. If $\tilde{t} \in \mathbb{R}^s$ is such that, for some $\tilde{y} \in \mathbb{Z}^{\bar{m}}, \tilde{y} \geq 0$, the inequality

$$\sum |\tilde{y}_i| \leq \gamma \sum |\tilde{t}_r| + \delta$$

holds as an equation, then $Y(t)$ changes on any neighbourhood of $\tilde{t}$. But this has no impact on the result of the minimization, if we assume that the constants γ, δ were selected in such a way (sufficiently large) that the minimum in Proposition 4.2.2 is attained for a $y \in \mathbb{Z}^{\bar{m}}$ such that

$$\sum |y_i| < \gamma \sum |t_r| + \delta.$$

Hence, the discontinuities of Φ are concentrated in points $t \in \mathbb{R}^s$ where, for some $y \in \mathbb{Z}_+^{\bar{m}}$, at least one of the inequalities

$$\tilde{d}_k^T(t - Wy) \geq 0 \, , \ k = 1, \ldots, \tilde{\ell}$$

holds as an equation.
The set of discontinuity points of Φ is thus contained in a countable union of hyperplanes in $\mathbb{R}^s$, more specifically, in a union of translates of hyperplanes determined by the facets of the cone $W'(\mathbb{R}_+^{m'})$.
By the rationality of W', the vectors $\tilde{d}_k$ ($k = 1, \ldots, \tilde{\ell}$) are rational, too. Since also

W is rational, this implies that there exists a constant $\varepsilon_0 > 0$ such that (for all $k = 1, \ldots, \tilde{\ell}$)

$$|\tilde{d}_k^T W y_1 - \tilde{d}_k^T W y_2| > \varepsilon_0 \qquad \text{whenever} \qquad y_1, y_2 \in \mathbb{Z}^{\bar{m}},\ \tilde{d}_k^T W y_1 \neq \tilde{d}_k^T W y_2.$$

Hence, for any $t \in \mathbb{R}^s$, there exists a neighbourhood $U(t)$ such that $Y(t') \subseteq Y(t)$ for any $t' \in U(t)$. This implies that $\liminf\limits_{t' \to t} \Phi(t') \geq \Phi(t)$, i.e. Φ is a lower semicontinuous function on $\mathbb{R}^s$ (cf. also [12], p. 133). An example (for the pure-integer case) showing how this lower semicontinuity is lost if the constraint matrix contains irrational entries can be constructed from the example given at page 58 in [4].

One key point in our analysis is that, in (1.2), not only the right-hand side z but also the technology matrix A can be stochastic. Therefore, the stochastic program $P(\mu)$ contains a joint probability distribution μ of z and A. Moreover, marginal and conditional distributions of μ will be important for our purposes. For convenience, we collect these notions here; further details can be found in textbooks on probability theory ([24], [38]).
Let $\pi_{\mathbb{R}^s}$ and $\pi_{\mathbb{R}^{ms}}$ denote the projections from $\mathbb{R}^S$ to $\mathbb{R}^s$ and $\mathbb{R}^{ms}$, respectively. The induced measures $\mu_1 = \mu \circ \pi_{\mathbb{R}^s}^{-1}$, $\mu_2 = \mu \circ \pi_{\mathbb{R}^{ms}}^{-1}$ are then referred to as the marginal distributions of μ with respect to z and A, respectively. By $\mu_1^2(A, .)$ we denote the (regular) conditional distribution of z given A. It has the following properties

(2.1) $\mu_1^2(A, .)$ is a probability measure on $\mathbb{R}^s$ for any $A \in \mathbb{R}^{ms}$;

(2.2) the function $\mu_1^2(., B_1) : \mathbb{R}^{ms} \to [0, 1]$ is measurable for any Borel set B_1 in $\mathbb{R}^s$;

(2.3) for any Borel set B in $\mathbb{R}^S$ it holds
$\mu(B) = \int\limits_{\mathbb{R}^{ms}} \int\limits_{R^s} \mathbf{1}_B(z, A) \mu_1^2(A, dz) \mu_2(dA)$
where $\mathbf{1}_B$ denotes the indicator function of B.

The above family of probability measures $\mu_1^2(A, .)$ indeed exists, since μ, as a probability measure on a Euclidean space, satisfies the general assumptions for the existence of a (regular) conditional distribution (cf. Theorem 10.2.2 in [24], Satz 5.3.21 in [38]).

4.3 Continuity of the Expected Recourse Function

We impose the following general assumptions to have (1.1), (1.2) well defined:

(A1) For all $t \in \mathbb{R}^s$ there exist $y \in \mathbb{Z}^{\bar{m}}$, $y' \in \mathbb{R}^{m'}$ such that $y \geq 0$, $y' \geq 0$ and $Wy + W'y' = t$.

(A2) There exists a $u \in \mathbb{R}^s$ such that $W^T u \leq q$, $W'^T u \leq q'$.

(A3) It holds that $\int\limits_{\mathbb{R}^S} (\|z\| + \|A\|)\mu(d(z, A)) < +\infty$.

In (A3), $\|z\|$ denotes the Euclidean norm of z and $\|A\|$ the induced matrix norm of A. In the context of (stochastic) linear programming (A2) is called "dual feasibility". Assumption (A1) is the natural extension of the complete-recourse assumption for stochastic programs with (non-integer) recourse and, therefore, referred to as "complete (mixed-) integer recourse". Assumption (A3), i.e. the finiteness of the first moment of μ, is basic for (non-integer) stochastic linear programs too (cf. [46], [131] and Chapter 2).

Proposition 4.3.1 *Assume (A1) – (A3), then $Q(., \mu)$ is a real-valued lower semicontinuous function on $\mathbb{R}^m$.*

Proof: Assumptions (A1), (A2) together with the duality theorem of linear programming and Lemma 7.1 in [5] imply that $\Phi(z - Ax) \in \mathbb{R}$ for all $z \in \mathbb{R}^s$, $A \in \mathbb{R}^{ms}$, $x \in \mathbb{R}^m$ (see also Proposition I.6.7 in [71]). Furthermore, Φ is measurable as a lower semicontinuous function on $\mathbb{R}^s$ (see Section 4.2). Assumption (A2) implies that $\Phi(0) = 0$, and we obtain in light of Proposition 4.2.1

$$\begin{aligned} |Q(x, \mu)| &\leq \int\limits_{\mathbb{R}^S} |\Phi(z - Ax) - \Phi(0)|\mu(d(z, A)) \\ &\leq \alpha \int\limits_{\mathbb{R}^S} \|z - Ax\|\mu(d(z, A)) + \beta \int\limits_{\mathbb{R}^{(m+1)}} \mu(d(z, A)) \\ &\leq \alpha \int\limits_{\mathbb{R}^S} \|z\|\mu(d(z, A)) + \alpha\|x\| \int\limits_{\mathbb{R}^{(m+1)s}} \|A\|\mu(d(z, A)) + \beta. \end{aligned}$$

Hence, $Q(., \mu)$ is a real-valued function on $\mathbb{R}^m$.
To verify the lower semicontinuity let $x \in \mathbb{R}^m$ and $\{x_n\}$ be a sequence in $\mathbb{R}^m$ converging to x. Denote $r := \max\limits_{n \in \mathbb{N}} \|x_n\| < +\infty$.
In view of Proposition 4.2.1 and $\Phi(0) = 0$ we have

$$\begin{aligned} \Phi(z - Ax_n) &\geq \Phi(0) - |\Phi(z - Ax_n) - \Phi(0)| \\ &\geq -\alpha\|z - Ax_n\| - \beta \\ &\geq -\alpha\|z\| - \alpha r\|A\| - \beta. \end{aligned}$$

Therefore, and by (A3), the function $h_o(z,A) := -\alpha\|z\| - \alpha r\|A\| - \beta$ is an integrable minorant of all the functions $h_n(z,A) := \Phi(z - Ax_n)$ $(n \in \mathbb{N})$.
Now we have

$$\begin{aligned} Q(x,\mu) &= \int \Phi(z-Ax)\mu(d(z,A)) \\ &\leq \int \liminf_{n\to\infty} \Phi(z-Ax_n)\mu(d(z,A)) \\ &\leq \liminf_{n\to\infty} \int \Phi(z-Ax_n)\mu(d(z,A)) \\ &= \liminf_{n\to\infty} Q(x_n,\mu). \end{aligned}$$

Here, the first estimate follows from the lower semicontinuity of Φ and the second is a consequence of Fatou's Lemma which works since we have the above minorant. Thus, $Q(.,\mu)$ is lower semicontinuous at x. ■

Let us remark that, by the above proposition, $P(\mu)$ is a "proper" model in the sense that one minimizes a lower semicontinuous function and, if the feasible set C is compact, for instance, the infimum of the objective is finite and actually attained.
To formulate a sufficient condition for the continuity of $Q(.,\mu)$ at some point $x \in \mathbb{R}^m$ we introduce the set $E(x)$ of all those $(z,A) \in \mathbb{R}^S$ such that Φ is discontinuous at $z - Ax$. $E(x)$ is measurable for all $x \in \mathbb{R}^m$ ([9], p. 225).

Proposition 4.3.2 *Assume (A1) – (A3) and let $x \in \mathbb{R}^m$ be such that $\mu(E(x)) = 0$, then $Q(.,\mu)$ is continuous at x.*

Proof: Let $\{x_n\}_{n=1}^{\infty}$ be a sequence converging to x. Denote $r := \max_{n\in\mathbb{N}} \|x_n\|$. Proposition 4.2.1 yields

$$\begin{aligned} |\Phi(z-Ax_n)| &= |\Phi(z-Ax_n) - \Phi(0)| \\ &\leq \alpha\|z-Ax_n\| + \beta \\ &\leq \alpha\|z\| + \alpha r\|A\| + \beta. \end{aligned}$$

In view of (A3), therefore, the function $h_o(z,A) := \alpha\|z\| + \alpha r\|A\| + \beta$ is an integrable majorant of all the functions $|h_n(z,A)|$ where $h_n(z,A) := \Phi(z - Ax_n)$.
Due to $\mu(E(x)) = 0$, it holds

$$h_n(z,A) \underset{n\to\infty}{\longrightarrow} h(z,A) := \Phi(z-Ax) \qquad \mu - \text{almost surely},$$

and Lebesgue's dominated convergence theorem works:

$$\begin{aligned}\lim_{n\to\infty} Q(x_n,\mu) &= \lim_{n\to\infty}\int\limits_{I\!R^S} \Phi(z-Ax_n)\mu(d(z,A))\\ &= \int\limits_{I\!R^S} \lim_{n\to\infty} \Phi(z-Ax_n)\mu(d(z,A))\\ &= \int\limits_{I\!R^S} \Phi(z-Ax)\mu(d(z,A))\\ &= Q(x,\mu).\end{aligned}$$

■

Corollary 4.3.3 *Assume (A1) – (A3) and let the conditional distribution $\mu_1^2(A,.)$ of z given A be absolutely continuous with respect to the Lebesgue measure on $I\!R^s$ for μ_2-almost all $A \in I\!R^{ms}$, ($\mu_2 := \mu \circ \pi_2^{-1}$); then $Q(.,\mu)$ is a continuous function on $I\!R^m$.*

Proof: As a consequence of Proposition 4.2.2 we obtained in Section 4.2 that the set of discontinuity points of Φ is contained in a countable union $\mathcal{H}$ of hyperplanes in $I\!R^s$. Therefore, for any $x \in I\!R^m$, $E(x) \subset E_1(x)$ where
$E_1(x) := \{(z,A) \in I\!R^S : z - Ax \in \mathcal{H}\}$. By (2.3),

$$\begin{aligned}\mu(E_1(x)) &= \int\limits_{I\!R^{ms}} \int\limits_{I\!R^s} \mathbf{1}_{E_1(x)}(z,A)\mu_1^2(A,dz)\mu_2(dA)\\ &= \int\limits_{I\!R^{ms}} \int\limits_{Ax+\mathcal{H}} \mu_1^2(A,dz)\mu_2(dA).\end{aligned}$$

Since $\mu_1^2(A,.)$ is absolutely continuous μ_2-almost surely, we now have
$\int\limits_{Ax+\mathcal{H}} \mu_1^2(A,dz) = 0$ for μ_2-almost all $A \in I\!R^{ms}$.
Hence, $\mu(E_1(x)) = 0$. This implies $\mu(E(x)) = 0$ for arbitrary $x \in I\!R^m$, and Proposition 4.3.2 yields the assertion. ■

Remark 4.3.4 *If z and A are independent random variables then $\mu_1^2(A,.)$ is absolutely continuous (for μ_2-almost all $A \in I\!R^{ms}$) if already the marginal distribution μ_1 has this property. Indeed, $\mu_1^2(A,.)$ then coincides μ_2-almost surely with μ_1. Another instance where Corollary 4.3.3 works is given when there is a joint density of z and the random components of A (i.e. those which are not constant μ-almost surely).*

If z and A are not independent, then it is not sufficient to claim that μ_1 is absolutely continuous when wishing to satisfy the assumptions of Corollary 4.3.3. Indeed, let $\mu \in \mathcal{P}(\mathbb{R}^2)$ be the uniform distribution concentrated on the line segment $\mathrm{conv}\{(0,0), (\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2})\}$. Then both μ_1 and μ_2 coincide with the uniform distribution on $\mathrm{conv}\{0, \frac{1}{2}\sqrt{2}\}$ and, for $0 \leq A \leq \frac{1}{2}\sqrt{2}$, $\mu_1^2(A,.)$ coincides with the measure concentrated at A.

However, if μ_1 is absolutely continuous and μ_2 is discrete (with countably many mass points) then $\mu_1^2(A,.)$ is absolutely continuous μ_2-almost surely. To see this, let $B_1 \subset \mathbb{R}^s$ have Lebesgue measure zero. Then

$$\begin{aligned} 0 &= \mu_1(B_1) = \mu(B_1 \times \mathbb{R}^{ms}) = \int\limits_{\mathbb{R}^{ms}} \int\limits_{R^s} \mathbf{1}_{B_1 \times \mathbb{R}^{ms}} \mu_1^2(A, dz) \mu_2(dA) \\ &= \sum_{j=1}^{\infty} \int\limits_{\mathbb{R}^s} \mathbf{1}_{B_1}(z) \mu_1^2(A_j, dz) \cdot p_j = \sum_{j=1}^{\infty} p_j \cdot \mu_1^2(A_j, B_1) \end{aligned}$$

with suitable mass points A_j and probabilities $p_j > 0$ $(j = 1, 2, \ldots)$. This implies $\mu_1^2(A_j, B_1) = 0$ for all j, and $\mu_1^2(A,.)$ is absolutely continuous μ_2-almost surely.

Remark 4.3.5 *Proposition 4.3.2 and Corollary 4.3.3 extend Theorem 5.1 in [118] where additional assumptions on Φ and μ are made. In the recent paper [60] a similar analysis is carried out for simple integer recourse. The authors also show how to verify the crucial assumption in Proposition 4.3.2 when having information about conditional distributions given certain components of the random technology matrix A.*

The Lipschitz continuity of $Q(., \mu)$ will first be investigated for the case where only the right-hand side z is random.

Recall from Section 2 that the discontinuity points of Φ are contained in

$$\bigcup_{k=1}^{\tilde{\ell}} \bigcup_{y \in \mathbb{Z}^{\bar{m}}} \{Wy + \mathcal{H}_k\}$$

where

$$\mathcal{H}_k = \{t \in \mathbb{R}^s : \tilde{d}_k^T t = 0\}.$$

By the rationality of W and W', the complement of the above union of hyperplanes admits a representation $\bigcup\limits_{i \in I} \mathfrak{P}_i$ where I is countable and $\mathrm{cl}\, \mathfrak{P}_i$ is a polyhedron for each $i \in I$. Proposition 4.2.2 yields that, on each of the sets $\mathfrak{P}_i$, the function Φ can be represented as the pointwise minimum of a family of Lipschitz continuous functions whose Lipschitz constants are bounded by $L_o := \max_{j \in \{1, \ldots, \ell\}} \|d_j\|$. Hence, on each of the sets $\mathfrak{P}_i$, the function Φ is Lipschitz continuous with constant L_o.

