SIAM J. OPTIMIZATION
Vol. 6, No. 2, pp. 531–547, May 1996

014

LIPSCHITZ STABILITY FOR STOCHASTIC PROGRAMS WITH COMPLETE RECOURSE*

WERNER RÖMISCH† AND RÜDIGER SCHULTZ‡

Abstract. This paper 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 [*Math. Programming*, 67 (1994), pp. 99–108] Shapiro 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 behavior 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).

Key words. stochastic programs with recourse, Lipschitz stability, empirical distributions, asymptotic analysis

AMS subject classifications. 90C15, 90C31

1. Introduction. We study quantitative stability and asymptotic properties (of estimates via empirical measures) of optimal solutions to stochastic programs with complete recourse. The latter are given by

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

where

$$Q_\mu(\zeta) = \int_{I\!R^s} Q(z-\zeta)\mu(\mathrm{d}z) \tag{1.2}$$

and

$$Q(t) = \min\{q^T v : Wv = t, \, v \geq 0\}. \tag{1.3}$$

For the data we assume that $g : I\!R^m \to I\!R$ is a convex function, $C \subset I\!R^m$ is a nonempty 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 which we assume will vary in $\mathcal{M}_1(I\!R^s)$—the space of all Borel probability measures on $I\!R^s$ with finite first moment, i.e., $\int_{I\!R^s} \|z\|\mu(\mathrm{d}z) < +\infty$ for all $\mu \in \mathcal{M}_1(I\!R^s)$. Further assumptions that ensure (1.1)–(1.3) to be well defined will be given in §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)

* Received by the editors April 2, 1993; accepted for publication (in revised form) October 10, 1994. This research was supported by the Schwerpunktprogramm “Anwendungsbezogene Optimierung und Steuerung” of the Deutsche Forschungsgemeinschaft.

† Humboldt-Universität zu Berlin, Institut für Mathematik, Unter den Linden 6, 10099 Berlin, Germany (romisch@mathematik.hu-berlin.de).

‡ Konrad-Zuse-Zentrum für Informationstechnik Berlin, Heilbronner Strasse 10, 10711 Berlin, Germany (schultz@zib-berlin.de).

decision y. For a more detailed introduction to this class of models, including a basic analysis of the function Q_μ in (1.2), we refer to [6], [31]. Here we only mention that Q_μ is convex whenever it is well defined.

In the present paper, we concentrate on studying the impact of changes in the underlying probability measure μ on the problem (1.1). To this end, we assign to $\mu \in \mathcal{M}_1(\mathbb{R}^s)$ 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 setup (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 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 [3], [4], [7], [12], [19], [22], [26], [28], [30], [32], [33].

The subsequent analysis is entirely concerned with quantitative continuity properties of the optimal set mapping ψ. As in our earlier work [22], [23], 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 first mention that the example does fit the setting of our central stability estimate; in particular, the function Q_μ is here strongly convex on a suitable subset (cf. Theorem 2.4).

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

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

Our investigations have been stimulated by recent results of Shapiro. In [29] 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 [2] of the function $Q_\nu - Q_\mu$ (cf. (1.2)) at points belonging to a suitable neighbourhood. Here Q_ν corresponds to the perturbed problem $P(\nu)$, $\nu \in \mathcal{M}_1(\mathbb{R}^s)$.

In the present paper we introduce a "subgradient distance" for $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^s)$ based on the above maximal norm (cf. (2.1)). 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 nonempty polyhedron, and Q_μ is strongly convex on a suitable neighbourhood of $A(\psi(\mu))$. Then the Lipschitz upper semicontinuity extends to the Lipschitz continuity of the Hausdorff distance of solution sets (Theorem 2.4). Since the subgradient distance of $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^s)$ 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 2.5, Remark 2.6), this leads to a powerful tool for quantitative statements on the stability of optimal solutions. In §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, previous results are extended to the case where solution sets are not necessarily singletons. No additional assumptions on the underlying probability measure μ are required for the large deviation result.

Some further propositions and examples supplement and illustrate the main issue of the paper. In Proposition 2.3 the persistence of optimal solutions under perturbations in the "subgradient distance" is addressed. Proposition 2.15 displays some handy conclusions for the special case of "simple recourse." Examples in §2 show that Shapiro's assumptions in [29] do not guarantee the lower semicontinuity of ψ (Example 2.11), that Theorem 2.4 is lost for general convex g and C (Example 2.9), and that the setting of Theorem 2.4 does not guarantee stability of the optimal value (Example 2.12).

Compared to Shapiro's analysis in [29], we confine ourselves to a specific problem class for which we eventually obtain stronger results (Lipschitz continuity instead of Lipschitz upper semicontinuity). Our stability conditions can be verified comparatively easily (strong convexity of Q_μ, to be verified using Proposition 2.13), and in fact they turn out to be sufficient for Shapiro's second-order growth condition to hold. Moreover, the persistence established in Proposition 2.3 further develops the results in [29].

