STABILITY ANALYSIS FOR STOCHASTIC PROGRAMS

Werner RÖMISCH and Rüdiger SCHULTZ

Humboldt-Universität Berlin, Sektion Mathematik, P.O. Box 1297, D-1086 Berlin, Germany

For stochastic programs with recourse and with (several joint) probabilistic constraints, respectively, we derive quantitative continuity properties of the relevant expectation functionals and constraint set mappings. This leads to qualitative and quantitative stability results for optimal values and optimal solutions with respect to perturbations of the underlying .probability distributions. Earlier stability results for stochastic programs with recourse and for those with probabilistic constraints are refined and extended, respectively. Emphasis is placed on equipping sets of probability measures with metrics that one can handle in specific situations. To illustrate the general stability results we present possible consequences when estimating the original probability measure via empirical ones.

Keywords: Stochastic programs with recourse, stochastic programs with probabilistic constraints, distribution sensitivity, probability metrics.

1. Introduction

When formulating a stochastic programming model, one tacitly assumes the underlying probability distribution to be given. In practical situations, however, this is rarely the case; moreover, one often has to live with incomplete information and approximations. Furthermore, also under full information about the underlying distributions one is led to approximations, since exact computation of expectations and probabilities typically arising in stochastic programming is beyond the present numerical capabilities for a large class of distributions (e.g. multivariate continuous ones). These circumstances motivate a stability analysis for optimal values and optimal solutions to stochastic programs with respect to perturbations of the underlying probability distributions (cf. [10,17,31,33,43]). In the present paper, we pursue this for two basic problem classes in stochastic programming – for stochastic programs with recourse and for stochastic programs with probabilistic (or chance) constraints. We lay stress on structural properties of expectation functionals and of certain multifunctions defined by probabilities, on implications of these properties with respect to stability, on a proper selection of metrics in spaces of probability measures to guarantee the structural properties, on the one hand, and to be able to compute (or to estimate) distances of probability measures in specific situations, on the other hand.

In section 2, we analyze stochastic programs with linear and quadratic re-course, respectively. We prove a general Lipschitz continuity result for a mapping which assigns to a probability measure a certain integral expression, where the metric on the space of probability measures is an L_p-Wasserstein metric [9,15,26]. This leads to (local) upper semicontinuity of optimal solutions and to (local) Lipschitz continuity of optimal values for the mentioned recourse problems. For stochastic programs with linear recourse and random right-hand sides we estab-lish (local) Hölder continuity of the Hausdorff distance of optimal solutions. Compared to [33,36] where a similar analysis was carried out with respect to the bounded Lipschitz metric (cf. [9,15] for details about the metric) the progress consists of less restrictive integrability conditions for the involved measures and of better possibilities to estimate distances between probability measures. The above mentioned Hölder continuity result for solution sets extends a result in [35] (theorem 4.4) to recourse problems with a convex quadratic portion in the objective. Among the consequences of our general stability results for specific situations we present an asymptotic convergence result for optimal solution sets when the true distribution is estimated by empirical ones.

Section 3 is devoted to stochastic programs with several joint probabilistic constraints. Using discrepancies (cf. [12]), we identify a suitable metric to ensure a certain Lipschitz property for the mapping which assigns the probabilistic constraint set to a probability measure. As implications for stability we again obtain (local) upper semicontinuity of optimal solutions and (local) Lipschitz continuity of optimal values, thus extending results in [33,34,36] to problems with several probabilistic constraints. We apply the general result to a model where the measure has a certain convexity property and to a specific chance constrained program with random technology matrix.

Let us add some further bibliographical comments: Approximation techniques for solving stochastic programs with complete information on the underlying measures are presented in [4,12,37,42]. When having only incomplete information about the measures it is possible to use parametric as well as non-parametric statistical estimators instead of the true distributions. For parametric estimators, the stability of the stochastic program with respect to changes of finite dimen-sional parameters is essential (cf. e.g. [10]). In connection with non-parametric statistical estimators there exist a number of contributions to stochastic program-ming [11,18,19,38,40]. The conclusions in the present paper with respect to asymptotic properties of statistical estimators fit into this line of research.