Proposition 4.3.6 *Assume (A1) – (A3) and that μ is absolutely continuous with respect to the Lebesgue measure on $I\!R^s$. Assume further that for any non-singular linear transformation $B \in L(I\!R^s, I\!R^s)$ the one-dimensional marginal distributions of $\mu \circ B$ have bounded densities which, outside some bounded interval, are monotonically decreasing with growing absolute value of the argument.*
Then Q is Lipschitz continuous on any bounded subset of $I\!R^m$.

Proof: Let x', x'' belong to some bounded subset D of $I\!R^m$. Define

$$\mathcal{S} = \mathcal{S}(x', x'') := \{z \in I\!R^s : \exists\, i \in I \quad z - Ax' \in \mathfrak{P}_i,\ z - Ax'' \in \mathfrak{P}_i\}.$$

Then

$$\begin{aligned}
&|Q(x', \mu) - Q(x'', \mu)| \\
&\quad \leq \int\limits_{I\!R^s} |\Phi(z - Ax') - \Phi(z - Ax'')| \mu(dz) \\
&\quad = \int\limits_{\mathcal{S}} |\Phi(z - Ax') - \Phi(z - Ax'')| \mu(dz) \\
&\qquad + \int\limits_{I\!R^s \setminus \mathcal{S}} |\Phi(z - Ax') - \Phi(z - Ax'')| \mu(dz) \\
&\quad \leq L_o \cdot \|A\| \cdot \|x' - x''\| \cdot \mu(\mathcal{S}) + \int\limits_{I\!R^s \setminus \mathcal{S}} (\alpha \|Ax' - Ax''\| + \beta)\, \mu(dz) \\
&\quad \leq (L_o + \alpha) \|A\| \|x' - x''\| \ +\ \beta \cdot \mu(I\!R^s \setminus \mathcal{S})
\end{aligned}$$

where we have used Proposition 4.2.1 and the Lipschitz property of Φ discussed above.
Consider the half spaces

$$\mathcal{H}_k^* = \{t \in I\!R^s : \tilde{d}_k^T t \geq 0\}, \quad k = 1, \ldots, \tilde{\ell}$$

and define

$$\mathcal{H}_k^{\triangle} = \bigcup_{y \in Z\!\!\!Z^{\bar{m}}} \{\{Ax' + Wy + \mathcal{H}_k^*\} \ \triangle\ \{Ax'' + Wy + \mathcal{H}_k^*\}\}$$

where $\triangle$ denotes the set-theoretic symmetric difference. Then

$$I\!R^s \setminus \mathcal{S} \ \subset\ \bigcup_{k=1}^{\tilde{\ell}} \mathcal{H}_k^{\triangle}. \tag{3.1}$$

Fix some $k \in \{1, \ldots, \tilde{\ell}\}$ and let $B_k \in L(I\!R^s, I\!R^s)$ be a non-singular linear transformation sending $\mathcal{H}_k^*$ to the half space $\{t \in I\!R^s : t_1 \geq 0\}$. Let $\tilde{\theta}_k = \theta_{\mu,k}^{(1)}$ be a

one-dimensional marginal density of the first component with respect to the image measure $\mu \circ B_k^{-1}$ fulfilling the boundedness and monotonicity assumptions listed in the proposition.
Without loss of generality, we assume that $[B_k Ax'']_1 > [B_k Ax']_1$. Denote $\tau'_{k,y} = [B_k(Ax' + Wy)]_1$ and $\tau''_{k,y} = [B_k(Ax'' + Wy)]_1$. Then it holds

$$\mu(\mathcal{H}_k^{\triangle}) = |\det B_k^{-1}| \sum_y \int_{\tau'_{k,y}}^{\tau''_{k,y}} \tilde{\theta}_k(\tau)\, d\tau. \tag{3.2}$$

Since D is bounded, there exists a finite subset $I_o \subset \mathbb{Z}^{\bar{m}}$ (independent on x', x'') such that the intervals $[\tau'_{k,y}, \tau''_{k,y}]$ meet the interval in the monotonicity assumption for $\tilde{\theta}_k$ for at most the elements y in I_o.
Without loss of generality, let us consider only those $y \notin I_o$ such that $\tau'_{k,y} > 0$. For the remaining $y \notin I_o$ a similar monotonicity argument applies. It holds

$$\sum_{y \notin I_o} \int_{\tau'_{k,y}}^{\tau''_{k,y}} \tilde{\theta}_k(\tau)\, d\tau \;\leq\; \|B_k\|\|A\|\|x' - x''\| \sum_{y \notin I_o} \tilde{\theta}_k(\tau'_{k,y}). \tag{3.3}$$

Let us show that $\sum_{y \notin I_o} \tilde{\theta}_k(\tau'_{k,y})$ is finite: Indeed, by the rationality of W and W', there are no accumulation points of $\{\tau'_{k,y}\}_{y \notin I_o}$. Hence, there exists an $\varepsilon > 0$ such that the monotonicity assumption on $\tilde{\theta}_k$ implies

$$1 \;\geq\; \sum_{y \notin I_o} \int_{\tau'_{k,y} - \varepsilon}^{\tau'_{k,y}} \tilde{\theta}_k(\tau)\, d\tau \;\geq\; \sum_{y \notin I_o} \int_{\tau'_{k,y} - \varepsilon}^{\tau'_{k,y}} \tilde{\theta}_k(\tau'_{k,y})\, d\tau \;=\; \varepsilon \sum_{y \notin I_o} \tilde{\theta}_k(\tau'_{k,y}).$$

For $y \in I_o$ the boundedness assumption on $\tilde{\theta}_k$ yields the existence of a bound $L_k > 0$ such that

$$\sum_{y \in I_o} \int_{\tau'_{k,y}}^{\tau''_{k,y}} \tilde{\theta}_k(\tau)\, d\tau \leq L_k \cdot \text{card } I_o \cdot \|B_k\| \cdot \|A\| \cdot \|x' - x''\|. \tag{3.4}$$

By (3.1) – (3.4) there exists a constant $L_* > 0$ such that

$$\mu(\mathbb{R}^s \setminus \mathcal{S}) \;\leq\; L_* \|A\| \|x' - x''\| \quad \text{for all } x', x'' \in D.$$

Together with the estimate from the beginning this completes the proof. ∎

The following examples show that the above proposition is no longer valid when omitting either the boundedness or the monotonicity assumptions on the marginal densities. In the examples the second-stage program is always as in (1.3). Hence, $\Phi(t) = \lceil t \rceil$.

Example 4.3.7 *Let Φ and A be given as in (1.3) and $\mu \in \mathcal{P}(I\!R)$ with the density*

$$\theta(\tau) = \tau^{-1/2} \quad \textit{for} \quad 0 < \tau \leq 1/4,$$

which is unbounded.
We obtain

$$Q(x,\mu) = \begin{cases} 1 & \textit{if } -\frac{3}{4} \leq x \leq 0 \\ 1 - 2\sqrt{x} & \textit{if } 0 \leq x \leq \frac{1}{4}, \end{cases}$$

which is not Lipschitz continuous on neighbourhoods of $x_o = 0$.

Example 4.3.8 *Let Φ and A be given as in (1.3) and $\mu \in \mathcal{P}(I\!R)$ with the density*

$$\theta(\tau) = \begin{cases} 1/n & \textit{for } \tau \in \left[n, n + \frac{1}{n^2} \cdot c\right], \; n = 1, 2, ... \\ 0 & \textit{else}, \end{cases}$$

where $c := \Big(\sum\limits_{n=1}^{\infty} \frac{1}{n^3} \Big)^{-1}$.
Obviously, the monotonicity assumption in Proposition 4.3.6 is not fulfilled.
We show that $Q(.,\mu)$ is not Lipschitz continuous on neighbourhoods of $x_o = 0$. First, observe that (A3) holds in view of

$$\int\limits_1^{\infty} \tau\theta(\tau) d\tau \leq \sum_{n=1}^{\infty} (n+1) \cdot \frac{1}{n} \cdot \frac{1}{n^2} \cdot c = c\frac{\pi^2}{6} + 1.$$

For arbitrary $x \in I\!R$, $0 < x < 1$ we have

$$Q(0,\mu) - Q(x,\mu) = \sum_{n=1}^{\infty} \int\limits_n^{n+x} \theta(\tau) d\tau \geq \sum_{n=1}^{\bar{n}(x)} \frac{1}{n} \cdot x,$$

where $\bar{n}(x) = \left\lfloor \sqrt{\frac{c}{x}} \right\rfloor$.
Consider $x_k = \frac{1}{k^2} \cdot c$ $(k = 1, 2, \ldots)$. The above yields

$$\frac{1}{x_k}(Q(0,\mu) - Q(x_k,\mu)) \geq \sum_{n=1}^{k} \frac{1}{n}.$$

Hence, for $k \to \infty$, the left-hand side tends to infinity, showing that $Q(.,\mu)$ is not Lipschitz continuous on neighbourhoods of $x_o = 0$.

The monotonicity assumption in Proposition 4.3.6 is, of course, fulfilled for a measure μ whose support (i.e. the smallest closed subset in $I\!R^s$ with μ-measure one) is bounded. The proof of Proposition 4.3.6 has shown that, instead of for arbitrary non-singular transformations, the assumptions have to hold only for specific transformations related to $W'(I\!R_+^{m'})$. Let us remark that there are counterexamples (cf. e.g. [95]) showing that boundedness of marginal densities is not implied by boundedness of the density of the original measure and that boundedness of marginal densities is not preserved under linear transformations of the original measure in general.

To present a class of probability measures which Proposition 4.3.6 applies to, we introduce the notation $M_r^\lambda(a,b)$, for $r \in I\!R \setminus \{0\}$, $\lambda \in [0,1]$, $a,b \geq 0$, cf. [17]:

$$M_r^\lambda(a,b) := \begin{cases} (\lambda a^r + (1-\lambda)b^r)^{1/r} & \text{if } a \cdot b > 0 \\ 0 & \text{if } a \cdot b = 0. \end{cases}$$

By passing to the limit this can be extended to $r = 0$, $r = -\infty$:

$$\begin{aligned} M_o^\lambda(a,b) &= a^\lambda b^{1-\lambda} \quad (\text{if } a \cdot b > 0) \quad \text{and} \\ M_{-\infty}^\lambda(a,b) &= \min\{a,b\}. \end{aligned}$$

A Borel probability measure $\mu \in \mathcal{P}(I\!R^s)$ is called r-convex, $r \in [-\infty, +\infty)$ (cf. [13], [17], [78]) if, for each $\lambda \in [0,1]$,

$$\mu(\lambda C_1 + (1-\lambda)C_2) \geq M_r^\lambda(\mu(C_1), \mu(C_2)) \tag{3.5}$$

holds for all Borel sets $C_1, C_2 \subset I\!R^s$ such that the Minkowski sum $\lambda C_1 + (1-\lambda)C_2$ is Borel. For $r = 0$ and $r = -\infty$, μ is also called logarithmic-concave and quasi-concave, respectively. Since $M_r^\lambda(a,b)$ is increasing in r, with the remaining variables fixed, the sets $\mathcal{M}_r$ of all r-convex measures are decreasing if r is increasing. In [13], Theorem 3.2, and [84], Theorem 1, it is shown that $\mu \in \mathcal{M}_r$ ($r \in (-\infty, 0]$) if and only if μ has a density θ such that

$$\theta(\lambda\tau_1 + (1-\lambda)\tau_2) \geq M_{r/(1-rs)}^\lambda(\theta(\tau_1), \theta(\tau_2))$$

for all $\lambda \in [0,1]$, $\tau_1, \tau_2 \in I\!R^s$.
From the literature, a number of multivariate probability distributions are known to be r-convex for some $r \in (-\infty, 0]$, e.g. the (non-degenerate) multivariate normal distribution and the t-distribution (cf. [13], p.113).

Proposition 4.3.9 *Assume that $\mu \in \mathcal{M}_r$ for some $r \in (-\infty, 0]$ and that the support of μ is the whole of $I\!R^s$. Then the hypotheses of Proposition 4.3.6 are satisfied.*

Proof: First observe that, for each non-singular $B \in L(\mathbb{R}^s, \mathbb{R}^s)$, we have $\mu \circ B \in \mathcal{M}_r$ and $\operatorname{supp}(\mu \circ B) = \mathbb{R}^s$. Hence, we are done when having verified the hypotheses on the one-dimensional marginal distributions in Proposition 4.3.6 for all measures in $\mathcal{M}_r$. Since $\mathcal{M}_o \subset \mathcal{M}_r$ for each $r < 0$, we may restrict ourselves to $r < 0$. Let $\mu \in \mathcal{M}_r$ and $\mu^{(i)}$ $(i \in \{1, \ldots, s\})$ be the one-dimensional marginal distribution of the i-th component. Let $C^{(i)} \subset \mathbb{R}$ be a Borel set, then, by definition,

$$\mu^{(i)}(C^{(i)}) = \mu(\mathbb{R} \times \ldots \times \mathbb{R} \times C^{(i)} \times \mathbb{R} \times \ldots \times \mathbb{R}),$$

and we obtain by (3.5) that

$$\mu^{(i)} \in \mathcal{M}_r^{(i)},$$

where $\mathcal{M}_r^{(i)} \subset \mathcal{P}(\mathbb{R})$ denotes the set of all Borel probability measures on $\mathbb{R}$ that satisfy (3.5).

Hence, by the theorem quoted above from [13], [84], $\mu^{(i)}$ has a density function $\theta = \theta^{(i)} : \mathbb{R} \to \mathbb{R}$ such that $\theta^{\frac{r}{1-r}}$ is convex $(r < 0)$. Furthermore, $\operatorname{supp} \mu = \mathbb{R}^s$ implies $\operatorname{supp} \mu^{(i)} = \mathbb{R}$.

Therefore, the function $\theta^{\frac{r}{1-r}}$ is continuous on $\mathbb{R}$, and, in view of the strict monotonicity of the transformation $t \to t^{\frac{r}{1-r}}$ $(t > 0)$, also θ is a continuous function on $\mathbb{R}$. Thus, for $\mu^{(i)}$ there exists a density function which is bounded on compact subsets of $\mathbb{R}$.

Since $\theta = \left(\theta^{\frac{r}{1-r}}\right)^{\frac{1-r}{r}}$ and $t \to t^{\frac{1-r}{r}}$ is strictly monotonically decreasing $(t > 0, r < 0)$, local minimizers of $\theta^{\frac{r}{1-r}}$ are local maximizers of θ and vice versa. Since $\theta^{\frac{r}{1-r}}$ is convex and $\int_{\mathbb{R}} \theta(\tau) d\tau = 1$, all local maximizers of θ are global ones, and the set of global maximizers is a bounded interval, which we denote by Ξ.

If there were local maximizers of θ outside Ξ, these would be local minimizers of $\theta^{\frac{r}{1-r}}$ outside Ξ, in contradiction to the convexity of $\theta^{\frac{r}{1-r}}$. Hence, θ fulfils the monotonicity property in Proposition 4.3.6.

Since, furthermore, θ is bounded on compact sets, this implies boundedness of θ on $\mathbb{R}$. ■

Using conditional distributions Proposition 4.3.6 can be extended to the general case where both z and A are random.