Compared to [22], [23], where the stability analysis is based on the L_1-Wasserstein distance and where Hölder continuity (with exponent 1/2) is obtained, the present paper leads to Lipschitz continuity. To illustrate our improvements over previous results, let us first mention that there exist sequences of measures where the results from both [22], [23] and the present paper lead to the same convergence rates (cf. the discussion after Proposition 2.13). On the other hand, there are important specific modes of perturbation (contaminated distributions, empirical measures) where the Hölder result in [22], [23] yields the rate 1/2, whereas the present approach leads to the rate 1 (Proposition 2.14, §3).

2. Stability.

The following basic assumptions are well known to ensure that the function Q_μ in (1.2) is well defined and convex on $I\!R^m$ (cf. [6], [31]):

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

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

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

Here "∂" denotes Clarke's subdifferential [2]. Since both Q_ν and Q_μ 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 pseudometric on $\mathcal{M}_1(I\!R^s)$. 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 [2]). The following lemma provides some more insight into d.

LEMMA 2.1. *Let $h_1, h_2 : I\!R^s \to I\!R$ be locally Lipschitzian. Then it holds for arbitrary $\zeta \in I\!R^s$ that $d_H(\partial h_1(\zeta), \partial h_2(\zeta)) \leq \sup\{\|z^*\| : z^* \in \partial(h_1 - h_2)(\zeta)\}$, where d_H denotes the Hausdorff distance of sets.*

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

$$\partial h_i(\zeta) = \text{conv}\, \mathcal{L}_{h_i}(\zeta),$$

where

$$\mathcal{L}_{h_i}(\zeta) = \{z : \text{ there exist } \zeta_n \in \text{Diff}(h_1) \cap \text{Diff}(h_2) \text{ such that } \zeta_n \to \zeta \text{ and } h_i'(\zeta_n) \to z \text{ as } n \to \infty\}.$$

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

Assume that

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

for some $\zeta \in I\!R^s$.

This implies

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

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

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

for all $z_2^* \in \mathcal{L}_{h_2}(\zeta)$.

Since $z_{1,0}^* \in \mathcal{L}_{h_1}(\zeta)$, there exists a sequence of points $\zeta_n \in \text{Diff}(h_1) \cap \text{Diff}(h_2)$ such that $h_1'(\zeta_n) \to z_{1,0}^*$ as $n \to \infty$.

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

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

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

In our quantitative stability analysis for optimal solutions of perturbed stochastic programs, d will be the distance that measures "how far" away a perturbation $P(\nu)$ is from the original program $P(\mu)$. In the context of stochastic programming Shapiro [29] has also used information contained in the definition of d to derive quantitative stability properties. Kummer [15] has obtained results on the quantitative stability of general convex programs based on the Hausdorff distance of subgradients, which appears in Lemma 2.1. Our considerations start with the following result by Shapiro [29].

THEOREM 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 $g(x) + Q_\mu(Ax) \geq \varphi(\mu) + \alpha \cdot \text{dist}(x, \psi(\mu))^2$ 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 $\nu \in \mathcal{M}_1(\mathbb{R}^s)$*:*

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

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

Theorem 2.2 asserts the upper Lipschitz continuity of the solution set mapping ψ with respect to the pseudometric d. It does not contain the persistence of optimal solutions, i.e., it is not clear whether the perturbed program $P(\nu)$ has a nonempty set of optimal solutions if $d(\mu, \nu; \operatorname{cl} U_o)$ is sufficiently small. The next proposition answers this question.

PROPOSITION 2.3. *Suppose* (A1)–(A3) *and that* $\psi(\mu)$ *is nonempty and bounded. Let* $U_o \subset \mathbb{R}^m$ *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 $\nu \in \mathcal{M}_1(\mathbb{R}^s)$ *such that* $d(\mu, \nu; \operatorname{cl} U_o) < \delta$.

Proof. We introduce the following notation:

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

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 [1] imply that there exists a constant $\delta' > 0$ such that

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

To apply results on the stability of certain generalized equations [15], we introduce the set-valued mappings $\Gamma_\nu : \operatorname{cl} U_o \to \mathbb{R}^m$, $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, 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 [20].

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 [15] are fulfilled. Proposition 3 in [15] 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 $B_m \subset \mathbb{R}^m$ denotes the closed unit ball.