2. Wasserstein metrics and stability of recourse problems

Let (Z, ρ) be a separable metric space, denote by $\mathscr{P}(Z)$ the set of all Borel probability measures on Z and consider a mapping $h: Z \to \mathbb{R}$ which is

Lipschitzian on bounded sets, i.e.

$$L_h(r) := \sup\left\{ \frac{|h(z) - h(\tilde{z})|}{\rho(z, \tilde{z})} : z, \tilde{z} \in B(\theta, r), z \neq \tilde{z} \right\} < \infty$$

for each $r > 0$, where $B(\theta, r) := \{ z \in Z: \rho(z, \theta) \leqslant r \}$ and $\theta \in Z$ is some distinguished element.

We begin by studying quantitative continuity of the mapping which assigns to $\mu \in \mathscr{P}(Z)$ the integral $\int_Z h(z) \mu(\mathrm{d}z)$. Subsequently, these results will serve to establish stability of stochastic programs with recourse. Studying quantitative continuity in the above setting requires a suitable metric on $\mathscr{P}(Z)$. In this respect, we address Wasserstein metrics.

The L_p-Wasserstein metric W_p ($p \geqslant 1$) is defined as follows:

$$W_p(\mu, \nu) := \left[\inf\left\{ \int_{Z \times Z} \rho(z, \tilde{z})^p \eta(\mathrm{d}z, \mathrm{d}\tilde{z}) : \eta \in D(\mu, \nu) \right\} \right]^{1/p}$$

for all

$$\mu, \nu \in \mathscr{M}_p(Z) := \left\{ \mu \in \mathscr{P}(Z): \int_Z \rho(z, \theta)^p \mu(\mathrm{d}z) < \infty \right\}.$$

where

$$D(\mu, \nu) := \left\{ \eta \in \mathscr{P}(Z \times Z): \eta \circ \pi_1^{-1} = \mu, \eta \circ \pi_2^{-1} = \nu \right\}$$

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

In [15] it is shown that $(\mathscr{M}_p(Z), W_p)$ is a metric space. In [26], the following equivalence is established:

Let $\mu \in \mathscr{M}_p(Z)$ and $\mu_n \in \mathscr{P}(Z)$ ($n = 1, 2, \ldots$); then $W_p(\mu_n, \mu) \to 0$ as $n \to \infty$ iff the sequence $\{\mu_n\}$ converges weakly to μ and

$$\lim_{n \to \infty} \int_Z \rho(z, \theta)^p \mu_n(\mathrm{d}z) = \int_Z \rho(z, \theta)^p \mu(\mathrm{d}z).$$

A sequence $\{\mu_n\}$ is said to converge weakly to μ, if for any bounded continuous function $g: Z \to \mathbb{R}$ we have

$$\int_Z g(z) \mu_n(\mathrm{d}z) \to \int_Z g(z) \mu(\mathrm{d}z) \quad \text{as } n \to \infty$$

(cf. [3]).

Concerning the announced continuity we state

PROPOSITION 2.1

Let $h: Z \to \mathbb{R}$ be Lipschitzian on bounded subsets of Z. Then, for all μ, $\nu \in \mathscr{M}_p(Z)$

$$\left| \int_Z h(z) \mu(\mathrm{d}z) - \int_Z h(z) \nu(\mathrm{d}z) \right| \leqslant \left(M_q(\mu) + M_q(\nu) \right) \cdot W_p(\mu, \nu),$$