Proposition 4.3.10 *Assume (A1) – (A3) and let the conditional distribution $\mu_1^2(A, .)$ of z given A be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^s$ for μ_2-almost all $A \in \mathbb{R}^{ms}$, $(\mu_2 := \mu \circ \pi_2^{-1})$.Assume further that for any non-singular linear transformation $B \in L(\mathbb{R}^s, \mathbb{R}^s)$ and for μ_2-almost all $A \in \mathbb{R}^{ms}$ the* *one-dimensional marginal distributions of $\mu_1^2(A,.)\circ B$ have densities which are uniformly bounded with respect to A and which, outside some bounded interval not depending on A, are monotonically decreasing with growing absolute value of the argument. Then $Q(.,\mu)$ is Lipschitz continuous on any bounded subset of $I\!R^m$.*

Proof: Let $D \subset I\!R^m$ be bounded and $x', x'' \in D$. Then we have

$$\begin{aligned}
&|Q(x',\mu) - Q(x'',\mu)| \\
&\quad = \Big| \int\limits_{I\!R^S} (\Phi(z - Ax') - \Phi(z - Ax''))\mu(d(z,A)) \Big| \\
&\quad = \Big| \int\limits_{I\!R^{ms}} \int\limits_{I\!R^s} (\Phi(z - Ax') - \Phi(z - Ax''))\mu_1^2(A,dz)\mu_2(dA) \Big| \\
&\quad \leq \int\limits_{I\!R^{ms}} \Big| \int\limits_{I\!R^s} \Phi(z - Ax')\mu_1^2(A,dz) - \int\limits_{I\!R^s} \Phi(z - Ax'')\mu_1^2(A,dz) \Big| \mu_2(dA).
\end{aligned}$$

Our assumptions and Proposition 4.3.6 (cf. also its proof) imply that there exists a constant $\tilde{L} > 0$ (independent of A) such that

$$\Big| \int\limits_{I\!R^s} \Phi(z - Ax')\mu_1^2(A,dz) - \int\limits_{I\!R^s} \Phi(z - Ax'')\mu_1^2(A,dz) \Big| \leq \tilde{L} \cdot \|A\| \cdot \|x' - x''\|.$$

Hence we obtain

$$|Q(x',\mu) - Q(x'',\mu)| \leq \tilde{L} \int\limits_{I\!R^{ms}} \|A\|\mu_2(dA) \cdot \|x' - x''\|.$$

Note that

$$\begin{aligned}
\int\limits_{I\!R^{ms}} \|A\|\mu_2(dA) &= \int\limits_{I\!R^{ms}} \|A\| \int\limits_{I\!R^s} \mu_1^2(A,dz)\mu_2(dA) \\
&= \int\limits_{I\!R^S} \|A\|\mu(d(z,A)).
\end{aligned}$$

Therefore, assumption (A3) yields the assertion. ■

If the random variables z and A are independent then the assumptions of Proposition 4.3.10 can be verified using Proposition 4.3.6 since conditional and marginal distributions coincide.

In the case of dependent random variables z and A the verification of the assumptions in Proposition 4.3.10 is not so obvious. However, at least for the situation where we have a joint density for z and the random components of A we can calculate a density of $\mu_1^2(A,.)$ as a quotient of the joint density and the marginal density for A. The one-dimensional case of simple integer recourse is treated in detail in [60], [61], [66]. Beside continuity statements the authors derive sufficient conditions for the differentiability of $Q(.,\mu)$ and descriptions of the convex hull of the epigraph of $Q(.,\mu)$.

Let us now study the continuity of Q as a function jointly of $x \in I\!R^m$ and $\mu \in \mathcal{P}(I\!R^S)$ - the set of all Borel probability measures on $I\!R^S$. While at $I\!R^m$ we have the usual convergence, a suitable notion on $\mathcal{P}(I\!R^S)$ is that of weak convergence of probability measures which covers a number of specific convergence modes for probability measures (e.g. pointwise convergent densities, discretizations via conditional expectations, convergence of empirical measures). A sequence $\{\mu_n\}$ of probability measures in $\mathcal{P}(I\!R^S)$ is said to converge weakly to $\mu \in \mathcal{P}(I\!R^S)$, i.e. $\mu_n \xrightarrow{w} \mu$, if for any bounded continuous function $h : I\!R^S \to I\!R$ we have

$$\int_{I\!R^S} h(\xi)\mu_n(d\xi) \to \int_{I\!R^S} h(\xi)\mu(d\xi) \quad \text{as} \quad n \to \infty.$$

A detailed description of the topology of weak convergence of probability measures can be found in the monograph [9].
For notational convenience we introduce the following subset of probability measures

$$\Delta_{p,K}(I\!R^S) = \{\nu \in \mathcal{P}(I\!R^S) : \int_{I\!R^S} \|(z,A)\|^p \nu(d(z,A)) \le K\}$$

where $p > 1$ and $K > 0$ are fixed real numbers.

Proposition 4.3.11 *Assume (A1), (A2) and let $\mu \in \Delta_{p,K}(I\!R^S)$ for some $p > 1$, $K > 0$. If the conditional distribution $\mu_1^2(A,.)$ of z given A is absolutely continuous with respect to the Lebesgue measure on $I\!R^s$ for μ_2-almost all $A \in I\!R^{ms}$, then Q, as a function from $I\!R^m \times \Delta_{p,K}(I\!R^S)$ to $I\!R$, is continuous on $I\!R^m \times \{\mu\}$.*

Proof: Take an arbitrary $x \in I\!R^m$ and consider sequences $\{x_n\}$, $\{\mu_n\}$ in $I\!R^m$ and $\Delta_{p,K}(I\!R^S)$, respectively, such that $x_n \to x$ and $\mu_n \xrightarrow{w} \mu$ as $n \to \infty$. We introduce functions $h_n : I\!R^S \longrightarrow I\!R$ and $h : I\!R^S \longrightarrow I\!R$ defined by

$$h_n(z,A) = \Phi(z - Ax_n) \qquad \text{and} \qquad h(z,A) = \Phi(z - Ax)$$

which are measurable due to the lower semicontinuity of Φ. Consider the set $E_o(x)$ of all those $(z,A) \in I\!R^S$ such that there exists a sequence $\{(z_n,A_n)\}_{n=1}^{\infty}$ in $I\!R^S$ with

$$(z_n,A_n) \to (z,A) \qquad \text{and} \qquad h_n(z_n,A_n) \not\to h(z,A).$$

In our situation it holds $E_o(x) = E(x)$ with $E(x)$ as in Proposition 4.3.2. Indeed, the inclusion $E_o(x) \subseteq E(x)$ is easy to see. For the reverse inclusion let $(z, A) \in E(x)$ and $\{t_n\}_{n=1}^{\infty}$ be a sequence in $I\!R^s$ such that $t_n \to z{-}Ax$ and $\Phi(t_n) \not\to \Phi(z{-}Ax)$ as $n \to \infty$. Now consider the sequence $\{z_n, A_n)\}_{n=1}^{\infty}$ given by $A_n = A$ and $z_n = t_n + Ax_n$. Then it holds

$$(z_n, A_n) \to (z, A) \quad \text{and} \quad h_n(z_n, A_n) = \Phi(z_n - A_n x_n) = \Phi(t_n) \not\to \Phi(z - Ax) = h(z, A).$$

Hence, $(z, A) \in E_o(x)$.

Our assumption now implies $\mu(E_o(x)) = 0$ (cf. proof of Corollary 4.3.3) and we can apply Rubin's Theorem ([9], Theorem 5.5). This yields

$$\mu_n \circ h_n^{-1} \xrightarrow{w} \mu \circ h^{-1} \qquad \text{as} \quad n \to \infty. \tag{3.6}$$

To end up with

$$\int\limits_{I\!R^S} h_n(z, A)\mu_n(d(z, A)) \underset{n\to\infty}{\longrightarrow} \int\limits_{I\!R^S} h(z, A)\mu(d(z, A)), \tag{3.7}$$

which, of course, is just the assertion, we will show that

$$\lim_{a\to\infty} \sup_n \int\limits_{\{(z,A):|h_n(z,A)|\geq a\}} |h_n(z, A)|\mu_n(d(z, A)) = 0. \tag{3.8}$$

Since $p > 1$, it holds

$$\begin{aligned} &\int\limits_{I\!R^S} |h_n(z, A)|^p \mu_n(d(z, A)) \\ &\quad\geq \int\limits_{\{|h_n(z,A)|\geq a\}} |h_n(z, A)| \cdot |h_n(z, A)|^{p-1} \mu_n(d(z, A)) \\ &\quad\geq a^{p-1} \int\limits_{\{|h_n(z,A)|\geq a\}} |h_n(z, A)|\mu_n(d(z, A)). \end{aligned}$$

Therefore

$$\int\limits_{\{|h_n(z,A)|\geq a\}} |h_n(z, A)|\mu_n(d(z, A)) \leq a^{1-p} \int\limits_{I\!R^S} |h_n(z, A)|^p \mu_n(d(z, A)). \tag{3.9}$$

Proposition 4.2.1 and $h_n(0) = 0$ imply

$$|h_n(z, A)|^p \leq (\alpha\|z\| + \alpha\|x_n\| \cdot \|A\| + \beta)^p.$$

Since $\{x_n\}$ is bounded and all μ_n belong to $\Delta_{p,K}(I\!R^S)$ we now have a positive constant c such that

$$\int_{I\!R^S} |h_n(z,A)|^p \mu_n(d(z,A)) \le c \qquad \text{for all} \quad n \in I\!N.$$

Using (3.9) we thus obtain (3.8). Finally, (3.6) and Theorem 5.4 in [9] yield (3.7), and the proof is complete. ∎

Remark 4.3.12 *It is straightforward to replace the above assumption on* $\mu_1^2(A,.)$ *by* $\mu(E(x)) = 0$ *(cf. Proposition 4.3.2) and to end up with continuity of* $Q : I\!R^m \times \Delta_{p,K}(I\!R^S) \to I\!R$ *at* (x,μ).

Proposition 4.3.11 extends corresponding results for non-integer stochastic programs in [48], [88]. From an example in [88] it is also clear that the joint continuity of Q is lost if there is no assumption finally leading to the uniform integrability in (3.8). We have achieved this by claiming that $\mu \in \Delta_{p,K}(I\!R^S)$, $p > 1$, $K > 0$.
We close this section with an example illustrating the difficulties that occur when aiming at quantitative continuity results for $Q(x,.)$ as a function on a suitable (metric) space of probability measures. For non-integer stochastic programs such results can be obtained when equipping a suitable subset of $\mathcal{P}(I\!R^S)$ with the Wasserstein metric (Proposition 1.3.2 in Chapter 1). We will give an example that, for stochastic integer programs, there is no Hölder continuity estimate for $Q(x,.)$ with respect to the Wasserstein distance. This also means that there cannot be a Hölder estimate with respect to the Prokhorov and the Dudley (or β-) metric, respectively ([81], [92]).
For the comfort of the reader we briefly introduce the Wasserstein distance $W_1(\mu,\nu)$ of two probability measures μ and ν belonging to

$$\mathfrak{M}_1(I\!R^S) := \{\mu' \in \mathcal{P}(I\!R^S) : \int_{I\!R^S} \|z'\|\mu'(dz') < +\infty\}.$$

It is given by

$$W_1(\mu,\nu) = \inf\Big\{ \int_{I\!R^S \times I\!R^S} \|z' - z''\|\eta(dz',dz'') : \eta \in D(\mu,\nu)\Big\}$$

where

$$D(\mu,\nu) := \{\eta \in \mathcal{P}(I\!R^S \times I\!R^S) : \eta \circ \pi_1^{-1} = \mu,\ \eta \circ \pi_2^{-1} = \nu\}$$

and π_1, π_2 denote the first and second projections, respectively.
For details we refer to [81] where it is also shown that for $\mu, \nu \in \mathfrak{M}_1(I\!R)$

$$(3.10) \qquad W_1(\mu,\nu) = \int_{-\infty}^{\infty} |F_\mu(t) - F_\nu(t)| dt$$

where F_μ, F_ν denote the distribution functions of μ and ν, respectively.

Example 4.3.13 *Let in (1.2)* $\Phi(z-x) = \min\{y : y \geq z-x,\ y \in Z\!\!\!Z\}$*, i.e.* $A = 1$ *non-random, and let* μ *be a distribution of* z *with support contained in the closed interval* $[-1/2, +1/2]$*. It can be computed (cf. also the explicit formulae in [66]) that for* $x \in [-1/4, 1/4]$ *we have* $Q(x,\mu) = 1 - F_\mu(x)$*, where* F_μ *again denotes the distribution function of* μ*. Consider* $\mu_n \in \mathcal{P}(I\!R)$ $(n \geq 1)$ *with support in* $[-1/2, 1/2]$ *and continuous distribution function* F_{μ_n} *fulfilling*

$$F_{\mu_n}(t) = \begin{cases} t^{1/n} + \frac{1}{2} & \text{for} \quad 0 \leq t \leq r_n \\ -|t|^{1/n} + \frac{1}{2} & \text{for} \quad -r_n \leq t \leq 0 \end{cases}$$

with a suitably fixed real number $r_n > 0$*,* $r_n < (\frac{1}{2})^n$*. Let* $\varepsilon_n > 0$ *such that* $\varepsilon_n < r_n$*. We construct perturbations* μ_{n,ε_n} *of* μ_n *whose distribution functions coincide with those of* μ_n *for* $t \notin [-\varepsilon_n, \varepsilon_n]$ *and which are defined on* $[-\varepsilon_n, \varepsilon_n]$ *as follows*

$$F_{\mu_{n,\varepsilon_n}}(t) = \begin{cases} \frac{1}{2} - \varepsilon_n^{1/n} & \text{for} \quad -\varepsilon_n \leq t \leq 0 \\ \frac{1}{2} - \varepsilon_n^{1/n} + 2\varepsilon_n^{\frac{1-n}{n}} \cdot t & \text{for} \quad 0 \leq t \leq \varepsilon_n. \end{cases}$$

Using (3.10) we compute $W_1(\mu_n, \mu_{n,\varepsilon_n}) = \varepsilon_n^{\frac{n+1}{n}}$*. On the other hand,*
$|Q(0,\mu_n) - Q(0,\mu_{n,\varepsilon_n})| = |F_{\mu_n}(0) - F_{\mu_{n,\varepsilon_n}}(0)| = \varepsilon_n^{1/n}$*. Hence*
$|Q(0,\mu_n) - Q(0,\mu_{n,\varepsilon_n})| = W_1(\mu_n, \mu_{n,\varepsilon_n})^{1/n+1}$*. Since the construction was possible for any* $n \in I\!N$*,* $n \geq 1$*, there is no* W_1*-based Hölder estimate for* $Q(x, .)$*.*

4.4 Stability

In this section we study consequences of the above continuity results for the stability of

$$P(\mu) \qquad \min\{g(x) + Q(x,\mu) : x \in C\}$$

when the underlying measure μ is subjected to perturbations. Of course, $P(\mu)$ is a non-convex program, and, hence, also local minimizers should be included into the analysis. Therefore, beside Berge's classical stability theory for abstract parametric

programs [7], local stability results from [57] and [86] will be the main tools for our investigations. We will see that, having the continuity properties of Section 4.3 at one's disposal and using the techniques of [7], [57], [86], it is only a small step to arrive at the desired stability of $P(\mu)$.
Let $V \subset I\!R^m$ be an arbitrary subset and $\mathrm{cl}\, V$ denote the closure of V. Then we introduce the following localized versions for the optimal-value function and the solution set mapping:

$$\begin{aligned} \varphi_V(\mu) &:= \inf\{g(x) + Q(x,\mu) : x \in C \cap \mathrm{cl}\, V\} \\ \psi_V(\mu) &:= \{x \in C \cap \mathrm{cl}\, V : \ g(x) + Q(x,\mu) = \varphi_V(\mu)\}. \end{aligned}$$

A central observation in [57], [86] is that local minimizers of parametric programs may behave unstable when directly transferring assumptions from global stability analysis. For local stability analysis it turns out crucial that considerations include all local minimizers that are, in some sense, nearby the minimizers one is interested in. This leads to the concept of a complete local minimizing set (CLM set) coined in [86], which can be formulated in our terminology as follows:
Given $\mu \in \mathcal{P}(I\!R^s)$, a non-empty set $M \subset I\!R^m$ is called a CLM set for $P(\mu)$ with respect to an open set $V \subset I\!R^m$ if $M \subset V$ and $M = \psi_V(\mu)$. Of course, the set of global minimizers is always a CLM set; further examples are strict local minimizers. For more details consult [57], [86].
Considering $P(\mu)$ as a parametric program whose parameter space is $\mathcal{P}(I\!R^S)$ endowed with the topology of weak convergence of probability measures (cf. Section 4.3 and [9]) we have the following result:

Proposition 4.4.1 *Assume (A1), (A2), let $\mu \in \Delta_{p,K}(I\!R^S)$ for some $p > 1$, $K > 0$ and let the conditional distribution $\mu_1^2(A,.)$ of z given A be absolutely continuous with respect to the Lebesgue measure on $I\!R^s$ for μ_2-almost all $A \in I\!R^{ms}$. Suppose further that $M \subset I\!R^m$ is a CLM set for $P(\mu)$ with respect to some bounded open set $V \subset I\!R^m$, i.e. $M = \psi_V(\mu)$.*
Then

(i) the function φ_V (from $\Delta_{p,K}(I\!R^S)$ to $I\!R$) is continuous at μ;

(ii) the multifunction ψ_V (from $\Delta_{p,K}(I\!R^S)$ to $I\!R^m$) is Berge upper semicontinuous at μ, i.e. for any open set G in $I\!R^m$ with $G \supset \psi_V(\mu)$ there exists a neighbourhood U of μ in $\Delta_{p,K}(I\!R^S)$ such that $\psi_V(\mu') \subset G$ whenever $\mu' \in U$;

(iii) there exists a neighbourhood U' of μ in $\Delta_{p,K}(I\!R^S)$ such that for all $\mu' \in U'$ we have that $\psi_V(\mu')$ is a CLM set for $P(\mu')$ with respect to V.