For the definition of admissibility we refer to [15]. 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 $\nu \in \mathcal{M}_1(\mathbb{R}^s)$ such that $d(\mu, \nu; \mathrm{cl}\, U_o) < \tilde{\delta}$. Lemma 2.1 now implies that

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

By the regularity of Γ_μ, this yields that $\psi(\nu)$ is nonempty whenever $d(\mu, \nu; \mathrm{cl}\, U_o) < \tilde{\delta}$. Now select $\delta > 0$ such that $\delta < \min\{\delta', \tilde{\delta}\}$. Let $\nu \in \mathcal{M}_1(\mathbb{R}^s)$ such that $d(\mu, \nu; \mathrm{cl}\, U_o) < \delta$. Then it holds that $\psi(\nu) \cap \mathrm{cl}\, U_o \neq \emptyset$. 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 that

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

Hence, there exists a $\tilde{d} \in \mathbb{R}^m$, $\|\tilde{d}\| < \delta'$ 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 2.2 extends to the Lipschitz continuity of ψ with respect to the Hausdorff distance of sets and the pseudometric d of probability measures.

While Shapiro derived Theorem 2.2 via a general variational principle, we impose an additional structure which enables us to use more specific techniques leading to Lipschitz continuity instead of Lipschitz upper semicontinuity. Our techniques combine estimates for strongly convex functions with Lipschitz continuity results for optimal solutions to perturbed (convex) quadratic programs.

THEOREM 2.4. *Suppose* (A1)–(A3) *and that $\psi(\mu)$ is nonempty and bounded. Let g be convex quadratic and C be convex polyhedral. Assume that there exists a convex open subset V of $\mathbb{R}^s$ such that $A(\psi(\mu)) \subset V$ and the function Q_μ 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 $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $d(\mu, \nu; U) < \delta$.

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

$$Q_\mu(\lambda\zeta + (1-\lambda)\tilde{\zeta}) \leq \lambda Q_\mu(\zeta) + (1-\lambda) Q_\mu(\tilde{\zeta}) - \kappa\lambda(1-\lambda)\|\zeta - \tilde{\zeta}\|^2.$$

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 $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $d(\mu, \nu; U) < \delta$ (Proposition 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 $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $d(\mu, \nu; U) < \delta$ that

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

Furthermore, the compactness of C_o guarantees

$$\begin{aligned} \min_x\{g(x) + Q_\nu(Ax) : x \in C_o\} &= \min_{x,y}\{g(x) + Q_\nu(y) : Ax = y,\ x \in C_o\} \\ &= \min_y\{Q_\nu(y) + \min_x\{g(x) : Ax = y,\ x \in C_o\} : y \in A(C_o)\}. \end{aligned}$$

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

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

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

The multifunction $\bar{X}(\cdot)$ is Lipschitzian on its effective domain $\operatorname{dom} \bar{X} = \{y \in \mathbb{R}^s : \bar{X}(y) \neq \emptyset\}$ (Satz 4.3.3 in [13], Theorem 4.2 in [14]). Therefore, there exists a constant $L_o > 0$ such that

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

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

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

Decrease, if necessary, $\delta > 0$ such that $\psi(\nu) \subset U_o$ whenever $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $d(\mu, \nu; U) < \delta$ (Proposition 2.3). Let $y \in \bar{Y}(\nu)$. Then $\bar{X}(y) \subset \psi(\nu)$. Since $\psi(\nu) \subset U_o$ and $\{y\} = A\bar{X}(y)$, it follows that $y \in A(U_o) \subset V$. Consider the point $\frac{1}{2}(y + y_\mu)$ belonging to $A(C_o) \cap V$.

Then

$$\begin{aligned} G(y_\mu, \mu) &\leq G\left(\frac{1}{2}(y + y_\mu), \mu\right) \\ &\leq \frac{1}{2} G(y, \mu) + \frac{1}{2} G(y_\mu, \mu) - \frac{\kappa}{4} \|y - y_\mu\|^2 \end{aligned}$$

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

This implies (if $\|y - y_\mu\| \neq 0$)

$$\begin{aligned} \frac{\kappa}{2} \|y - y_\mu\| &\leq \frac{G(y, \mu) - G(y_\mu, \mu)}{\|y - y_\mu\|} \\ &= \frac{(G(y_\mu, \nu) - G(y_\mu, \mu)) - (G(y_\mu, \nu) - G(y, \mu))}{\|y - y_\mu\|} \\ &\leq \frac{(G(y_\mu, \nu) - G(y_\mu, \mu)) - (G(y, \nu) - G(y, \mu))}{\|y - y_\mu\|} \\ &\leq \frac{|(Q_\nu - Q_\mu)(y_\mu) - (Q_\nu - Q_\mu)(y)|}{\|y - y_\mu\|}. \end{aligned}$$

By Lebourg's mean value theorem [2] there exists a point y^* on the line segment $[y_\mu, y]$ (belonging entirely to $A(U)$) such that the above estimate continues

$$\leq \sup \left| \left\langle \partial(Q_\nu - Q_\mu)(y^*), \frac{y_\mu - y}{\|y_\mu - y\|} \right\rangle \right|.$$

Hence,

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

Together with (2.5) this completes the proof. □

COROLLARY 2.5. *Adopt the setting of Theorem* 2.4. *Then there exist nonsingular 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 $\nu \in \mathcal{M}_1(I\!R^s)$ *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 [2] we obtain the following representation of $d(\mu, \nu; U)$:

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

where E contains those $x \in I\!R^m$ such that $Q_\nu - Q_\mu$ is not differentiable at Ax and $A(E)$ has Lebesgue measure zero.