where $p > 1$, $p^{-1} + q^{-1} = 1$ and

$$M_q(\mu) := \left(\int_Z L_h(\rho(z, \theta))^q \mu(dz) \right)^{1/q}.$$

Proof

Let $\eta \in D(\mu, \nu)$; then it holds

$$\left| \int_Z h(z)\mu(dz) - \int_Z h(\tilde{z})\nu(d\tilde{z}) \right|$$

$$\leq \int_{Z \times Z} |h(z) - h(\tilde{z})| \, \eta(dz, d\tilde{z})$$

$$\leq \int_{Z \times Z} L_h(\max\{\rho(z, \theta), \rho(\tilde{z}, \theta)\}) \cdot \rho(z, \tilde{z}) \eta(dz, d\tilde{z})$$

$$\leq \left(\int_{Z \times Z} [L_h(\max\{\rho(z, \theta), \rho(\tilde{z}, \theta)\})]^q \eta(dz, d\tilde{z}) \right)^{1/q}$$

$$\times \left(\int_{Z \times Z} \rho(z, \tilde{z})^p \eta(dz, d\tilde{z}) \right)^{1/p}.$$

By

$$L_h(\max\{\rho(z, \theta), \rho(\tilde{z}, \theta)\}) \leq L_h(\rho(z, \theta)) + L_h(\rho(\tilde{z}, \theta))$$

we can continue

$$\leq \left[\left(\int_Z [L_h(\rho(z, \theta))]^q \mu(dz) \right)^{1/q} + \left(\int_Z [L_h(\rho(\tilde{z}, \theta))]^q \nu(d\tilde{z}) \right)^{1/q} \right]$$

$$\times \left(\int_{Z \times Z} \rho(z, \tilde{z})^p \eta(dz, d\tilde{z}) \right)^{1/p}.$$

Passing to the infimum with respect to $\eta \in D(\mu, \nu)$ yields the assertion. $\square$

Remark 2.2.

For globally Lipschitzian h, i.e. in case $L_h(r) \leq L_h = \text{const.}$ for all $r > 0$, the above proof yields

$$\left| \int_Z h(z)\mu(dz) - \int_Z h(z)\nu(dz) \right| \leq L_h \cdot W_p(\mu, \nu) \quad (\mu, \nu \in \mathcal{M}_p(Z)),$$

and this remains valid even for $p = 1$. Note that, in this situation, the result may also be gained as a consequence of the Kantorovich–Rubinstein theorem (e.g. [15], p. 233).

COROLLARY 2.3

Let $\mu \in \mathcal{M}_2(\mathbb{R}^s)$. If there exists $L_0 > 0$ such that $L_h(r) \leqslant L_0 r$ for all $r > 0$, then there exist $L > 0$ and $\delta > 0$ such that

$$\left| \int_{\mathbb{R}^s} h(z)\mu(\mathrm{d}z) - \int_{\mathbb{R}^s} h(z)\nu(\mathrm{d}z) \right| \leqslant L W_2(\mu, \nu),$$

whenever $\nu \in \mathcal{M}_2(\mathbb{R}^s)$, $W_2(\mu, \nu) < \delta$.

Proof

Specifying proposition 2.1 ($p = 2$ and $Z = \mathbb{R}^s$ with norm $\| \cdot \|$), we obtain

$$\left| \int_{\mathbb{R}^s} h(z)\mu(\mathrm{d}z) - \int_{\mathbb{R}^s} h(z)\nu(\mathrm{d}z) \right|$$

$$\leqslant L_0 \left(\left(\int_{\mathbb{R}^s} \| z \|^2 \mu(\mathrm{d}z) \right)^{1/2} + \left(\int_{\mathbb{R}^s} \| z \|^2 \nu(\mathrm{d}z) \right)^{1/2} \right) \cdot W_2(\mu, \nu),$$

whenever $\mu, \nu \in \mathcal{M}_2(\mathbb{R}^s)$. The characterization of convergence with respect to W_2 now yields that there exist $\delta > 0$ and $M > 0$ such that

$$\int_{\mathbb{R}^s} \| z \|^2 \nu(\mathrm{d}z) \leqslant M \quad \text{whenever} \quad W_2(\mu, \nu) < \delta.$$