Proof: Using Proposition 4.3.11 the proof of (i) and (ii) follows the lines of Berge's theory (cf. also [4], proof of Theorem 4.2.2) and is therefore not repeated here. When verifying (iii) the non-emptiness of $\psi_V(\mu')$ is gained via the lower semi-continuity of $Q(.,\mu')$ (Proposition 4.3.1); the CLM property then follows from (ii). ■

Let us add a few comments on the above proposition:
If the location of the (bounded) CLM set $\psi_V(\mu)$ is known, it could be helpful to know that the assumption on $\mu_1^2(A,.)$ can be relaxed to claiming that $\mu(E(x)) = 0$ for any $x \in C \cap \mathrm{cl}\, V$ (cf. Remark 4.3.12). Indeed, for the mentioned analysis along the lines of Berge the continuity of Q is only needed on $(C \cap \mathrm{cl}\, V) \times \{\mu\}$.
To see that the continuity assumption on $\mu_1^2(A,.)$ can not be relaxed in general we consider the following example where only the right-hand side z is random (cf. (1.3)):

$$P(\mu) \qquad \min\{Q(x,\mu) : x \le 0\},$$

where

$$\begin{aligned} Q(x,\mu) &= \int_{I\!R} \Phi(z-x)\mu(dz), \\ \Phi(t) &= \min\{y : y \ge t,\ y \in Z\!\!\!Z\}. \end{aligned}$$

Let μ be the discrete probability measure with mass 1 at zero and consider a sequence $\{\mu_n\}$ in $\mathcal{P}(I\!R)$ where μ_n assigns mass 1 to z_n with $z_n > 0, z_n \to 0$. The sequence $\{\mu_n\}$ weakly converges to μ. Moreover, it holds $Q(x,\mu) = \lceil -x \rceil$ and $Q(x,\mu_n) = \lceil z_n - x \rceil$. We consider global minimizers (i.e. $V = I\!R$) and obtain $\varphi(\mu) = 0, \varphi(\mu_n) = 1$ for all n, showing that φ is not continuous at μ.
If one relaxes the CLM property of M to assuming that M is a bounded set of local minimizers to $P(\mu)$ then it is also possible to construct counterexamples. Here, the perturbed programs have no local minimizers at all near M how "small" the perturbation is ever taken.
When analyzing (iii) it is clear that in view of the lower semicontinuity of $Q(.,\mu')$ (Proposition 4.3.1) and the compactness of $\mathrm{cl}\, V$ the sets $\psi_V(\mu')$ are always non-empty. In this context, the essence of (iii) is that non-emptiness of $\psi_V(\mu')$ is not enforced by restricting the objective to a compact, but that, for μ' sufficiently close to μ, the sets $\psi_V(\mu')$ again consist of local minimizers to $P(\mu')$.
The analogous result to Proposition 4.4.1 for two-stage stochastic programs without integer requirements was derived in [48], [88].

In [1] the authors investigated the stability of general stochastic programs involving discontinuous integrands. Compared with their fairly comprehensible stability conditions, Proposition 4.4.1 rather focuses on conditions which are verifiable for stochastic programs with mixed-integer recourse.
Proposition 4.4.1 may also be read as a general justification for numerical procedures that rely on approximating the distribution μ by simpler ones. For instance, discretizing μ via conditional expectations ([49], [10]) yields, a weakly convergent sequence of probability measures, provided that support partitions become arbitrarily small. Proposition 4.4.1 then ensures convergence of local optimal values and optimal solutions. Of course, up to now there are no comprehensive algorithms to solve stochastic integer programs with discrete probability distributions. However, in the recent paper [61] some substantial progress was made for stochastic programs with simple integer recourse.

Acknowledgement
I wish to thank Willem K. Klein Haneveld (University of Groningen), Werner Römisch (Humboldt University Berlin), Leen Stougie (University of Amsterdam) and Maarten H. van der Vlerk (University of Groningen) for beneficial discussions. Moreover, I am indebted to two anonymous referees for their helpful comments.

Chapter 5

Rates of Convergence in Stochastic Programs with Complete Integer Recourse

Abstract
The stability of stochastic programs with mixed-integer recourse under perturbations of the integrating probability measure is considered from a quantitative viewpoint. Objective-function values of perturbed stochastic programs are related to each other via a variational distance of probability measures based on a suitable Vapnik-Červonenkis class of Borel sets in a Euclidean space. This leads to Hölder continuity of local optimal values. In the context of estimation via empirical measures the general results imply qualitative and quantitative statements on the asymptotic convergence of local optimal values and optimal solutions.

5.1 Introduction

Consider the following two-stage stochastic integer program:

$$P(\mu) \qquad \min\{g(x) + Q(x,\mu) : x \in C\},$$

where

$$(1.1) \qquad Q(x,\mu) = \int_{I\!R^s} \Phi(z - Ax)\mu(\mathrm{d}z)$$

and

$$(1.2) \quad \Phi(t) = \min\{q^T y + q'^T y' : Wy + W'y' = t,\ y' \geq 0,\ y \geq 0,\ y' \in I\!R^{m'},\ y \in Z\!\!Z^{\bar{m}}\}.$$

We assume that $g : I\!R^m \to I\!R$ is continuous, $C \subset I\!R^m$ non-empty, closed, $q \in I\!R^{\bar{m}}$, $q' \in I\!R^{m'}$, that $W \in L(I\!R^{\bar{m}}, I\!R^s)$, $W' \in L(R^{m'}, I\!R^s)$ are matrices with rational entries and that μ belongs to $\mathcal{P}(I\!R^s)$ – the set of all Borel probability measures on $I\!R^s$.
The model $P(\mu)$ arises from a minimization problem with uncertain constraint parameters whose realizations are not known when having to fix the (first-stage) decision variable x. Infeasibilities t occuring after the realization of the uncertain parameters can be compensated at cost $\Phi(t)$ by the second-stage optimization procedure (1.2). Altogether, $P(\mu)$ aims at finding a first stage decision x such that the sum of the first stage costs $g(x)$ and the expected compensation (or recourse) costs $Q(x, \mu)$ becomes minimal.
The above model essentially differs from traditional two-stage stochastic programs (cf. [46], [131]) by the integrality constraints in the second stage. Whereas integrality in the first stage (if at all) can be dealt with by fairly conventional means ([135]), its presence in the second stage is much more cumbersome since the integrand Φ in (1.1) is discontinuous. However, there are several examples in the literature showing that integrality of second-stage decisions has to or at least should be included into the model ([65], [83], [109]).
Structural properties of stochastic programs with integer recourse have been studied in [1], [60], [66], [83], [104], [118]. The present chapter extends convergence results from [1] and from Chapter 4 by rates of convergence with respect to suitable distances of probability measures. More specifically, we consider $P(\mu)$ as a parametric program with the parameter μ varying in some (metric) space of probability measures. Then we derive Hölder estimates for Q (cf. (1.1)) as a function of the integrating probability measure which leads to corresponding results for local optimal values of $P(\mu)$. Our analysis is motivated by the incomplete information on the underlying measure μ that is often encountered (e.g. [27]). If μ is approximated by empirical measures then our general results specify to qualititative and quantitative results on the asymptotic convergence of (local) optimal values and optimal solutions. Another motivation for studying $P(\mu)$ as a parametric program in μ is given by numerical techniques that rely on approximating μ by simpler measures ([10], [49]).

When studying the quantitative continuity of Q as a function of μ a proper distance of probability measures has to be selected at the very beginning. Here, we want to understand by "proper" that the distance should both fit to the discontinuous integrand Φ and metrize (possibly under mild additional hypotheses) the weak convergence of probability measures ([9]). The difficulty of selecting a suitable probability distance is illustrated by Example 4.3.13 in Chapter 4 which shows that probability metrics like the Wasserstein distance ([81]), which led to convergence rates for $Q(x, .)$ in the non-integer recourse case ([93]), fail when it comes to integer recourse.

In the present chapter we propose the following variational distance (or discrepancy) for quantitative investigations of integer recourse models:

$$\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu, \nu) := \sup\{|\mu(B) - \nu(B)| : \ B \in \mathcal{B}_{\mathcal{K}}\}$$

where $\mathcal{B}_{\mathcal{K}}$ is a suitable class of convex Borel sets in $I\!R^s$ to be specified in Section 5.3. If the second stage (1.2) is a pure-integer linear program, then $\mathcal{B}_{\mathcal{K}}$ can be taken as the class of all lower left orthants in $I\!R^s$, and $\alpha_{\mathcal{B}_{\mathcal{K}}}$ coincides with the uniform (or Kolmogorov-Smirnov) distance of distribution functions. In the general case, $\mathcal{B}_{\mathcal{K}}$ still can be selected as a Vapnik-Červonenkis class of subsets of $I\!R^s$, which allows some interesting conclusions on the asymptotic behaviour of estimators based on empirical measures (see Section 5.5 for details).

The chapter is organized as follows: In Section 5.2 we collect some prerequisites from parametric integer programming on the value function Φ (cf. (1.2)). Section 5.3 contains the central Hölder estimates for $Q(x, .)$. Consequences for the quantitative stability of local optimal values of $P(\mu)$ are derived in Section 5.4. In Section 5.5 we elaborate the special case where μ is estimated via empirical measures. Throughout, $\|.\|$ denotes the ℓ^∞-norm in the Euclidean space under consideration; $\mathfrak{B}(t, r)$ denotes the closed ball around t with radius r (with respect to $\|.\|$).

5.2 Properties of the Value Function

As prerequisites for our subsequent considerations we have to collect some properties of the value function Φ (cf. (1.2)). The basic literature in this respect consists of [4], [5], [12]. Let us assume that, for each $t \in I\!R^s$, the constraint set $M(t) \subseteq Z\!\!\!Z^{\bar{m}} \times I\!R^{m'}$ in (1.2) is non-empty and that $\Phi(0) = 0$. Then $\Phi(t) \in I\!R$ for all $t \in I\!R^s$ (cf. e.g. [71], Prop. 1.6.7). Since Φ is discontinuous, in general, it is interesting to ask for continuity regions. Denoting by $\mathrm{pr}_1(M(t))$ the projection of $M(t)$ to $Z\!\!\!Z^{\bar{m}}$ we have the following result:

Lemma 5.2.1 ([4], Theorem 3.4.3, Theorem 5.6.5).
The restrictions of Φ to subsets of $I\!R^s$ where $\mathrm{pr}_1(M(.))$ is constant are continuous.

Subsets of $\mathbb{R}^s$ where the projection $\mathrm{pr}_1(M(.))$ is constant can be described via $\mathcal{K} := W'(\mathbb{R}_+^{m'})$ – the positive span of W':

Lemma 5.2.2 ([4], Lemma 5.6.1, Lemma 5.6.2, Theorem 5.6.3).
There exists a countable partition $\mathbb{R}^s = \bigcup_{i=1}^{\infty} \mathcal{B}_i$ *such that*

(i) $\mathrm{pr}_1(M(.))$ *is constant on* $\mathcal{B}_i$ *and, hence,* $\Phi|_{\mathcal{B}_i}$ *is continuous;*

(ii) *each of the sets* $\mathcal{B}_i$ *has a representation* $\mathcal{B}_i = \{t_i + \mathcal{K}\} \setminus \bigcup_{j=1}^{N_o} \{t_{ij} + \mathcal{K}\}$ *where* $t_i, t_{ij} \in \mathbb{R}^s$ $(i \in \mathbb{N},\ j = 1, \ldots, N_o)$, *and* N_o *does not depend on* i.

In [4] the rationality of both W and W' is employed to establish the above result. The authors first show that $\mathcal{B}_i$ can be represented as the (set-theoretic) difference of an (infinite) intersection and an (infinite) union of polyhedral cones. Utilizing the rationality of W they show that the infinite intersection can be replaced by just one cone; the rationality of W' implies that the infinite union of cones can be replaced by a finite one. In this context, we refer to an example at page 58 in [4] where some pathologies are illustrated that can occur if W contains irrational elements.
Lemma 5.2.2 implies a representation of $\Phi|_{\mathcal{B}_i}$. It holds

$$\begin{aligned} \Phi|_{\mathcal{B}_i}(t) &= \min\{q^T y + q'^T y' : W'y' = t - Wy;\ y \in \mathrm{pr}_1(M(t)),\ y' \geq 0\} \\ &= \min\{q^T y + \tilde{\Phi}(t - Wy) : y \in \mathrm{pr}_1(M(t))\} \end{aligned} \tag{2.1}$$

where $\tilde{\Phi}(\tilde{t}) := \min\{q'^T y' : W'y' = \tilde{t},\ y' \geq 0\}$ denotes a value function of a linear program with parameters in the right-hand side of the constraints. The assumptions on Φ imposed at the beginning imply that $\tilde{\Phi}(\tilde{t}) \in \mathbb{R}$ for any $\tilde{t} \in \mathrm{pos}\, W' = \mathcal{K}$. Furthermore, it is well-known from the literature ([72], [126]) that $\tilde{\Phi}|_{\mathcal{K}}$ admits a representation

$$\tilde{\Phi}|_{\mathcal{K}}(\tilde{t}) = \max_{i=1,\ldots,\tilde{N}} \tilde{d}_i^T \tilde{t}$$

where $\tilde{d}_i$ $(i = 1, \ldots \tilde{N})$ are determined by q' and W'.
Thus, the representation (2.1) says that $\Phi|_{\mathcal{B}_i}$ is the pointwise minimum of countably many continuous, piecewise linear functions. Moreover, the above argument proves the following lemma:

Lemma 5.2.3 $\Phi|_{\mathcal{B}_i}$ *is Lipschitz continuous with a constant* $L_o > 0$ *not depending on* i.

Let us now turn to some proximity results for the value function Φ and for optimal solutions of (1.2).