Recall that the integrand Q in (1.2) is a piecewise linear convex function on $I\!R^s$ and that there exist basis submatrices $B_1, \ldots, B_\ell$ of W such that the simplicial cones $B_1(I\!R^s_+); \ldots, B_\ell(I\!R^s_+)$ are linearity regions of Q (in general not the maximal ones) (cf. [6], [16], Satz 6.7). Of course, $\bigcup_{i=1}^{\ell} B_i(I\!R^s_+) = I\!R^s$, and B_i $(i = 1, \ldots, \ell)$ can be chosen in such a way that $B_i(I\!R^s_+) \cap B_j(I\!R^s_+)$, $i \neq j$, is always contained in a hyperplane in $I\!R^s$. Thus, $\mathcal{H} := I\!R^s \setminus \bigcup_{i=1}^{\ell} \operatorname{int} B_i(I\!R^s_+)$ is contained in a finite union of hyperplanes in $I\!R^s$.

Let us now confirm that, for some single hyperplane $\mathcal{H}_o \subset I\!R^s$, the set $\mathcal{Z}_{\nu,o} := \{\zeta \in I\!R^s : \nu(\zeta + \mathcal{H}_o) > 0\}$ has Lebesgue measure zero. It holds that

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

where $a : I\!R^m \to I\!R$ denotes the linear transformation induced by a^T and a^{-1} is the preimage.

Hence, $\mathcal{Z}_{\nu,o} = \{\zeta \in I\!R^s : \nu \circ a^{-1}(\{a^T\zeta\}) > 0\}$. Now $\nu \circ a^{-1}$ is a probability measure on $I\!R$ and $a^T\zeta$ 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

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

have Lebesgue measure zero.

Now select E in (2.6) as the preimage $A^{-1}(\mathcal{Z}_\mu \cup \mathcal{Z}_\nu)$. Then both Q_ν and Q_μ are differentiable at Ax for all $x \in U \setminus E$ and we have for those x

$$\begin{aligned}\nabla(Q_\nu - Q_\mu)(Ax) &= \nabla\left(\int_{I\!R^s \setminus \{Ax + \mathcal{H}\}} Q(z - Ax)\nu(\mathrm{d}z) - \int_{I\!R^s \setminus \{Ax + \mathcal{H}\}} Q(z - Ax)\mu(\mathrm{d}z)\right) \\ &= \int_{I\!R^s \setminus \{Ax + \mathcal{H}\}} \nabla Q(z - Ax)(\nu - \mu)\mathrm{d}z,\end{aligned}$$

where the first identity is valid because $\nu(Ax + \mathcal{H}) = \mu(Ax + \mathcal{H}) = 0$. The second identity follows from Lebesgue's theorem on dominated convergence. Indeed, since Q is globally Lipschitzian on $I\!R^s$, the difference quotients related to ∇Q are bounded above by a uniform constant yielding the integrable majorant making Lebesgue's theorem work.

We continue the above identity as follows:

$$\nabla(Q_\nu - Q_\mu)(Ax) = \int_{Ax + \bigcup\limits_{i=1}^{\ell} \operatorname{int} B_i(I\!R^s_+)} \nabla Q(z - Ax)(\nu - \mu)(\mathrm{d}z)$$

$$= \sum_{i=1}^{\ell} d_i(\nu - \mu)(Ax + B_i(I\!R^s_+))$$

$$= \sum_{i=1}^{\ell} d_i(F_{\nu\circ(-B_i)}(-B_i^{-1}Ax) - F_{\mu\circ(-B_i)}(-B_i^{-1}Ax)),$$

where $-d_i$ is the gradient of Q on int $B_i(I\!R^s_+)$, $i = 1, \ldots, \ell$. Note that for the second identity we have also used that $\mu(Ax + \text{int } B_i(I\!R^s_+)) = \mu(Ax + B_i(I\!R^s_+))$, $\nu(Ax + \text{int } B_i(I\!R^s_+)) = \nu(Ax + B_i(I\!R^s_+))$ for all $x \in U \setminus E$.

The assertion now immediately follows from Theorem 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 2.6. The above estimate is closely related to Theorem 2.1 in [29], where the author uses the normal cones $\bar{C}_j$ $(j = 1, \ldots, k)$ to the set $\{u \in I\!R^s : W^T u \leq q\}$ at its vertices v_j $(j = 1, \ldots, k)$. From linear parametric programming it is known ([16], Satz 6.7) that each of the cones $\bar{C}_j$ $(j = 1, \ldots, k)$ is the union of certain cones $B_i(I\!R^s_+)$ $(i \in \{1, \ldots, \ell\})$ arising in Corollary 2.5.