This completes the proof. $\square$

The recourse problems, whose stability we are going to analyze, all fit into the following class of parametric programs:

$$\mathbf{P}(\mu) \qquad \min\{ F(x, \mu): x \in C \},$$

where $C \subseteq \mathbb{R}^m$ is nonempty, closed, convex and $F: \mathbb{R}^m \times \mathscr{P}(\mathbb{R}^s) \to \mathbb{R}$ is given by

$$F(x, \mu) := \int_{\mathbb{R}^s} f(x, z)\mu(\mathrm{d}z),$$

with some function $f: \mathbb{R}^m \times \mathbb{R}^s \to \mathbb{R}$ and some $\mu \in \mathscr{P}(\mathbb{R}^s)$.

For the moment, let us assume all ingredients to be taken in such a way that the above integral exists. Precise assumptions will be given when considering the specified problems. The (global) optimal value of $\mathbf{P}(\mu)$ and the set of (global) minimizers are denoted $\phi(\mu)$ and $\psi(\mu)$, respectively.

THEOREM 2.4

Fix $\mu \in \mathcal{M}_2(\mathbb{R}^s)$ and let

(1) $\psi(\mu)$ be nonempty and bounded;
(2) the function $f(\cdot, z)$ be convex for each $z \in \mathbb{R}^s$;
(3) there exists $V \subset \mathbb{R}^m$, open and bounded, and $L_0 > 0$ such that $\psi(\mu) \subset V$ and

$$| f(x, z) - f(x, \tilde{z}) | \leqslant L_0 \max\{ \| z \|, \| \tilde{z} \| \} \cdot \| z - \tilde{z} \|,$$

whenever $x \in \mathrm{cl}\, V$ and $z, \tilde{z} \in \mathbb{R}^s$.

Then the multifunction $\psi(\cdot)$ (from $(\mathcal{M}_2(\mathbb{R}^s), W_2)$ into $\mathbb{R}^m$) is upper semicontinuous at μ, and there exist constants $L > 0$ and $\delta > 0$ such that $\psi(\nu) \neq \emptyset$ and

$$|\phi(\nu) - \phi(\mu)| \leq L \cdot W_2(\mu, \nu),$$

whenever $\nu \in \mathcal{M}_2(\mathbb{R}^s)$, $W_2(\mu, \nu) < \delta$.

Proof

We show that the assertion follows from lemma A.2 in the appendix. Assumption (a) in lemma A.2 is a consequence of (1), and (b), (c) hold since the constraint set C is closed and does not depend on the parameter. Hence it remains to verify (d). Indeed, for arbitrary $x, \tilde{x} \in \mathrm{cl}\, V$ and $\nu \in \mathcal{M}_2(\mathbb{R}^s)$ we have

$$|F(x, \mu) - F(\tilde{x}, \nu)| \leq |F(x, \mu) - F(\tilde{x}, \mu)| + |F(\tilde{x}, \mu) - F(\tilde{x}, \nu)|.$$

In view of (2), the function F is convex in x and thus Lipschitzian on bounded sets. Hence, the first term on the right is bounded above by $\tilde{L} \|x - \tilde{x}\|$ with a suitable constant $\tilde{L} > 0$. To see that there exist constants $\hat{L} > 0$, $\delta > 0$ such that the second term on the right is bounded above by $\hat{L} \cdot W_2(\mu, \nu)$ whenever $W_2(\mu, \nu) < \delta$, we apply corollary 2.3 (for $h(z) := f(\tilde{x}, z)$), whose assumptions are fulfilled in view of (3). □

Remark 2.5

If $f(x, \cdot)$ is globally Lipschitzian (uniformly with respect to $x \in \mathrm{cl}\, V$), then, by remark 2.2, the above remains valid for μ, $\nu \in \mathcal{M}_1(\mathbb{R}^s)$ with respect to W_1.