Lemma 5.2.4 ([5], Theorem 8.1; [12], Theorem 2.1).
There exist constants $\beta_1 > 0$, $\beta_2 > 0$, $\gamma_1 > 0$, $\gamma_2 > 0$ such that for all $t_1, t_2 \in I\!R^s$ we have

(i) $$|\Phi(t_1) - \Phi(t_2)| \leq \beta_1 \|t_1 - t_2\| + \gamma_1,$$

(ii) *if $(y_1, y_1') \in Z\!\!Z^{\bar{m}} \times I\!R^{m'}$ is optimal in (1.2) with respect to parameter t_1, then there exists a $(y_2, y_2') \in Z\!\!Z^{\bar{m}} \times I\!R^{m'}$ optimal with respect to t_2 such that*

$$\|(y_1, y_1') - (y_2, y_2')\| \leq \beta_2 \|t_1 - t_2\| + \gamma_2.$$

A first consequence of the above lemma which will turn out useful later on is the following.

Lemma 5.2.5 *There exists a number $N_1 \in I\!N$ such that for any $t_o \in I\!R^s$ the ball $\mathfrak{B}(t_o, 1)$ is intersected by at most N_1 different continuity regions $\mathcal{B}_i$ of Φ.*

Proof: Let $t_o \in I\!R^s$ be arbitrary and (y_o, y_o') be optimal in (1.2) with respect to t_o. Let $t \in \mathfrak{B}(t_o, 1)$. By Lemma 5.2.4(ii) there exists a $(y, y') \in Z\!\!Z^{\bar{m}} \times I\!R^{m'}$ optimal with respect to t such that

$$\|(y_o y_o') - (y, y')\| \leq \beta_2 \|t_o - t\| + \gamma_2 \leq \beta_2 + \gamma_2 =: r_o,$$

i.e. for any $t \in \mathfrak{B}(t_o, 1)$ there exists an optimal (y, y') whose y-component belongs to $B(y_o, r_o) \cap Z\!\!Z^{\bar{m}}$.
Recalling the representation (2.1) we obtain that for any $t \in \mathfrak{B}(t_o, 1)$

$$\Phi(t) = \min\{q^T y + \tilde{\Phi}(t - Wy) : y \in \mathrm{pr}_1(M(t)) \cap B(y_o, r_o)\}.$$

Since there are only finitely many different subsets of $\mathrm{pr}_1(M(t)) \cap B(y_o, r_o)$ the proof is complete. ∎

A situation deserving special attention is that of pure integer recourse, i.e. Φ is given by

$$\Phi(t) = \min\{q^T y : Wy \geq t, y \in Z\!\!Z_+^{\bar{m}}\}.$$

Introducing slack variables $y' \in I\!R^s$ this fits into the above setting and we have $\mathcal{K} = -I\!R_+^s$. Furthermore, $q' = 0$ which leads to $\tilde{\Phi}(\tilde{t}) = 0$ for all $\tilde{t} \in \mathcal{K}$, and Lemma 5.2.3 specifies to $\Phi|_{\mathcal{B}_i}$ being constant for all i.

5.3 Rates of Convergence for Expectation Functions

The following basic assumptions are imposed throughout. They guarantee that (1.1) and (1.2) are well-defined.

(A1) For all $t \in I\!R^s$ there exist $y \in Z\!\!Z^{\bar{m}}$, $y' \in I\!R^{m'}$ such that $y \geq 0$, $y' \geq 0$ and $Wy + W'y' = t$.

(A2) There exists a $u \in I\!R^s$ such that $W^T u \leq q$ and $W'^T u \leq q'$.

(A3) It holds that $\int\limits_{I\!R^s} \|z\|\mu(\mathrm{d}z) < +\infty$.

Note that these assumptions can be read as natural extensions of counterparts for linear stochastic programs with non-integer recourse ([46], [131]). In Chapter 4 it is shown that (A1) – (A3) imply that $Q(.,\mu)$ is a lower semicontinuous real-valued function on $I\!R^m$. Let us further remark that (A2) is equivalent to $\Phi(0) = 0$ (which appeared as an assumption in Section 5.2) and that (A1) in particular implies that $\mathcal{K} = \mathrm{pos}\, W'$ has a non-empty interior.

We now address the problem of finding upper estimates for $|Q(x,\mu) - Q(x,\nu)|$ in terms of the underlying probability measures $\mu, \nu \in \mathcal{P}(I\!R^s)$. Proposition 4.3.11 in Chapter 4 contains a sufficient condition for the joint continuity of Q in x and μ with respect to the product topology of the usual one on $I\!R^m$ and that of weak convergence on $\mathcal{P}(I\!R^s)$.

In what follows we will show that a certain discrepancy (or variational distance of probability measures) leads to Hölder estimates of $Q(x,.)$. Given a class $\mathcal{B}_o \subset \mathcal{B}(I\!R^s)$ of Borel sets in $I\!R^s$, the discrepancy $\alpha_{\mathcal{B}_o}(\mu,\nu)$ is defined by

$$\alpha_{\mathcal{B}_o}(\mu,\nu) := \sup\{|\mu(B) - \nu(B)| : \ B \in \mathcal{B}_o\}.$$

Popular instances of $\mathcal{B}_o$ in the literature are the families of all lower left orthants or of all convex Borel sets in $I\!R^s$ (cf. e.g. [8]). For us, a class in between these two families will be important: Let $\mathcal{B}_{\mathcal{K}} \subseteq \mathcal{B}(I\!R^s)$ denote the class of all (closed) bounded polyhedra in $I\!R^s$ each of whose facets (i.e. $(s-1)$-dimensional faces) parallels a facet of $\mathcal{K} = \mathrm{pos}\, W'$ or a facet of $\bigtimes\limits_{i=1}^{s} [0,1]$.

The discrepancy $\alpha_{\mathcal{B}_{\mathcal{K}}}$ is then even a metric on $\mathcal{P}(I\!R^s)$ which can be seen as follows: From the definition of a discrepancy $\alpha_{\mathcal{B}_o}$ we deduct the basic properties of a metric, except the one that $\alpha_{\mathcal{B}_o}(\mu,\nu) = 0$ implies $\mu = \nu$. Let $\mathcal{B}_{orth}$ denote the family of all (closed) lower left orthants in $I\!R^s$ and $\mathcal{B}_{box}$ denote the family of all boxes, i.e.

all (closed) bounded polyhedra in $I\!R^s$ whose facets are parallel to facets of $\bigtimes_{i=1}^{s}[0,1]$. Then it holds for all $\mu, \nu \in \mathcal{P}(I\!R^s)$

$$\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu) \geq \alpha_{\mathcal{B}_{box}}(\mu,\nu) \geq \alpha_{\mathcal{B}_{orth}}(\mu,\nu)$$

where the last inequality follows from the monotonicity of μ and ν on ascending sequences of sets. Therefore, $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu) = 0$ implies $\alpha_{\mathcal{B}_{orth}}(\mu,\nu) = 0$. Since $\alpha_{\mathcal{B}_{orth}}$ is just the uniform (or Kolmogorov-Smirnov) distance of distribution functions (which is known to be a metric on $\mathcal{P}(I\!R^s)$,[81]), this implies $\mu = \nu$.
Let us further introduce

$$\Delta_{p,K}(I\!R^s) := \Big\{\nu \in \mathcal{P}(I\!R^s) : \int\limits_{I\!R^s} \|z\|^p \nu(\mathrm{d}z) \leq K\Big\}$$

where $p > 1$ and $K > 0$ are fixed constants.
Then, the following estimate with respect to the discrepancy $\alpha(\mu,\nu) := \alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu)$ is valid:

Proposition 5.3.1 *Suppose* (A1), (A2) *and let $D \subset I\!R^m$ be non-empty and bounded.*
Then there exist constants $L > 0$ and $\delta > 0$ such that

$$\sup_{x \in D} |Q(x,\mu) - Q(x,\nu)| \leq L \cdot \alpha(\mu,\nu)^{\frac{p-1}{p(s+1)}}$$

whenever $\mu, \nu \in \Delta_{p,K}(I\!R^s)$, $\alpha(\mu,\nu) < \delta$.

Proof: Let $x \in D$ be arbitrary and $\mu, \nu \in \Delta_{p,K}(I\!R^s)$ such that $\alpha(\mu,\nu) < \delta_o := 1$. The bound δ_o will be further shrinked in the course of the proof. Define the radii $R := \alpha(\mu,\nu)^{-1/p(s+1)}$ and $r := \alpha(\mu,\nu)^{\frac{p-1}{p(s+1)}}$.
It holds

$$\begin{aligned} (3.1) \qquad |Q(x,\mu) - Q(x,\nu)| &= \Big| \int\limits_{I\!R^s} \Phi(z - Ax)(\mu - \nu)(\mathrm{d}z) \Big| \\ &\leq \Big| \int\limits_{Ax+B_o} \Phi(z - Ax)(\mu - \nu)(\mathrm{d}z) \Big| + \Big| \int\limits_{I\!R^s \setminus \{Ax+B_o\}} \Phi(z - Ax)(\mu - \nu)(\mathrm{d}z) \Big| \end{aligned}$$

where $B_o = \mathfrak{B}(0,R)$.
Now split B_o into ($l^\infty-$) balls of radius r. The splitting is carried out in a disjunctive way, i.e. B_o is partitioned into disjoint (measurable) sets whose closures are closed balls with radius r. In the splitting, balls of radius r are used as long as possible;

possible gaps are filled with maximal balls of radius less than r. Hence, the number of elements in the partition of B_o is bounded above by $(\frac{R}{r}+1)^s \leq (\frac{2R}{r})^s$.
Each element of the above partition intersects certain continuity regions $\mathcal{B}_i$ of Φ (cf. Lemma 5.2.2); by Lemma 5.2.5 it intersects at most N_1 of such regions. From Lemma 5.2.2(ii) it can be seen that each $\mathcal{B}_i$ splits into disjoint (measurable) subsets whose closures are polyhedra with facets parallel to suitable facets of $\mathcal{K}$. Moreover, the number of such subsets can be bounded above by a constant only depending on N_o (cf. Lemma 5.2.2(ii)) and the number of facets of $\mathcal{K}$. The constant does not depend on i.
Altogether, B_o is splitted into disjoint subsets B_j $(j = 1, \ldots, N)$ whose closures are polyhedra with facets parallel to suitable facets of $\mathcal{K}$ or of $\bigtimes_{i=1}^{s}[0,1]$, i.e. the closures belong to $\mathcal{B}_{\mathcal{K}}$. Furthermore, there exists a constant $\kappa > 0$ which is independent on R and r such that $N \leq \kappa \cdot (\frac{R}{r})^s$. From each of the sets B_j $(j = 1, \ldots, N)$ we pick an element b_j and continue (3.1):

$$
\begin{aligned}
(3.2)\ & |Q(x,\mu) - Q(x,\nu)| \\
\leq\ & \Big|\sum_{j=1}^{N} \int_{Ax+B_j} (\Phi(z-Ax) - \Phi(b_j))(\mu-\nu)(\mathrm{d}z) + \sum_{j=1}^{N} \int_{Ax+B_j} \Phi(b_j)(\mu-\nu)(\mathrm{d}z)\Big| \\
& + \Big| \int_{I\!R^s \setminus \{Ax+B_o\}} \Phi(z-Ax)(\mu-\nu)(\mathrm{d}z)\Big| \\
\leq\ & \sum_{j=1}^{N} \Big(\int_{Ax+B_j} |\Phi(z-Ax) - \Phi(b_j)|\mu(\mathrm{d}z) + \int_{Ax+B_j} |\Phi(z-Ax) - \Phi(b_j)|\nu(\mathrm{d}z)\Big) \\
& + \sum_{j=1}^{N} |\Phi(b_j)| \cdot \Big| \int_{Ax+B_j} (\mu-\nu)(\mathrm{d}z)\Big| + \Big| \int_{I\!R^s \setminus \{Ax+B_o\}} \Phi(z-Ax)(\mu-\nu)(\mathrm{d}z)\Big| \\
\leq\ & L_o \cdot 2r(\mu(Ax+B_o) + \nu(Ax+B_o)) + \\
& + (\beta_1 R + \gamma_1) \cdot N \cdot \alpha(\mu,\nu) + \Big| \int_{I\!R^s \setminus \{Ax+B_o\}} \Phi(z-Ax)(\mu-\nu)(\mathrm{d}z)\Big|.
\end{aligned}
$$

The first member of the above sum results from Lemma 5.2.3 and the fact that diam $B_j \leq 2r$ for all $j = 1, \ldots, N$. Concerning the second member we refer to Lemma 5.2.4(i) and the fact that $\Phi(0) = 0$. Furthermore for any $j = 1, \ldots, N$

$$|\mu(Ax+B_j) - \nu(Ax+B_j)| \leq \sup\{|\mu(B) - \nu(B)| : B \in \mathcal{B}_{\mathcal{K}}\}$$

where, if necessary, $Ax + B_j$ is approximated by a monotone sequence of polyhedra in $\mathcal{B}_{\mathcal{K}}$ and the estimate is gained by passing to the limit.

Now suppose in addition that $\alpha(\mu,\nu) \le \delta_1$ where δ_1 is selected such that $\beta_1 R \ge \gamma_1$ and

$$\mathbb{R}^s \setminus \{Ax + B_o\} \subset \{\|z\| \in \mathbb{R}^s : \|z\| \ge \frac{1}{2}R\} \quad \text{for all} \quad x \in D$$

(note that D is bounded).
Then we can continue (3.2) as follows

$$\begin{aligned}
(3.3)\ & |Q(x,\mu) - Q(x,\nu)| \\
&\le 4L_o r + 2\beta_1 R \cdot N \cdot \alpha(\mu,\nu) + \int\limits_{\|z\|\ge\frac{1}{2}R} |\Phi(z-Ax)|\mu(\mathrm{d}z) + \int\limits_{\|z\|\ge\frac{1}{2}R} |\Phi(z-Ax)|\nu(\mathrm{d}z) \\
&\le 4L_o r + 2\beta_1 R \cdot N \cdot \alpha(\mu,\nu) + \int\limits_{\|z\|\ge\frac{1}{2}R} (\beta_1\|z-Ax\| + \gamma_1)(\mu+\nu)(\mathrm{d}z) \\
&\qquad\qquad (\text{by Lemma 5.2.4(i) and } \Phi(0)=0) \\
&\le 4L_o \cdot r + 2\beta_1 R \cdot N \cdot \alpha(\mu,\nu) + \int\limits_{\|z\|\ge\frac{1}{2}R} (\beta_1\|z\| + \beta_1\|Ax\| + \gamma_1)(\mu+\nu)(\mathrm{d}z).
\end{aligned}$$

Let $\alpha(\mu,\nu) \le \delta_2$ where δ_2 is selected such that $\beta_1\|z\| \ge \beta_1\|Ax\| + \gamma_1$ for all $x \in D$ and all $\|z\| \ge \frac{1}{2}R$. Then we can continue

$$\le 4L_o r + 2\beta_1 R \cdot N \cdot \alpha(\mu,\nu) + 2\beta_1 \int\limits_{\|z\|\ge\frac{1}{2}R} \|z\|(\mu+\nu)(\mathrm{d}z).$$

Now recall that $\mu,\nu \in \Delta_{p,K}(\mathbb{R}^s)$. This implies

$$\begin{aligned}
2K &\ge \int\limits_{\mathbb{R}^s} \|z\|^p(\mu+\nu)(\mathrm{d}z) \ge \int\limits_{\|z\|\ge\frac{1}{2}R} \|z\| \cdot \|z\|^{p-1}(\mu+\nu)(\mathrm{d}z) \\
&\ge \left(\frac{1}{2}R\right)^{p-1} \int\limits_{\|z\|\ge\frac{1}{2}R} \|z\|(\mu+\nu)(\mathrm{d}z).
\end{aligned}$$

Hence we can continue (3.3)

$$(3.4) \qquad |Q(x,\mu) - Q(x,\nu)| \le 4L_o r + 2\beta_1 R \cdot N \cdot \alpha(\mu,\nu) + 2\beta_1 \cdot 2K \cdot 2^{(p-1)} \cdot R^{-(p-1)}.$$

The above estimate holds for all $x \in D$ and all $\mu,\nu \in \Delta_{p,K}(\mathbb{R}^s)$ such that $\alpha(\mu,\nu) \le \delta := \min\{\delta_o,\delta_1,\delta_2\}$. Inserting $N \le \kappa(\frac{R}{r})^s$, $R = \alpha(\mu,\nu)^{-1/p(s+1)}$ and $r = \alpha(\mu,\nu)^{\frac{p-1}{p(s+1)}}$ yields the assertion. ∎

Corollary 5.3.2 *Suppose* (A1), (A2) *and that there exists a bounded set $S \subset I\!R^s$ such that both the supports of $\mu, \nu \in \mathcal{P}(I\!R^s)$ are contained in S. Let $D \subseteq I\!R^m$ be non-empty and bounded. Then there exist constants $L > 0$ and $\delta > 0$ such that*

$$\sup_{x \in D} |Q(x,\mu) - Q(x,\nu)| \leq L \cdot \alpha(\mu,\nu)^{1/s+1}$$

whenever $\alpha(\mu,\nu) < \delta$.