We prefer to use the cones $B_i(I\!R^s_+)$ since these are simplicial cones, which allows a direct relation to distribution functions.

Remark 2.7. Consider the right-hand side of the estimate in Corollary 2.5 and take the suprema with respect to $t \in I\!R^s$ 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 $F_{\nu\circ(-B_i)}$ $(i = 1, \ldots, \ell)$.

Remark 2.8. Theorem 2.4 remains valid under any hypothesis on g and C leading to Lipschitz continuity of the multifunction $\bar{X}(y) := \text{argmin}\{g(x) : Ax = y,\ x \in C\}$.

The next example shows that (already for contaminated distributions) Theorem 2.4 is lost for a general closed convex set $C \subset I\!R^m$. Another counterexample involving the function g can be constructed following the guidelines of Remark 2.9 in [22].

Example 2.9. In (1.1)–(1.3) let $m = 2$, $s = 1$, $g(x) \equiv 0$, $A = (1, 0)$, $C = \{x \in I\!R^2 : (x_2)^2 \leq x_1\}$, $q = (1, 1)^T$, $W = (1, -1)$, and μ be the uniform distribution on $[-1/2, 1/2]$. Let δ_1 denote the probability measure on $I\!R$ 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 Q_μ holds for $V = (-1/2, 1/2)$. 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 $U \subset I\!R^2$ taken as the closed ball around zero with radius $1/2$ (for instance), one confirms that $d(\mu, \mu_t; U) = \text{const} \cdot t$, i.e., the assertion of Theorem 2.4 does not hold.

Note that in Example 2.9 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 2.2 without adding assumptions on g and C.

Remark 2.10. Using similar techniques as in the proof of Theorem 2.7 in [22], it can be shown that the assumptions from Theorem 2.4 imply that the second-order growth condition (2.3) in Theorem 2.2 holds.

The following example concludes the comparison between Theorems 2.2 and 2.4. This example demonstrates that the setting in Theorem 2.2 is the more general one. Indeed, Theorem 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 2.4.

Example 2.11. In (1.1)–(1.3) let $m = s = 1$, $g(x) \equiv 0$, $A = 1$, $C = [-1,1]$, $q = (1,1)^T$, $W := (1,-1)$ ("simple recourse"), and μ be the uniform distribution on $[-1,-1/2] \cup [1/2,1]$. Let $\bar{\mu}$ be the uniform distribution on $[-1/2,1/2]$ 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 $\psi(\mu) = [-1/2,1/2]$ and that

$$Q_\mu(x) \geq Q_\mu(x_\mu) + [\mathrm{dist}(x, \psi(\mu))]^2$$

for all $x \in (-1,1)$ and all $x_\mu \in \psi(\mu)$. Hence (2.3) is fulfilled and Theorem 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 2.4; moreover, ψ is not lower semicontinuous at μ.

The next example is interesting because it shows that Theorem 2.4 does not ensure the stability of the optimal value.

Example 2.12. In (1.1)–(1.3) let $m = s = 1$, $g(x) \equiv 0$, $C = [-1,1]$, $A = 1$, $q = (1,1)$, $W = (1,-1)$, $\mu = \delta_o$, 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 $Q_\mu(x) = |x|$ and, thus, $\varphi(\mu) = 0$, $\psi(\mu) = \{0\}$. Furthermore, $Q_{\mu_n}(x) = (1-\frac{1}{n})|x| + \frac{1}{n}(n^2 - x)$. Therefore, $\psi(\mu_n) = \{0\}$, $\varphi(\mu_n) = n$. The assumptions of Proposition 2.3 and Theorem 2.4 are fulfilled, but $\varphi(\mu_n) \to \infty$ as $n \to \infty$.

The following result (established in [24]) provides a handy tool for checking the strong convexity of Q_μ needed in Theorem 2.4.

PROPOSITION 2.13 ([24], Theorem 2.2). *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* $\mathrm{dist}(t,V) \leq \varrho$.

Then Q_μ is strongly convex on V.

In [22], [23] 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 $\mathcal{M}_1(I\!R^s)$ [17]. In fact, Theorem 2.7 in [22] 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 2.4 in the present paper. Furthermore, [22] contains an example (Remark 2.9) showing the optimality of the convergence rate 1/2. We now analyze this example with respect to the pseudometric $d(\mu,\nu;U)$. The setting is as in Examples 2.11 and 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

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

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 2.4 are fulfilled. In [22] we computed $W_1(\mu,\mu_n) = \varepsilon_n^2$, which showed the 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 2.4 does not outperform Theorem 2.7 in [22]. However, for certain specific modes of perturbation, Theorem 2.4 yields stronger estimates than Theorem 2.7 in [22] (contaminated distributions, asymptotic properties of nonparametric estimators; see the analysis that follows).