Remark 2.6

Using the same approach, it is possible to cover more general situations where the Lipschitz modulus of $f(x, \cdot)$ grows faster than linearly. The crucial point is to "compensate" the growth of the Lipschitz modulus by more restrictive moment conditions on μ and ν. Of course, the stability results one then obtains are with respect to less convenient metrics (W_p with $p > 2$).

Now, let us discuss instances of $f(x, z)$ which lead to recourse models in stochastic programming. As a basic structure we always assume

$$f(x, z) = g(x) + \tilde{Q}(x, z),$$

with some convex function $g: \mathbb{R}^m \to \mathbb{R}$.

In models with *linear recourse* we have

$$\tilde{Q}(x, z) = \min\{q^\mathsf{T} y: Wy = b - Ax, \; y \geq 0\}, \tag{2.1}$$

where $q \in \mathbb{R}^{s_1}$, $W \in L(\mathbb{R}^{s_1}, \mathbb{R}^{s_2})$ such that $W(\mathbb{R}^{s_1}_+) = \mathbb{R}^{s_2}$ (complete recourse), $A \in L(\mathbb{R}^m, \mathbb{R}^{s_2})$, $b \in \mathbb{R}^{s_2}$. (Throughout, superscript T denotes transposition.) The vector z is formed by the components of q, A and b. We suppose the supports of the measure μ and all its perturbations ν to be contained in the set

$$\mathscr{Z} = \{z \in \mathbb{R}^s: \exists u \in \mathbb{R}^{s_2}, \; W^\mathsf{T} u \leq q\} \quad (s = s_1 + (m+1)s_2).$$

Then we have $\tilde{Q}(x, z) \in \mathbb{R}$ whenever $x \in \mathbb{R}^m$, $z \in \mathscr{Z}$. Now, the only further assumptions one must impose to achieve the stability asserted in theorem 2.4 are that $\mu \in \mathscr{M}_2(\mathbb{R}^s)$ and that $\psi(\mu)$ is nonempty and bounded. Indeed, assumptions (2) and (3) in theorem 2.4 are both consequences of well-known facts in linear parametric programming (consult [16] for (2) and lemma 3.2 in [36] for (3)).

Moreover, if q in (2.1) is non-random and $\{ u \in \mathbb{R}^{s_2}: W^{\mathrm{T}}u \leqslant q \} \neq \varnothing$, then, with z formed by the components of b and A, the function $\tilde{Q}(x, \cdot)$ is globally Lipschitzian uniformly with respect to x varying in a compact set. Thus, remarks 2.2 and 2.5 apply, and we can relax the assumption $\mu \in \mathscr{M}_2(\mathbb{R}^s)$ to $\mu \in \mathscr{M}_1(\mathbb{R}^s)$ ($s := (m+1)s_2$) which leads to stability for perturbations $\nu \in \mathscr{M}_1(\mathbb{R}^s)$ with respect to the metric W_1.

A specific instance of (2.1) to be dealt with subsequently in this paper is the *linear recourse model with random right-hand side*, which is given by

$$\tilde{Q}(x, z) = \min\{ q^{\mathrm{T}}y: Wy = z - Ax, \ y \geqslant 0 \}, \tag{2.2}$$

where A is non-random and $\mathscr{Z}$ coincides with $\mathbb{R}^s$ ($s = s_2$). Of course, also for this model the statements made in theorem 2.4 remain valid when considering $(\mathscr{M}_1(\mathbb{R}^s), W_1)$ as parameter space.