Proof: In the same notation as in the proof of Proposition 5.3.1 we define the radius $r := \alpha(\mu,\nu)^{1/s+1}$ and put $R > 0$ as a constant such that $S \subset Ax + B_o$ for any $x \in D$. Repeating the proof of Proposition 5.3.1 until (3.2) and inserting the expressions for r and $N \leq \kappa \cdot (\frac{R}{r})^s = \kappa \cdot R^s \cdot \alpha(\mu,\nu)^{-s/s+1}$ yields the desired estimate. ■

Remark 5.3.3 *A class $\mathcal{B}$ of Borel sets in $I\!R^s$ is called a μ-uniformity class if* $\sup\{|\mu(B) - \mu_n(B)| : B \in \mathcal{B}\} \to 0$ *holds for every sequence $\{\mu_n\}$ in $\mathcal{P}(I\!R^s)$ converging weakly to $\mu \in \mathcal{P}(I\!R^s)$ (cf.*[8]*). Theorem* 2.11 *in* [8] *says that the family $\mathcal{B}_c$ of all convex Borel sets in $I\!R^s$ is a μ-uniformity class if μ is absolutely continuous with respect to the Lebesgue measure in $I\!R^s$. Since $\mathcal{B}_{\mathcal{K}} \subset \mathcal{B}_c$, this implies that $\alpha(\mu_n,\mu) \to 0$ provided that $\{\mu_n\}$ converges weakly to μ and μ is absolutely continuous.*

Remark 5.3.4 *In Chapter 4, Proposition 4.3.11, it is stated that Q, as a function from $I\!R^m \times \Delta_{p,K}(I\!R^s)$ to $I\!R$, is continuous on $I\!R^m \times \{\mu\}$, provided that μ is absolutely continuous and $\Delta_{p,K}(I\!R^s)$ is equipped with weak convergence of probability measures. By the above remark this result is now obtained as a conclusion of Proposition* 5.3.1 *which, therefore, can be read as a quantification of the continuity result in Chapter 4.*

For the situation of pure integer recourse, i.e. the case where Φ in (1.2) is given by

$$\Phi(t) = \min\{q^T y : Wy \geq t,\ y \in Z\!\!Z_+^{\bar{m}}\}$$

some substantial improvements of Proposition 5.3.1 are possible.
First recall (cf. Section 5.2) that here $\mathcal{K} = -I\!R^s_+$, hence $\mathcal{B}_{\mathcal{K}}$ consists of all bounded polyhedra whose facets parallel those of $\bigtimes_{i=1}^{s}[0,1]$, i.e. of all boxes in $I\!R^s$.The discrepancy $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu)$ can now be bounded above by the Kolmogorov-Smirnov distance

$$\|F_\mu - F_\nu\|_\infty := \sup\{|F_\mu(t) - F_\nu(t)| : t \in I\!R^s\}$$

of the distribution functions F_μ and F_ν:

Lemma 5.3.5 *If $\mathcal{K} = -I\!R^s_+$ then $\alpha_{\mathcal{B}_\mathcal{K}}(\mu, \nu) \le 2^s \cdot \|F_\mu - F_\nu\|_\infty$.*

Proof: Let $B \in \mathcal{B}_\mathcal{K}$ and assume a representation $B = \bigtimes_{i=1}^{s} [\underline{b}_i, \bar{b}_i]$. Then B has exactly 2^s vertices b_j $(j = 1, \ldots, 2^s)$. Let $\varepsilon > 0$ and consider $B_\varepsilon = \bigtimes_{i=1}^{s} (\underline{b}_i - \varepsilon, \bar{b}_i]$ with vertices $b_{j,\varepsilon}$ $(j = 1, \ldots, 2^s)$. According to a well-known formula there exist $n_{j,\varepsilon} \in \{0, 1\}$ $(j = 1, \ldots, 2^s)$ such that $\mu(B_\varepsilon) = \sum_{j=1}^{2^s} (-1)^{n_{j,\varepsilon}} F_\mu(b_{j,\varepsilon})$.
Therefore

$$|\mu(B_\varepsilon) - \nu(B_\varepsilon)| \le 2^s \cdot \|F_\mu - F_\nu\|_\infty$$

for all $\varepsilon > 0$ and all $B \in \mathcal{B}_\mathcal{K}$.
Let $\varepsilon \downarrow 0$, the continuity of μ and ν on monotone sequences of sets then implies

$$|\mu(B) - \nu(B)| \le 2^s \cdot \|F_\mu - F_\nu\|_\infty \quad \text{for all} \;\; B \in \mathcal{B}_\mathcal{K}.$$

■

Another peculiarity we can benefit from is that here Lemma 5.2.3 holds with $L_o = 0$, i.e. Φ is constant on each of the continuity sets $\mathcal{B}_i$.

Proposition 5.3.6 *Let $P(\mu)$ have pure integer recourse. Suppose* (A1), (A2) *and let $D \subset I\!R^m$ be non-empty and bounded. Then there exist constants $L > 0$ and $\delta > 0$ such that*

$$\sup_{x \in D} |Q(x, \mu) - Q(x, \nu)| \le L \cdot \|F_\mu - F_\nu\|_\infty^{\frac{p-1}{p+s}}$$

whenever $\mu, \nu \in \Delta_{p,K}(I\!R^s)$, $\|F_\mu - F_\nu\|_\infty < \delta$.

Proof: Let $x \in D$ be arbitrary and $\mu, \nu \in \Delta_{p,K}(I\!R^s)$ such that $\|F_\mu - F_\nu\|_\infty < \delta_o := 1$. Define the radii $R := \|F_\mu - F_\nu\|_\infty^{-1/s+p}$ and $r = 1$. Let $B_o = \mathfrak{B}(0, R)$. In view of Lemma 5.2.5 and Lemma 5.2.2(ii) there exists a constant $\kappa > 0$, independent on R, such that B_o is intersected by at most $\kappa \cdot R^s$ rectangular continuity regions of Φ. Now repeat the proof of Proposition 5.3.1 until (3.4) and take into account that $L_o = 0$ and $N \le \kappa \cdot R^s$. This yields

$$|Q(x, \mu) - Q(x, \nu)| \le 2\beta_1 R \cdot \kappa R^s \cdot 2^s \|F_\mu - F_\nu\|_\infty + 2\beta_1 \cdot 2K \cdot 2^{(p-1)} \cdot R^{-(p-1)} \tag{3.5}$$

provided that $\mu, \nu \in \Delta_{p,K}(I\!R^s)$, $\|F_\mu - F_\nu\|_\infty \le \delta$ with some properly chosen $\delta > 0$. Inserting $R := \|F_\mu - F_\nu\|_\infty^{-1/s+p}$ completes the proof. ■

Corollary 5.3.7 *Let $P(\mu)$ have pure integer recourse. Suppose* (A1), (A2) *and that there exists a bounded set $S \subset I\!R^s$ such that both the supports of $\mu, \nu \in \mathcal{P}(I\!R^s)$ are contained in S. Let $D \subset I\!R^m$ be non-empty and bounded. Then there exist constants $L > 0$ and $\delta > 0$ such that*

$$\sup_{x \in D} |Q(x, \mu) - Q(x, \nu)| \leq L \cdot \|F_\mu - F_\nu\|_\infty$$

whenever $\|F_\mu - F_\nu\|_\infty < \delta$.

Proof: Adopt the notation from the above proof and put $R > 0$ as a constant such that $S \subset Ax + B_o$ for any $x \in D$. We obtain the same estimate as in (3.5) with the difference that the second member of the sum on the right does not appear.

■

5.4 Rates of Convergence for Optimal Values

In contrast to stochastic programs with non-integer recourse, where $Q(., \mu)$ in (1.1) is always convex, integer recourse models obey local minimizers which are not necessarily global ones. The subsequent analysis is directed to convergence rates for optimal values if the underlying probability measure in $P(\mu)$ is subjected to perturbations. From the literature ([57], [86]) it is well-known that already qualitative investigations on the convergence of optimal values necessitate an exclusion of pathological types of local minimizers in the unperturbed problem (cf. also Chapter 4). This leads to the concept of a complete local minimizing set (CLM set) (cf. [57], [86]) which we will introduce next.

Let $V \subset I\!R^m$ be arbitrary and let $\mathrm{cl}\, V$ denote the closure of V. Recall the shape of $P(\mu)$ and consider the following localized optimal value and set of optimal solutions

$$\begin{aligned} \varphi_V(\mu) &:= \inf\{g(x) + Q(x, \mu) : x \in C \cap \mathrm{cl}\, V\}, \\ \psi_V(\mu) &:= \{x \in C \cap \mathrm{cl}\, V : g(x) + Q(x, \mu) = \varphi_V(\mu)\}. \end{aligned}$$

Given $\mu \in \mathcal{P}(I\!R^s)$, a non-empty set $M \subset I\!R^m$ is called a CLM set for $P(\mu)$ with respect to an open set $V \subset I\!R^m$ if $M \subset V$ and $M = \psi_V(\mu)$.

Obvious examples for CLM sets are the set of global minimizers and isolated local minimizers (cf. [57], [86] for further details).

Theorem 5.4.1 *Suppose* (A1), (A2)*, let $\mu \in \mathcal{P}(I\!R^s)$ be absolutely continuous with respect to the Lebesgue measure on $I\!R^s$ and let there exist constants $p > 1$ and $K > 0$*

such that $\mu \in \Delta_{p,K}(I\!R^s)$*. Assume further that* $M \subset I\!R^m$ *is a CLM set for* $P(\mu)$ *with respect to some bounded open set* $V \subset I\!R^m$*. Then there exist constants* $L > 0$ *and* $\delta > 0$ *such that*

$$|\varphi_V(\mu) - \varphi_V(\nu)| \leq L \cdot \alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu)^{\frac{p-1}{p(s+1)}}$$

whenever $\nu \in \Delta_{p,K}(I\!R^s)$*,* $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu) < \delta$*.*

Proof: Since μ is absolutely continuous, Remark 5.3.3 says that $\mathcal{B}_{\mathcal{K}}$ is a μ-uniformity class. In view of Proposition 4.4.1(ii), (iii) in Chapter 4 then there exists a $\delta > 0$ such that $\psi_V(\nu)$ is a CLM set for $P(\nu)$ satisfying $\emptyset \neq \psi_V(\nu) \subset V$ for all $\nu \in \Delta_{p,K}(I\!R^s)$, $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu) < \delta$.
Let $\nu \in \Delta_{p,K}(I\!R^s)$, $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu) < \delta$ and $x_\nu \in \psi_V(\nu)$, $x_\mu \in \psi_V(\mu) = M$. Then it holds

$$\varphi_V(\mu) \leq g(x_\nu) + Q(x_\nu,\mu) \leq \varphi_V(\nu) + |Q(x_\nu,\mu) - Q(x_\nu,\nu)|$$

and

$$\varphi_V(\nu) \leq g(x_\mu) + Q(x_\mu,\nu) \leq \varphi_V(\mu) + |Q(x_\mu,\nu) - Q(x_\mu,\mu)|.$$

Hence

$$|\varphi_V(\mu) - \varphi_V(\nu)| \leq \sup_{x \in C \cap \mathrm{cl}\, V} |Q(x,\mu) - Q(x,\nu)|. \tag{4.1}$$

Since $C \cap \mathrm{cl}\, V$ is a bounded subset, we can apply Proposition 5.3.1 which completes the proof. ∎

Remark 5.4.2 *Let us point out that in the above proof we have also shown that the sets* $\psi_V(\nu)$ *are CLM sets and, therefore, sets of local minimizers to* $P(\nu)$ *(i.e. under minimization over* C*!). Nonemptiness of* $\psi_V(\nu)$ *alone is not sufficient for the latter, since the minimization is restricted to* $C \cap \mathrm{cl}\, V$*.*

Remark 5.4.3 *Improved versions of Proposition* 5.4.1 *are straightforward when adopting the more specific setting as in Corollary* 5.3.2*, Proposition* 5.3.6 *and Corollary* 5.3.7*. The rates obtained there directly extend to the convergence of local optimal values.*

5.5 Asymptotic Convergence

This section is devoted to studying some implications of our previous results for the asymptotic convergence of (local) optimal values and optimal solutions when estimating the underlying measure μ in $P(\mu)$ by empirical measures. Given a sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ of independent $\mathbb{R}^s$-valued random variables on some probability space $(\Omega, \mathfrak{A}, \mathbf{P})$ with joint distribution μ, the empirical measures $\mu_n(\omega)$ $(\omega \in \Omega$, $n \in \mathbb{N})$ are defined by

$$\mu_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i(\omega)}$$

where $\delta_{\xi_i(\omega)}$ denotes the measure with unit mass at $\xi_i(\omega)$ (cf. [24],[39],[76],[117]). A classiscal result in probability theory (Glivenko-Cantelli Theorem) states that

$$\|F_{\mu_n(\omega)} - F_\mu\|_\infty \underset{n\to\infty}{\longrightarrow} 0 \quad \text{for } \mathbf{P}\text{-almost all } \omega \in \Omega \quad .$$

Recalling that weak convergence of probability measures in $\mathcal{P}(\mathbb{R}^s)$ is equivalent to pointwise convergence of the distribution functions at continuity points of the limit function, the Glivenko-Cantelli Theorem asserts some uniformity of weak convergence. In contrast to the uniformity of weak convergence reflected in Remark 5.3.3 we, here, do not need that the limit measure μ is absolutely continuous. Proposition 5.3.6, Corollary 5.3.7 and Remark 5.4.3 now allow immediate consequences for the asymptotic convergence of the objective function and local optimal values when estimating μ in $P(\mu)$ via empirical measures. Instead of elaborating this issue we take a more general stand and show how to extend the uniformity argument from lower left orthants (distribution functions) to the class $\mathcal{B}_\mathcal{K}$ introduced in Section 5.3. This will allow us to prove asymptotic convergence of local optimal values for the general case of complete mixed-integer recourse. A proper tool from probability theory in this respect are Vapnik-Červonenkis classes of Borel sets in $\mathbb{R}^s$ ([76], [122]).
Given a family $\mathcal{B}_o \subset \mathcal{B}(\mathbb{R}^s)$, let $\mathcal{V}(\mathcal{B}_o)$ be the smallest $k \in \mathbb{N}$ such that for every set $E \subset \mathbb{R}^s$ with k elements, not every subset of E is of the form $E \cap B$, $B \in \mathcal{B}_o$. $\mathcal{B}_o$ is called a Vapnik-Červonenkis class (VČ class) if $\mathcal{V}(\mathcal{B}_o) < +\infty$.