To further explain the essence of Theorem 2.4, let us mention that Theorem 2.7 in [22] also yields the convergence rate 1/2 when one replaces the Wasserstein distance W_1 by

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

where $U \subset I\!R^m$ is a suitable nonempty, convex, compact set. Therefore, the approach taken here differs from older ones because it measures the distance of the objectives in the original and the perturbed programs in terms of their subgradients rather 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 2.12. Hence, when aiming at the stability of the optimal value one should resort to a distance like d^*. In [15], Proposition 8, it is shown that convergence to zero of d^* does imply the same for d, provided that the original function Q_μ is differentiable.

PROPOSITION 2.14 (contaminated distributions). *Suppose* (A1)–(A3) *and that* $\psi(\mu)$ *is nonempty 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 the function* Q_μ *is strongly convex on* V. *Let* $\bar\mu \in \mathcal{M}_1(I\!R^s)$ *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 $Q_{\mu_t} - Q_\mu = t(Q_{\bar\mu} - Q_\mu)$. 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 2.4. □

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

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

where $\bar m = 2s$ and $q^+, q^- \in I\!R^s$. Shapiro has derived a similar result (Theorem 3.1 in [29]) by a direct estimate from Theorem 2.2.

PROPOSITION 2.15. *Let* $P(\mu)$ *be a simple recourse model,* $\psi(\mu)$ *be nonempty and bounded,* g *be convex quadratic, and* C *be a nonempty 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 $\nu \in \mathcal{M}_1(I\!R^s)$ *is chosen such that the right-hand side is sufficiently small.*

Proof. First note that the function Q in (2.7) is separable with respect to the components of t. Therefore, the functions Q_ν and Q_μ here depend only on the one-dimensional marginal distributions of ν and μ (cf. also [6], [31]), and we can assume without loss of generality that ν, μ are probability measures with independent one-dimensional marginals. Then our assumptions and Proposition 2.13 yield that Q_μ is strongly convex on some convex open set $V \supset A(\psi(\mu))$.

Here the cones $B_i(\mathbb{R}^s_+)$ $(i = 1, \ldots, 2^s)$ are orthants. To estimate $(\nu - \mu)(Ax + B_i(\mathbb{R}^s_+))$, let us fix some $B_i(\mathbb{R}^s_+)$ for which we assume without loss of generality that

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

Our independence assumption then yields

$$(\nu - \mu)(Ax + B_i(\mathbb{R}^s_+)) = \prod_{j=1}^{s_i} \nu_j((-\infty, (Ax)_j]) \cdot \prod_{j=s_i+1}^{s} \nu_j([(Ax)_j, +\infty)) - \prod_{j=1}^{s_i} \mu_j((-\infty, (Ax)_j]) \cdot \prod_{j=s_i+1}^{s} \mu_j([(Ax)_j, +\infty)).$$

Using the inequality

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

(which can be shown by induction), we obtain

$$|(\nu - \mu)(Ax + B_i(\mathbb{R}^s_+))| \leq \sum_{j=1}^{s_i} |F_{\nu_j}((Ax)_j) - F_{\mu_j}((Ax)_j)| + \sum_{j=s_i+1}^{s} |F^-_{\nu_j}((Ax)_j) - F^-_{\mu_j}((Ax)_j)|,$$

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 2.5) the superscripts can be dropped and we obtain

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

and the proof is completed as with Corollary 2.5. □

3. Applications to asymptotic analysis.

In this section, we show how to employ the Lipschitz stability result of §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 $\mathbb{R}^s$-valued random variables on a probability space $(\Omega, \mathfrak{A}, \mathbf{P})$ having joint distribution μ. Let δ_z denote the probability measure assigning unit mass to $z \in \mathbb{R}^s$. We consider the empirical measures

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

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 a $\mathfrak{A}$-measurable mapping due to Theorem 2K in [21].

PROPOSITION 3.1. *Under the assumptions of Theorem* 2.4 *it holds that*

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

where L and ℓ denote the Lipschitz modulus and the number of basis matrices, respectively, arising in Corollary 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 ([18], p. 302, [25]):

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)^{\frac{1}{2}}\sup_{t\in\mathbb{R}^s}|F_{\mu\circ(-B_j)}(t)-F_{\mu_n(\omega)\circ(-B_j)}(t)|\le\frac{1}{2}$$

$\mathbf{P}$-almost surely for all $j=1,\ldots,\ell$.

Hence, the estimate from Corollary 2.5 is valid for $\mathbf{P}$-almost all $\omega\in\Omega$ with $\nu:=\mu_n(\omega)$, provided that $n=n(\omega)\in\mathbb{N}$ is sufficiently large.