Models with *quadratic recourse* are given by

$$\tilde{Q}(x, z) = \max\{ -y^{\mathrm{T}}Hy + (b - Ax)^{\mathrm{T}}y: Dy \leqslant d, \ y \geqslant 0 \}, \tag{2.3}$$

where $H \in L(\mathbb{R}^{s_1}, \mathbb{R}^{s_1})$ symmetric, positive definite, $b \in \mathbb{R}^{s_1}$, $A \in L(\mathbb{R}^m, \mathbb{R}^{s_1})$, $D \in L(\mathbb{R}^{s_1}, \mathbb{R}^{s_2})$, $d \in \mathbb{R}^{s_2}$. The vector z is formed by the components of b, A and d; the set $\mathscr{Z}$ containing the supports of the measures μ and ν is

$$\mathscr{Z} = \{ z \in \mathbb{R}^s: \exists y \in \mathbb{R}^{s_1}, \ Dy \leqslant d, \ y \geqslant 0 \} \quad (s = (m+1)s_1 + s_2).$$

Again, assumptions (2) and (3) are implications of known convexity results and of representations for infima of (now) quadratic programs which depend on parameters (consult [33], proof of proposition 3.2). To have the stability asserted in theorem 2.4 it suffices hence to claim $\mu \in \mathscr{M}_2(\mathbb{R}^s)$ and $\psi(\mu)$ nonempty and bounded.

For stochastic programs with linear recourse and random right-hand side we can quantify the upper semicontinuity of $\psi(\cdot)$ asserted in theorem 2.4. In its setting, the following theorem differs from theorem 4.4 in [35] merely by allowing g to be convex quadratic instead of linear. However, since the refinement of the corresponding proof needs some effort, and to make the argument more transparent to the reader we will present the complete proof, although it sometimes parallels that in [35]. Since there is no hope of obtaining the following result for general convex g and C (cf. remark 2.9 below), subclasses of convex problems where the result still holds are interesting. Our motivation to consider the extension to convex quadratic g stems from a recourse model in optimal power dispatch about which we will report in a subsequent paper.

Before stating the theorem let us recall that a function $\Phi: \mathbb{R}^s \to \mathbb{R}$ is called strongly convex on a convex set $V \subset \mathbb{R}^s$, if there exists $\kappa > 0$ such that

$$\Phi(\lambda x + (1 - \lambda)\tilde{x}) \leqslant \lambda \Phi(x) + (1 - \lambda)\Phi(\tilde{x}) - \kappa \lambda (1 - \lambda) \| x - \tilde{x} \|^2,$$

whenever $x, \tilde{x} \in V$. Furthermore, let us introduce the notation

$$\tilde{h}(v) = \min\{q^T y: Wy = v, \ y \geqslant 0\},$$

with $v \in \mathbb{R}^s$ and q, W as in (2.2).

THEOREM 2.7

Adopt the general setting given above together with the specification (2.2). Let g be convex quadratic and C polyhedral. Fix $\mu \in \mathcal{M}_1(\mathbb{R}^s)$, let $\psi(\mu)$ be nonempty, bounded and let the function $Q_\mu: \mathbb{R}^s \to \mathbb{R}$, given by

$$Q_\mu(\zeta) = \int_{\mathbb{R}^s} \tilde{h}(z - \zeta) \mu(\mathrm{d}z),$$

be strongly convex on a convex open set $\tilde{V}$ containing $A(\psi(\mu))$. Then there exist constants $L^* > 0$ and $\delta^* > 0$ such that

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

whenever $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $W_1(\mu, \nu) < \delta^*$. (d_H denotes the Hausdorff distance.)

Proof

In a first step we are going to apply lemma A.3 in the appendix to the parametric program

$$\mathbf{P}(\nu) \qquad \min\{F(x, \nu): x \in C\},$$

with parameter space $(T, d) = (\mathcal{M}_1(\mathbb{R}^s), W_1)$, $t_0 = \mu$ and objective function

$$F(x, \nu) = g(x) + Q_\nu(Ax).$$

By assumption, the function g admits a representation

$$g(x) = x^T H x + c^T x,$$

with $H \in L(\mathbb{R}^m, \mathbb{R}^m)$ symmetric and positive semidefinite, and $c \in \mathbb{R}^m$. The assumptions on H imply that there exists $H^{1/2} \in L(\mathbb{R}^m, \mathbb{R}^m)$ symmetric and positive semidefinite, such that $H^{1/2} \cdot H^{1/2} = H$.