Lemma 5.5.1 ([39],[76],[122]).
If $\mathcal{B}_o \subset \mathcal{B}(\mathbb{R}^s)$ is a VČ class, then

$$\sup\{|\mu_n(\omega)(B) - \mu(B)| : B \in \mathcal{B}_o\} \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{for } \mathbf{P}\textit{-almost all } \omega \in \Omega.$$

It is well-known (e.g. [39],[76],[117]) that the family of all closed halfspaces in $\mathbb{R}^s$ forms a VČ class. Moreover, if $\mathcal{B}', \mathcal{B}'' \subseteq \mathcal{B}(\mathbb{R}^s)$ are VČ classes then this is also true for $\mathcal{B}' \cap \mathcal{B}''$ and $\mathcal{B}' \cup \mathcal{B}''$ where $\mathcal{B}' \overset{\cup}{\cap} \mathcal{B}'' := \{B' \overset{\cup}{\cap} B'' : B' \in \mathcal{B}',\ B'' \in \mathcal{B}''\}$ ([76],[117]). Now observe, that for each polyhedron in $\mathcal{B}_{\mathcal{K}}$ the number of facets is bounded above by $2s$ plus twice the number of facets of $\mathcal{K}$. Together with the facts just mentioned this implies that $\mathcal{B}_{\mathcal{K}}$ is a VČ class.
Furthermore, each element of $\mathcal{B}_{\mathcal{K}}$ can be represented as an upper level set of a continuous real-valued function on $\mathbb{R}^s$. Therefore (cf. Proposition 4.5 and the remark after it in [23]) $\mathcal{B}_{\mathcal{K}}$ possesses and abstract measurability property called $P\in$Suslin (consult [23], [62] for its definition).
These observations lead to the following result on the speed of the convergence asserted in Lemma 5.5.1 if we put $\mathcal{B}_o := \mathcal{B}_{\mathcal{K}}$ (law of iterated logarithm). It is a special case of Corollary 2.4 in [62].

Lemma 5.5.2 *It holds*

$$\limsup_{n\to\infty} \Big(\frac{n}{2\log\log n}\Big)^{1/2} \cdot \sup\{|\mu_n(\omega)(B) - \mu(B)| : B \in \mathcal{B}_{\mathcal{K}}\} \leq \frac{1}{2}$$

for $\mathbf{P}$*-almost all* $\omega \in \Omega$.

Employing the above preliminaries we obtain the following results:

Proposition 5.5.3 *Suppose* (A1), (A2), *let* $\mu \in \Delta_{p,K}(\mathbb{R}^s)$ *for some* $p > 1$, $K > 0$ *and assume that* $M \subset \mathbb{R}^m$ *is a CLM set for* $P(\mu)$ *with respect to some bounded open set* $V \in \mathbb{R}^m$. *Then it holds:*

(i) $\varphi_V(\mu_n(\omega)) \underset{n\to\infty}{\longrightarrow} \varphi_V(\mu)$ *for* $\mathbf{P}$*-almost all* $\omega \in \Omega$,

(ii) *for any open set* $\tilde{V} \subset \mathbb{R}^m$ *such that* $\psi_V(\mu) \subset \tilde{V}$ *and* $\mathbf{P}$*-almost all* $\omega \in \Omega$ *there exists an* $n_o(\omega) \in \mathbb{N}$ *such that* $\psi_V(\mu_n(\omega)) \subset \tilde{V}$ *for all* $n \geq n_o(\omega)$ *(upper semicontinuity of* ψ_V*),*

(iii) *for* $\mathbf{P}$*-almost all* $\omega \in \Omega$ *there exists an* $n_1(\omega) \in \mathbb{N}$ *such that for all* $n \geq n_1(\omega)$ *the set* $\psi_V(\mu_n(\omega))$ *is a CLM set for* $P(\mu_n(\omega))$ *with respect to* V.

Proof: To simplify the notation we often write μ_n instead of $\mu_n(\omega)$.
Since $C \cap \mathrm{cl}\, V$ is compact and $Q(., \mu_n)$, $Q(., \mu)$ are both lower semicontinuous on $\mathbb{R}^s$ (Proposition 4.3.1 in Chapter 4), the sets $\psi_V(\mu_n)$ and $\psi_V(\mu)$ are both non-empty. Let $x_n \in \psi_V(\mu_n)$ and $x \in \psi_V(\mu)$. As in (4.1) we obtain

$$(5.1) \qquad |\varphi_V(\mu_n) - \varphi_V(\mu)| \leq \sup_{x \in C\cap \mathrm{cl}\, V} |Q(x, \mu_n) - Q(x, \mu)|.$$

It holds (cf. e.g. [80], Corollary 2 in chapter 4.1)

$$\int_{\mathbb{R}^s} \|z\|^p \mu_n(\mathrm{d}z) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^s} \|z\|^p \mu(\mathrm{d}z) \quad \text{for } \mathbf{P}\text{-almost all } \omega \in \Omega. \tag{5.2}$$

Hence, for $\mathbf{P}$-almost all $\omega \in \Omega$ there exists an $n_3(\omega) \in \mathbb{N}$ such that $\mu_n(\omega) \in \Delta_{p,2K}(\mathbb{R}^s)$ for all $n \geq n_3(\omega)$, and Proposition 5.3.1 works in the present setting (with $\Delta_{p,2K}(\mathbb{R}^s)$ instead of $\Delta_{p,K}(\mathbb{R}^s)$). $\mathcal{B}_{\mathcal{K}}$ being a VČ class Lemma 5.5.1 and (5.1) together establish (i).

Since $C \cap \operatorname{cl} V$ is compact, upper semicontinuity of ψ_V is equivalent to closedness of ψ_V (cf. [4]), i.e. given $x_n \in \psi_V(\mu_n)$ such that $x_n \underset{n\to\infty}{\longrightarrow} \bar{x}$ we have to show that $\bar{x} \in \psi_V(\mu)$.

Let $\omega \in \Omega$ be such that (i), (5.2) and the assertion of Lemma 5.5.1 hold. Let $\varepsilon > 0$ be arbitrary and $n_4(\omega) \in \mathbb{N}$ such that for $n \geq n_4(\omega)$

$$g(\bar{x}) + Q(\bar{x}, \mu) \leq g(x_n) + Q(x_n, \mu) + \varepsilon/3 \tag{5.3}$$

$$\sup_{x \in C \cap \operatorname{cl} V} |Q(x, \mu_n) - Q(x, \mu)| \leq \varepsilon/3 \tag{5.4}$$

$$|\varphi_V(\mu_n) - \varphi_V(\mu)| \leq \varepsilon/3. \tag{5.5}$$

The validity of (5.3) follows from the lower semincontinuity of $Q(., \mu)$.

Now it holds for $n \geq n_4(\omega)$:

$$\begin{aligned} g(\bar{x}) + Q(\bar{x}, \mu) &\leq g(x_n) + Q(x_n, \mu) + \varepsilon/3 \\ &= g(x_n) + Q(x_n, \mu_n) - Q(x_n, \mu_n) + Q(x_n, \mu) + \varepsilon/3 \\ &\leq \varphi_V(\mu_n) + 2\varepsilon/3 \ \leq \ \varphi_V(\mu) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary this implies $\bar{x} \in \psi_V(\mu)$, and (ii) is shown.

To establish (iii) recall that $\psi_V(\mu_n(\omega))$ is non-empty due to the lower semicontinuity of $Q(., \mu_n)$ and the compactness of $C \cap \mathrm{cl} V$. The remaining part of the CLM property is a direct consequence of (ii). ■

Remark 5.5.4 *Compared to Proposition* 4.4.1 *in Chapter* 4 *it is interesting to note that the above result shows that in the context of estimation via empirical measures one can dispense with the smoothness assumption on* μ *when aiming at qualitative stability.*

Proposition 5.5.5 *Adopt the setting of Proposition* 5.5.3. *Then there exists a constant $c > 0$ such that for* $\mathbf{P}$*-almost all* $\omega \in \Omega$

$$\limsup_{n\to\infty} \Big(\frac{n}{2\log\log n}\Big)^{\frac{1}{2}\cdot\frac{p-1}{p(s+1)}} |\varphi_V(\mu_n(\omega)) - \varphi_V(\mu)| \le c.$$

Proof: By (5.2) we again have that $\mu_n(\omega) \in \Delta_{p,2K}(\mathbb{R}^s)$ for $\mathbf{P}$-almost all $\omega \in \Omega$ and $n \in \mathbb{N}$ sufficiently large. Since $\mathcal{B}_{\mathcal{K}}$ is a VČ class, Lemma 5.5.1 implies that for $\mathbf{P}$-almost all $\omega \in \Omega$ and $n \in \mathbb{N}$ sufficiently large $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu_n(\omega), \mu) < \delta$ where δ is taken according to Proposition 5.3.1 (with $\Delta_{p,2K}(\mathbb{R}^s)$ instead of $\Delta_{p,K}(\mathbb{R}^s)$). Hence, (5.1) and Proposition 5.3.1 yield

$$\begin{aligned}
&\Big(\frac{n}{2\log\log n}\Big)^{\frac{1}{2}\cdot\frac{p-1}{p(s+1)}} |\varphi_V(\mu_n(\omega)) - \varphi_V(\mu)| \\
&\qquad \le L \cdot \Big[\Big(\frac{n}{2\log\log n}\Big)^{1/2} \alpha(\mu_n(\omega), \mu)\Big]^{\frac{p-1}{p(s+1)}}
\end{aligned}$$

for $\mathbf{P}$-almost all $\omega \in \Omega$ and $n \in \mathbb{N}$ sufficiently large.
Taking the lim sup and employing Lemma 5.5.2 yields the assertion with
$c := L \cdot (\frac{1}{2})^{\frac{p-1}{p(s+1)}}$. ∎

Remark 5.5.6 *If* $\mathcal{B}_o \subset \mathcal{B}(\mathbb{R}^s)$ *is a VČ class, then there exists a function* $\pi : \mathbb{R}_+ \to \mathbb{R}_+$ *growing at most polynomially such that for any* $\varepsilon > 0$ *and all sufficiently large* $n \in \mathbb{N}$

$$P(\{\omega \in \Omega : \alpha_{\mathcal{B}_o}(\mu_n(\omega), \mu) \ge \varepsilon\}) \le 4\pi(2n)\exp(-n\varepsilon^2/8)$$

(cf. [117], *p.* 829*). Combining this result with Proposition* 5.4.1 *yields an upper estimate for* $P(\{\omega \in \Omega : |\varphi_V(\mu_n(\omega)) - \varphi_V(\mu)| \ge \varepsilon\})$.

Acknowledgement
I wish to thank Professors S.T. Rachev (University of California, Santa Barbara), W. Römisch (Humboldt-Universität Berlin) and W. Stute (Justus-Liebig-Universität Gießen) for beneficial discussions.

Bibliography

[1] Artstein, Z.; Wets, R.J-B.: Stability of stochastic programs with possibly discontinuous objective functions, SIAM Journal on Optimization (to appear).

[2] Attouch, H.; Wets, R.J-B: Quantitative stability of variational systems: I. The epigraphical distance, Transactions of the American Mathematical Society 328 (1991), 695-729.

[3] Attouch, H.; Wets, R.J-B: Quantitative stability of variational systems: II. A framework for nonlinear conditioning, SIAM Journal on Optimization 3 (1993), 359-381.

[4] Bank, B.; Guddat, J.; Klatte, D.; Kummer, B.; Tammer, K.: Nonlinear Parametric Optimization, Akademie-Verlag, Berlin 1982.

[5] Bank, B.; Mandel, R.: Parametric Integer Optimization, Akademie-Verlag, Berlin 1988.

[6] Beale, E.M.: On minimizing a convex function subject to linear inequalities, Journal of the Royal Statistical Society B (1955), 173-184.

[7] Berge, C.: Espaces topologiques, fonctions multivoques, Dunod, Paris 1959.

[8] Bhattacharya, R.N.; Ranga Rao, R.: Normal Approximation and Asymptotic Expansions, Wiley, New York 1976.

[9] Billingsley, P.: Convergence of Probability Measures, Wiley, New York 1968.

[10] Birge, J.R., Wets, R.J-B: Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse, Mathematical Programming Study 27 (1986), 54-102.

[11] Birge, J.R.; Wets, R.J-B (Eds.): Stochastic Programming, Parts I, II, Annals of Operations Research 30-31 (1991).

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

[13] Borell, C.: Convex set functions in d-space, Periodica Mathematica Hungarica 6 (1975), 111-136.

[14] Charnes, A.; Cooper, W.W.: Chance constrained programming, Management Science 5 (1959), 197-207.

[15] Clarke, F.H.: Optimization and Nonsmooth Analysis, Wiley, New York 1983.

[16] Collatz, L.; Wetterling, W.: Optimization Problems, Springer-Verlag, New York 1975.

[17] Dancs, S.; Uhrin, B.: On a class of integral inequalities and their measure-theoretic consequences, Journal of Mathematical Analysis and Applications 74 (1980), 388-400.

[18] Dantzig, G.B.: Linear programming under uncertainty, Management Science 1 (1955), 197-206.

[19] Deák, I.: Multidimensional integration and stochastic programming, Numerical Techniques for Stochastic Optimization (Y. Ermoliev, R.J-B Wets, Eds.), Springer-Verlag, Berlin 1988, 187-200.

[20] Dempster, M.A.H. (Ed.): Stochastic Programming, Academic Press, New York, 1980.

[21] Dempster, M.A.H.: On stochastic programming II: Dynamic problems under risk, Stochastics 25 (1988), 15-42.

[22] Dudley, R.M.: Probabilities and Metrics, Lecture Notes Series No. 45, Aarhus Universitet, Aarhus 1976.

[23] Dudley, R.M.: Central limit theorems for empirical measures, Annals of Probability 6 (1978), 899-929.

[24] Dudley, R.M.: Real Analysis and Probability, Wadsworth & Brooks/Cole, Pacific Grove, California 1989.

[25] Dupačová, J.: Stability in stochastic programming with recourse - estimated parameters, Mathematical Programming 28 (1984), 72-83.

[26] Dupačová, J.: Stability in stochastic programming - probabilistic constraints, Stochastic Optimization (V.I. Arkin, A. Shiraev, R.J-B Wets, Eds.), Springer-Verlag, Berlin 1986, 314-325.

[27] Dupačová, J.: Stochastic programming with incomplete information: a survey of results on postoptimization and sensitivity analysis, Optimization 18 (1987), 507-532.

[28] Dupačová, J.: Stability and sensitivity analysis for stochastic programming, Annals of Operations Research 27 (1990), 115-142.

[29] Dupačová, J.; Wets, R.J-B: Asymptotic behaviour of statistical estimators and of optimal solutions of stochastic optimization problems, The Annals of Statistics 16 (1988), 1517-1549.

[30] Eisner, M.J.; Olsen, P.: Duality for stochastic programming interpreted as L.P. in L_p-space, SIAM Journal Applied Mathematics 28 (1975), 779-792.

[31] Ermoliev, Y.M.; Norkin, V.I.: Normalized convergence in stochastic optimization, Annals of Operations Research 30 (1991), 187-198.

[32] Ermoliev, Y.M.; Wets, R.J-B (Eds.): Numerical Techniques for Stochastic Optimization, Springer-Verlag, Berlin, 1988.

[33] Evstigneev, I.V.: Measurable selection and dynamic programming, Mathematics of Operations Research 1 (1976), 267-272.

[34] Fiacco, A.V.: Introduction to Sensitivity and Stability Analysis in Nonlinear Programming, Academic Press, New York, 1983.

[35] Fiedler, O.; Römisch, W.: Stability in Multistage Stochastic Programming, Schwerpunktprogramm der Deutschen Forschungsgemeinschaft "Anwendungsbezogene Optimierung und Steuerung", Report No. 426 (1993).

[36] Flåm, S.D.; Schultz, R.: A new approach to stochastic linear programming, Numerical Functional Analysis and Optimization 14 (1993), 545-554.

[37] Frauendorfer, K.: Stochastic Two-Stage Programming, Lecture Notes in Economics and Mathematical Systems 392, Springer-Verlag, Berlin 1992.

[38] Gänssler, P.; Stute, W.: Wahrscheinlichkeitstheorie, Springer-Verlag, Berlin 1977.

[39] Gänssler, P.; Stute, W.: Empirical processes: A survey of results for independent and identically distributed random variables, Annals of Probability 7 (1979), 193-243.