Thus we have for $\mathbf{P}$-almost all $\omega\in\Omega$

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)^{\frac{1}{2}}d_H(\psi(\mu),\psi(\mu_n(\omega)))\le\frac{L\ell}{2}. \qquad \square$$

In [4], [12] the authors obtain consistency results under weak hypotheses on the optimization problems involved (based on the theory of epi-convergence). Proposition 3.1 supplements these results by giving, under stronger assumptions, the (optimal) rate of convergence for the solution sets. Compared to considerations in [29], we can dispense with the linear-independence assumptions imposed there. This became possible because we used simplical cones instead of more general ones (cf. Remark 2.6). Compared to [4], [33] we do not need the unique solvability of $P(\mu)$.

PROPOSITION 3.2. *Under the assumptions of Theorem* 2.4 *there exists a constant $\varepsilon_o>0$ such that it holds for all $\varepsilon\in(0,\varepsilon_o]$ that*

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

Proof. For brevity, we introduce the following notation:

$$\Phi_n(\omega):=\max_{j=1,\ldots,\ell}\eta_{n,j}(\omega),$$
$$\eta_{n,j}(\omega):=\sup_{t\in\mathbb{R}^s}|F_{\mu\circ(-B_j)}(t)-F_{\mu_n(\omega)\circ(-B_j)}(t)| \quad (\omega\in\Omega).$$

Now select $\varepsilon_o>0$ in such a way that $L\ell\Phi_n(\omega)<\varepsilon_o$ and Corollary 2.5 imply

$$d_H(\psi(\mu),\psi(\mu_n(\omega)))\le L\ell\Phi_n(\omega).$$

Then we have for each $\varepsilon\in(0,\varepsilon_o]$ and all $n\in\mathbb{N}$

$$\{\omega: d_H(\psi(\mu),\psi(\mu_n(\omega)))\ge\varepsilon\}\subseteq\Big\{\omega:\Phi_n(\omega)\ge\frac{\varepsilon_o}{L\ell}\Big\}\cup\Big\{\omega:\Phi_n(\omega)\ge\frac{\varepsilon}{L\ell}\Big\}$$
$$=\Big\{\omega:\Phi_n(\omega)\ge\frac{\varepsilon}{L\ell}\Big\}=\bigcup_{j=1}^{\ell}\Big\{\omega:\eta_{n,j}(\omega)\ge\frac{\varepsilon}{L\ell}\Big\}$$

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. [9], [25]) then implies that, for each $\delta > 0$ and $j \in \{1, \ldots, \ell\}$, there exist constants $C_j > 0$ such that

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

Hence, we obtain for any $n \in \mathbb{N}$ and any $\delta > 0$,

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

and the proof is complete. □

Compared to Theorem 4.6 in [8], which represents a large deviation result for more general stochastic programs, Proposition 3.2 requires only 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 in [30] which, in the context of two-stage stochastic programming, works for nonunique solutions but applies only to measures μ with bounded support.

Another substantial step in the asymptotic analysis of optimal solutions involves 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 [11], [27] 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 nonsmooth (cf. (1.2), (1.3)) and uniqueness of optimal solutions is rather exceptional (cf. Example 1.1) such that the results from [11], [27] do not apply.

From Theorem 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 [5]. For simple recourse models the lower estimate becomes more detailed (Remark 3.4).

PROPOSITION 3.3. *Under the assumptions of Theorem* 2.4 *there exist probability distribution functions* G_j, $j = 1, \ldots, \ell$, *on* $\mathbb{R}$ *such that it holds that*

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

for all $t \geq 0$, *where* L *and* ℓ *denote the Lipschitz modulus and the number of basis matrices, respectively, arising in Corollary* 2.5.

Proof. Let $\eta_{n,j}(\omega)$ be given as in the proof of Proposition 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 [10], Chap. 2.1.5 in [25]).

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

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

where $\delta > 0$ is selected according to Corollary 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 2.5 then yields the following chain of inequalities:

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

Hence we obtain the following estimate for all $t \geq 0$ and $n \in \mathbb{N}$:

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

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

$$\liminf_{n \to \infty} \mathbf{P}(\{\omega : n^{\frac{1}{2}} d_H(\psi(\mu), \psi(\mu_n(\omega))) < t\})$$

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

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

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

Remark 3.4. Unfortunately, the limit distributions G_j $(j = 1, \ldots, \ell)$ cannot be characterized in general for multidimensional distributions $F_{\mu\circ(-B_j)}$ (see [10]). However, in the case of simple recourse, we obtain the following estimate by using Proposition 2.15 instead of Corollary 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.

Acknowledgments. This paper has benefited from several discussions with Alexander Shapiro (Georgia Institute of Technology, Atlanta) and Diethard Klatte (University of Zurich). Moreover, the suggestions of a referee that led to a more transparent proof of Theorem 2.4 are gratefully acknowledged.