Then, for arbitrary $x, \tilde{x} \in \mathbb{R}^m$, $\lambda \in \mathbb{R}$, $0 \leqslant \lambda \leqslant 1$, we have the identity

$$(\lambda x + (1 - \lambda)\tilde{x})^T H(\lambda x + (1 - \lambda)\tilde{x})$$
$$= \lambda x^T H x + (1 - \lambda)\tilde{x}^T H \tilde{x} - \lambda(1 - \lambda) \| H^{1/2} x - H^{1/2}\tilde{x} \|^2,$$

where $\|\cdot\|$, as in the remainder of the proof, denotes the Euclidean norm. Strong convexity of $Q_\mu(\cdot)$ together with the above identity yields

$$
\begin{aligned}
F(\lambda x &+ (1-\lambda)\tilde{x}, \mu) \\
&\leqslant \lambda F(x, \mu) + (1-\lambda)F(\tilde{x}, \mu) - \lambda(1-\lambda)\|H^{1/2}x - H^{1/2}\tilde{x}\|^2 \\
&\quad - \kappa\lambda(1-\lambda)\|Ax - A\tilde{x}\|^2,
\end{aligned}
$$

for all x, $\tilde{x} \in \mathbb{R}^m$ such that Ax, $A\tilde{x} \in \tilde{V}$ and all $\lambda \in \mathbb{R}$, $0 \leqslant \lambda \leqslant 1$.

Now, consider a bounded open set V containing $\psi(\mu)$ with the property that $Ax \in \tilde{V}$ for all $x \in V$. With $\lambda = 1/2$ we obtain for arbitrary $x \in C \cap \operatorname{cl} V$ and arbitrary $x_* \in \psi(\mu)$

$$
\begin{aligned}
F(x_*, \mu) &\leqslant F(\tfrac{1}{2}(x + x_*), \mu) \\
&\leqslant \tfrac{1}{2}F(x, \mu) + \tfrac{1}{2}F(x_*, \mu) - \tfrac{1}{4}\|H^{1/2}x - H^{1/2}x_*\|^2 \\
&\quad - \tfrac{1}{4}\kappa\|Ax - Ax_*\|^2,
\end{aligned}
$$

whence

$$
F(x, \mu) \geqslant F(x_*, \mu) + \tfrac{1}{2}\|H^{1/2}x - H^{1/2}x_*\|^2 + \tfrac{1}{2}\kappa\|Ax - Ax_*\|^2.
$$

Thus, hypothesis (b) in lemma A.3 is fulfilled when setting $x_0 := x_*$ for some $x_* \in \psi(\mu)$ and $\alpha(\xi) := \tfrac{1}{2}\|H^{1/2}\xi\|^2 + \tfrac{1}{2}\kappa\|A\xi\|^2$, $\xi \in \mathbb{R}^m$.

To establish hypothesis (c) recall that, due to the general setting, $F(\cdot, \nu)$ is a convex function for all $\nu \in \mathcal{M}_1(\mathbb{R}^s)$. Furthermore, for each $x \in \mathbb{R}^m$ and each $\nu \in \mathcal{M}_1(\mathbb{R}^s)$

$$
\begin{aligned}
|F(x, \nu) - F(x, \mu)| &= |Q_\nu(Ax) - Q_\mu(Ax)| \\
&= \left| \int_{\mathbb{R}^s} \tilde{h}(z - Ax)\nu(\mathrm{d}z) - \int_{\mathbb{R}^s} \tilde{h}(z - Ax)\mu(\mathrm{d}z) \right|.
\end{aligned}
$$