[40] Garstka, S.J.: Distribution functions in stochastic programs with recourse: A parametric analysis, Mathematical Programming 6 (1974), 339-351.

[41] Guddat, J.: Stability in convex quadratic parametric programming, Mathematische Operationsforschung und Statistik 7 (1976), 223-245.

[42] Higle, J.L.; Sen, S.: Stochastic decomposition: an algorithm for two-stage linear programs with recourse, Mathematics of Operations Research 16 (1991), 650-669.

[43] Higle, J.L.; Sen, S.; Yakowitz, D.S.: Finite master programs in stochastic decomposition, Technical Report, SIE Department, University of Tucson, Arizona 1990.

[44] Hoffman, A.J.: On approximate solutions of systems of linear inequalities, Journal Research National Bureau of Standards 49 (1952), 263-265.

[45] Kall, P.: Approximations to stochastic programs with complete fixed recourse, Numerische Mathematik 22 (1974), 333-339.

[46] Kall, P.: Stochastic Linear Programming, Springer-Verlag, Berlin, 1976.

[47] Kall, P.; Stoyan, D.: Computational methods for solving two-stage stochastic linear programming problems with recourse including error bounds, Mathematische Operationsforschung und Statistik, Series Optimization 13 (1982), 431-447.

[48] Kall, P.: On approximations and stability in stochastic programming, Parametric Optimization and Related Topics (J. Guddat, H.Th. Jongen, B. Kummer, F. Nožička, Eds.), Akademie Verlag, Berlin 1987, 387-407.

[49] Kall, P.; Ruszczyński, A.; Frauendorfer, K.: Approximation techniques in stochastic programming, Numerical Techniques for Stochastic Optimization (Y. Ermoliev, R.J-B Wets, Eds.), Springer-Verlag, Berlin 1988, 33-64.

[50] Kaniovski, Y.M.; King, A.J.; Wets, R.J-B: Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems, IBM Research Division, Research Report, RC 18752, 1993.

[51] Kiefer, J.: On large deviations of the empiric D.F. of vector chance variables and a law of the iterated logarithm, Pacific Journal of Mathematics 11 (1961), 649-660.

[52] Kiefer, J.; Wolfowitz, J.: On the deviations of the empiric distribution function of vector chance variables, Transactions American Mathematical Society 87 (1958), 173-186.

[53] King, A.J.; Rockafellar, R.T.: Asymptotic theory for solutions in statistical estimation and stochastic programming, Mathematics of Operations Research 18 (1993), 148-162.

[54] King, A.J.; Wets, R.J-B: Epi-consistency of convex stochastic programs, Stochastics and Stochastics Reports 34 (1991), 83-92.

[55] Klatte, D.: Beiträge zur Stabilitätsanalyse nichtlinearer Optimierungsprobleme, Dissertation (B), Humboldt-Universität Berlin, Sektion Mathematik, 1984.

[56] Klatte, D.: On the Lipschitz behavior of optimal solutions in parametric problems of quadratic optimization and linear complementarity, Optimization 16 (1985), 819-831.

[57] Klatte, D.: A note on quantitative stability results in nonlinear optimization, Proceedings of the 19. Jahrestagung Mathematische Optimierung (K. Lommatzsch, Ed.) Seminarbericht Nr. 90, Humboldt-Univ. Berlin, Sektion Mathematik, 1987, 77-86.

[58] Klatte, D.: On quantitative stability for non-isolated minima, Control and Cybernetics (to appear).

[59] Klatte, D.; Kummer, B.: Stability properties of infima and optimal solutions of parametric optimization problems, Nondifferentiable Optimization: Motivations and Applications (V. Demyanov, D. Pallaschke, Eds.), Springer-Verlag, Berlin 1985, 215-229.

[60] Klein Haneveld, W.K; van der Vlerk, M.H.: On the expected value function of a simple integer recourse problem with random technology matrix, Journal of Computational and Applied Mathematics 56 (1994) (to appear).

[61] Klein Haneveld, W.K.; Stougie, L.; van der Vlerk, M.H.: On the convex hull of the simple integer recourse objective function, Research Memorandum 516, Institute of Economic Research, University of Groningen, 1993.

[62] Kuelbs, J.; Dudley, R.M.: Log log laws for empirical measures, Annals of Probability 8 (1980), 405-418.

[63] Kummer, B.: Generalized equations: Solvability and regularity, Mathematical Programming Study 21 (1984), 199-212.

[64] Langen, H.: Convergence of dynamic programming models, Mathematics of Operations Research 6 (1981), 493-512.

[65] Laporte, G.; Louveaux, F.V.: The integer L-shaped method for stochastic integer programs with complete recourse, Operations Research Letters 13 (1993), 133-142.

[66] Louveaux, F.V.; van der Vlerk, M.H.: Stochastic programs with simple integer recourse, Mathematical Programming 61 (1993), 301-326.

[67] Marti, K.: Konvexitätsaussagen zum linearen stochastischen Optimierungsproblem, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 18 (1971), 159-166.

[68] Marti, K.: Approximationen der Entscheidungsprobleme mit linearer Ergebnisfunktion und positiv homogener, subadditiver Verlustfunktion, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 31 (1975), 203-233.

[69] Marti, K. (Ed.): Stochastic Optimization, Numerical Methods and Technical Applications, Springer-Verlag, Lecture Notes in Economics and Mathematical Systems 379, 1992.

[70] Mayer, J.: Computational techniques for probabilistic constrained optimization problems, Stochastic Optimization, Numerical Methods and Technical Applications, (K. Marti, Ed.), Springer-Verlag, Lecture Notes in Economics and Mathematical Systems 379, 1992, 141-164.

[71] Nemhauser, G.L.; Wolsey, L.A.: Integer and Combinatorial Optimization, Wiley, New York 1988.

[72] Nožička, F.; Guddat, J.; Hollatz, H.; Bank, B.: Theorie der linearen parametrischen Optimierung, Akademie Verlag, Berlin 1974.

[73] Olsen, P.: Multistage stochastic programming with recourse as mathematical programming in an L_p-space, SIAM Journal on Control and Optimization 14 (1976), 111-124.

[74] Ortega, J.M.; Rheinboldt, W.C.: Iterative solution of nonlinear equations in several variables, Academic Press, New York 1970.

[75] Poljak, B.: Existence theorems and convergence of minimizing sequences in extremum problems with restrictions, Soviet Mathematics Doklady 7 (1966), 72-75.

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

[77] Prékopa, A.: Logarithmic concave measures with applications to stochastic programming, Acta Scientiarum Mathematicarum 32 (1971), 301-316.

[78] Prékopa, A.: Logarithmic concave measures and related topics, Stochastic Programming (M.A.H. Dempster, Ed.), Academic Press, New York 1980, 63-82.

[79] Prékopa, A.: Numerical solution of probabilistic constrained programming problems, Numerical Techniques for Stochastic Optimization (Y. Ermoliev, R.J-B Wets, Eds.), Springer-Verlag, Berlin 1988, 123-139.

[80] Rachev, S.T.: The Monge-Kantorovich mass transference problem and its stochastic applications, Theory of Probability and its Applications 29 (1984), 647-676.

[81] Rachev, S.T.: Probability Metrics and the Stability of Stochastic Models, Wiley, New York 1991.

[82] Richter, H.: Das Gesetz vom iterierten Logarithmus für empirische Verteilungsfunktionen im $\mathbb{R}^k$, Manuscripta mathematicae 11 (1974), 291-303.

[83] Rinnooy Kan, A.; Stougie, L.: Stochastic integer programming, Numerical Techniques for Stochastic Optimization (Y. Ermoliev, R.J-B Wets, Eds.) Springer-Verlag, Berlin 1988, 201-213.

[84] Rinott, Y.: On convexity of measures, Annals of Probability 4 (1976), 1020-1026.

[85] Robinson, S.M.: Some continuity properties of polyhedral multifunctions, Mathematical Programming Study 14 (1981), 206-214.

[86] Robinson, S.M.: Local epi-continuity and local optimization, Mathematical Programming 37 (1987), 208-222.

[87] Robinson, S.M.: Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity, Mathematical Programming Study 30 (1987), 45-66.

[88] Robinson, S.M.; Wets, R.J-B: Stability in two-stage stochastic programming, SIAM Journal on Control and Optimization 25 (1987), 1409-1416.

[89] Rockafellar, R.T.: Convex Analysis, Princeton University Press, Princeton 1970.

[90] Rockafellar, R.T.: Integral functionals, normal integrands and measurable selections, Nonlinear Operators and the Calculus of Variations (G.P. Gossez et al. Eds.), Lecture Notes in Mathematics, Vol 543, Springer-Verlag, New York 1976, 157-207.

[91] Rockafellar, R.T.; Wets, R.J-B: Nonanticipativity and L^1-martingales in stochastic optimization problems, Mathematical Programming Study 6 (1976), 170-187.

[92] Römisch, W.; Schultz, R.: Distribution sensitivity in stochastic programming, Mathematical Programming 50 (1991), 197-226.

[93] Römisch, W.; Schultz, R.: Stability analysis for stochastic programs, Annals of Operations Research 30 (1991), 241-266.

[94] Römisch, W.; Schultz, R.: Distribution sensitivity for certain classes of chance-constrained models with application to power dispatch, Journal of Optimization Theory and Applications 71 (1991), 569-588.

[95] Römisch, W.; Schultz, R.: Stability of solutions for stochastic programs with complete recourse, Mathematics of Operations Research 18 (1993), 590-609.

[96] Römisch, W.; Schultz, R.: Strong convexity and directional derivatives of marginal values in two-stage stochastic programming, Stochastic Optimization - Numerical Methods and Technical Applications II, (K. Marti, Ed.),

Springer-Verlag, Lecture Notes in Economics and Mathematical Systems (submitted).

[97] [*Chapter 3*] Römisch, W.; Schultz, R.: Lipschitz stability for stochastic programs with complete recourse, SIAM Journal on Optimization (submitted).

[98] Römisch, W.; Wakolbinger, A.: Obtaining convergence rates for approximations in stochastic programming, Parametric Optimization and Related Topics (J. Guddat, H.Th. Jongen, B. Kummer, F. Nožička, Eds.), Akademie Verlag, Berlin 1987, 327-343.

[99] Ruszczyński, A.: A regularized decomposition method for minimizing a sum of polyhedral functions, Mathematical Programming 35 (1986), 309-333.

[100] Ruszczyński, A.: A linearization method for nonsmooth stochastic programming problems, Mathematics of Operations Research 12 (1987), 32-49.

[101] Ruszczyński, A.: Interior point methods in stochastic programming, Working Paper WP-93-8, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1993.

[102] Ruszczyński, A.: Regularized decomposition of stochastic programs: algorithmic techniques and numerical results, Working Paper WP-93-21, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1993.

[103] Salinetti, G.: Approximations for chance-constrained programming problems, Stochastics 10 (1983), 157-179.

[104] [*Chapter 4*] Schultz, R.: Continuity properties of expectation functions in stochastic integer programming, Mathematics of Operations Research 18 (1993), 578-589.

[105] [*Chapter 2*] Schultz, R.: Strong convexity in stochastic programs with complete recourse, Journal of Computational and Applied Mathematics 56 (1994) (to appear).

[106] [*Chapter 4*] Schultz, R.: On structure and stability in stochastic programs with random technology matrix and complete integer recourse, Mathematical Programming (submitted).

[107] [*Chapter 5*] Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse, Schwerpunktprogramm der Deutschen Forschungsgemeinschaft "Anwendungsbezogene Optimierung und Steuerung", Report No. 483 (1993).

[108] Schultz, R.; Stougie, L.; van der Vlerk, M.H.: Two-stage stochastic integer programming: a survey, Statistica Neerlandica (submitted).

[109] Sen, S.; Doverspike, R.D.; Cosares, S.: Network planning with random demand, Bell Communications Research Tech. Memo, TM-OPT-022142, Red Bank 1992.

[110] Serfling, R.J.: Approximation Theorems of Mathematical Statistics, Wiley, New York 1980.

[111] Shapiro, A.: Asymptotic properties of statistical estimators in stochastic programming, The Annals of Statistics 17 (1989), 841-858.

[112] Shapiro, A.: On differential stability in stochastic programming, Mathematical Programming 47 (1990), 107-116.

[113] Shapiro, A.: Asymptotic analysis of stochastic programs, Annals of Operations Research 30 (1991), 169-186.

[114] Shapiro, A.: Perturbation analysis of optimization problems in Banach spaces, Numerical Functional Analysis and Optimization 13 (1992), 97-116.

[115] Shapiro, A.: Asymptotic behaviour of optimal solutions in stochastic programming, Mathematics of Operations Research 18 (1993), 829-845.

[116] Shapiro, A.: Quantitative stability in stochastic programming, Mathematical Programming (to appear).

[117] Shorack, G.R.; Wellner, J.A.: Empirical Processes with Applications to Statistics, Wiley, New York 1986.

[118] Stougie, L.: Design and analysis of algorithms for stochastic integer programming, Ph.D.Thesis, Center for Mathematics and Computer Science, Amsterdam 1985.

[119] Szantai, T.: Evaluation of a special multivariate gamma distribution function, Mathematical Programming Study 27 (1986), 1-16.

[120] Tintner, G.: Stochastic linear programming with applications to agricultural economics, Proceedings of the Second Symposium in Linear Programming (H.A. Antosiewicz, Ed.), Washington (1955), 197-207.

[121] Uryas'ev, S.P.: A differentiation formula for integrals over sets given by inclusion, Numerical Functional Analysis and Optimization 10 (1989), 827-841.

[122] Vapnik, V.N.; Červonenkis, A.V.: On the uniform convergence of relative frequencies of events to their probabilities, Theory of Probability and Applications 16 (1971), 264-280.

[123] Vasil'ev, F.P.: Numerical Methods for the Solution of Extremal Problems (in Russian), Moscow, Nauka 1988.

[124] Vogel, S.: Stability results for stochastic programming problems, Optimization 19 (1988), 269-288.

[125] Walkup, D.W.; Wets, R.J-B: A Lipschitzian characterization of convex polyhedra, Proceedings of the American Mathematical Society 23 (1969), 167-173.

[126] Walkup, D.W.; Wets, R.J-B: Lifting projections of convex polyhedra, Pacific Journal of Mathematics 28 (1969), 465-475.

[127] Wallace, S.: A piecewise linear upper bound on the network recourse function, Mathematical Programming 38 (1987), 133-146.

[128] Wallace, S.: A two-stage stochastic facility location problem, Numerical Techniques for Stochastic Optimization (Y. Ermoliev, R.J-B Wets, Eds.) Springer-Verlag, Berlin 1988, 489-514.

[129] Wang, J.: Distribution sensitivity analysis for stochastic programs with complete recourse, Mathematical Programming 31 (1985), 286-297.

[130] Wets, R.J-B: Stochastic programs with recourse: a basic theorem for multistage problems, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 21 (1972), 201-206.

[131] Wets, R.J-B: Stochastic programs with fixed recourse: the equivalent deterministic program, SIAM Review 16 (1974), 309-339.

[132] Wets, R.J-B: Stochastic programming: solution techniques and approximation schemes, Mathematical Programming - The State-of-the-Art 1982

(A. Bachem, M. Grötschel, B. Korte, Eds.), Springer-Verlag, Berlin 1983, 566-603.

[133] Wets, R.J-B: Stochastic programming, Handbooks in Operations Research and Management Science, Vol.1: Optimization (G.L. Nemhauser, A.H.G. Rinnooy Kan, M.J. Todd, Eds.), North-Holland, 1989, 573-629.

[134] Wets, R.J-B: Constrained estimation: Consistency and asymptotics, Applied Stochastic Models and Data Analysis 7 (1991), 17-32.

[135] Wollmer, R.M.: Two-stage linear programming under uncertainty with 0-1 integer first stage variables, Mathematical Programming 19 (1980), 279-288.