REFERENCES

[1] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, *Non-Linear Parametric Optimization*, Akademie-Verlag, Berlin, 1982.

[2] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[3] J. Dupačová, *Stability and sensitivity analysis for stochastic programming*, Ann. Oper. Res., 27 (1990), pp. 115–142.

[4] J. Dupačová and R. J.-B. Wets, *Asymptotic behaviour of statistical estimators and of optimal solutions of stochastic optimization problems*, Ann. Statist., 16 (1988), pp. 1517–1549.

[5] Y. M. Ermoliev and V. I. Norkin, *Normalized convergence in stochastic optimization*, Ann. Oper. Res., 30 (1991), pp. 187–198.

[6] P. Kall, *Stochastic linear programming*, Springer-Verlag, Berlin, 1976.

[7] ———, *On approximations and stability in stochastic programming*, in Parametric Optimization and Related Topics, J. Guddat, H. Th. Jongen, B. Kummer, and F. Nožička, eds., Akademie Verlag, Berlin, 1987, pp. 387–407.

[8] Y. M. Kaniovski, A. J. King, and R. J.-B. Wets, *Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems*, Ann. Oper. Res., 56 (1995), pp. 189–208.

[9] J. Kiefer, *On large deviations of the empiric D.F. of vector chance variables and a law of the iterated logarithm*, Pacific J. Math., 11 (1961), pp. 649–660.

[10] J. Kiefer and J. Wolfowitz, *On the deviations of the empiric distribution function of vector chance variables*, Trans. Amer. Math. Soc., 87 (1958), pp. 173–186.

[11] A. J. King and R. T. Rockafellar, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, Math. Oper. Res., 18 (1993), pp. 148–162.

[12] A. J. King and R. J.-B. Wets, *Epi-consistency of convex stochastic programs*, Stochastics Stochastics Rep., 34 (1991), pp. 83–92.

[13] D. Klatte, *Beiträge zur Stabilitätsanalyse nichtlinearer Optimierungsprobleme*, Dissertation B, Humboldt-Universität Berlin, Sektion Mathematik, Berlin, 1984.

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

[15] B. Kummer, *Generalized equations: Solvability and regularity*, Math. Programming Stud., 21 (1984), pp. 199–212.

[16] F. Nožička, J. Guddat, H. Hollatz, and B. Bank, *Theorie der linearen parametrischen Optimierung*, Akademie-Verlag, Berlin, 1974.

[17] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, UK, 1991.

[18] H. Richter, *Das Gesetz vom iterierten Logarithmus für empirische Verteilungsfunktionen im* $\mathbb{R}^k$, Manuscripta Math., 11 (1974), pp. 291–303.

[19] S. M. Robinson and R. J.-B. Wets, *Stability in two-stage stochastic programming*, SIAM J. Control Optim., 25 (1987), pp. 1409–1416.

[20] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[21] ———, *Integral functionals, normal integrands and measurable selections*, in Nonlinear Operators and the Calculus of Variations, G. P. Gossez et al., eds., Lecture Notes in Math., Vol. 543, Springer-Verlag, New York, 1976, pp. 157–207.

[22] W. Römisch and R. Schultz, *Stability analysis for stochastic programs*, Ann. Oper. Res., 30 (1991), pp. 241–266.

[23] ———, *Stability of solutions for stochastic programs with complete recourse*, Math. Oper. Res., 18 (1993), pp. 590–609.

[24] R. Schultz, *Strong convexity in stochastic programs with complete recourse*, J. Comput. Appl. Math., 56 (1994), pp. 3–22.

[25] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

[26] A. Shapiro, *Asymptotic properties of statistical estimators in stochastic programming*, Ann. Statist., 17 (1989), pp. 841–858.

[27] ———, *On differential stability in stochastic programming*, Math. Programming, 47 (1990), pp. 107–116.

[28] ———, *Asymptotic analysis of stochastic programs*, Ann. Oper. Res., 30 (1991), pp. 169–186.

[29] ———, *Quantitative stability in stochastic programming*, Math. Programming, 67 (1994), pp. 99–108.

[30] S. Vogel, *Stability results for stochastic programming problems*, Optimization, 19 (1988), pp. 269–288.

[31] R. J.-B. Wets, *Stochastic programs with fixed recourse: The equivalent deterministic program*, SIAM Rev., 16 (1974), pp. 309–339.

[32] ———, *Stochastic Programming*, in Handbooks in Operations Research and Management Science, Vol. 1, Optimization, G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, eds., North-Holland, Amsterdam, 1989, pp. 573–629.

[33] ———, *Constrained estimation: Consistency and asymptotics*, Appl. Stochastic Models Data Anal., 7 (1991), pp. 17–32.