From linear parametric programming it is known that $\tilde{h}$ is a piecewise linear convex real-valued function on $\mathbb{R}^s$; hence, $\tilde{h}$ is globally Lipschitzian with some constant $L_{\tilde{h}} > 0$. By the Kantorovich–Rubinstein theorem (e.g. [15], p. 233) we obtain for all $x \in \mathbb{R}^m$ and all $\nu \in \mathcal{M}_1(\mathbb{R}^s)$:

$$
\begin{aligned}
|F(x, \nu) &- F(x, \mu)| \\
&= L_{\tilde{h}} \left| \int_{\mathbb{R}^s} L_{\tilde{h}}^{-1} \cdot \tilde{h}(z - Ax)(\nu - \mu)(\mathrm{d}z) \right| \\
&\leqslant L_{\tilde{h}} \sup\left\{ \left| \int_{\mathbb{R}^s} \tilde{f}(z)(\nu - \mu)(\mathrm{d}z) \right| : \tilde{f}: \mathbb{R}^s \to \mathbb{R}, \right. \\
&\qquad\qquad\qquad \left. \tilde{f} \text{ is Lipschitzian with modulus less than or equal to } 1 \right\} \\
&\leqslant L_{\tilde{h}} W_1(\mu, \nu).
\end{aligned}
$$

This verifies hypothesis (c) in lemma A.3.

Since hypothesis (a) in lemma A.3 is also fulfilled, we obtain that there exist $L > 0$ and $\delta > 0$ such that $\psi(\nu) \neq \varnothing$ and

$$\sup_{x \in \psi(\nu)} \left\{ \left\| H^{1/2}x - H^{1/2}x_* \right\|^2 + \left\| Ax - Ax_* \right\|^2 \right\} \leqslant L \cdot W_1(\mu, \nu), \tag{2.4}$$

whenever $x_* \in \psi(\mu)$, $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $W_1(\mu, \nu) < \delta$.

In a second step let us consider the matrix $\bar{A} \in L(\mathbb{R}^m, \mathbb{R}^{m+s+1})$ given by

$$\bar{A}^\mathsf{T} = (A^\mathsf{T}, H^{1/2}, c).$$

Denote P the orthogonal projection from $\mathbb{R}^m$ onto $\mathscr{L} := \operatorname{im} \bar{A}^\mathsf{T}$, the image of $\bar{A}^\mathsf{T}$. Due to the choice of $\bar{A}$ we then obtain after some calculation that $F(x, \nu) = F(Px, \nu)$ whenever $x \in \mathbb{R}^m$, $\nu \in \mathcal{M}_1(\mathbb{R}^s)$. Furthermore, for each $\nu \in \mathcal{M}_1(\mathbb{R}^s)$ with $\psi(\nu) \neq \varnothing$, we have the representation

$$\psi(\nu) = C \cap \bigcup_{y \in P(\psi(\nu))} \left\{ y + \mathscr{L}^\perp \right\},$$

with $\mathscr{L}^\perp$ denoting the orthogonal complement of $\mathscr{L}$. To see this, let $x \in \psi(\nu)$. Then there exists $\bar{y} \in \mathscr{L}^\perp$ such that $x = Px + \bar{y}$. Furthermore, $x \in C$, implying $x \in C \cap \bigcup_{y \in P(\psi(\nu))} \{ y + \mathscr{L}^\perp \}$. Conversely, let $x \in C \cap \bigcup_{y \in P(\psi(\nu))} \{ y + \mathscr{L}^\perp \}$. Then $x \in C$, and there exists $y \in P(\psi(\nu))$ such that $x - y \in \mathscr{L}^\perp$. Therefore, $\bar{A}x = \bar{A}y$, which yields $F(x, \nu) = F(y, \nu)$. By $y \in P(\psi(\nu))$ there exists $\bar{y} \in \mathscr{L}^\perp$ such that $y + \bar{y} \in \psi(\nu)$. This implies $\phi(\nu) = F(y + \bar{y}, \nu) = F(y, \nu)$. Thus, $x \in C$ and $F(x, \nu) = \phi(\nu)$, which yields $x \in \psi(\nu)$ and verifies the asserted representation of $\psi(\nu)$.

Denoting $C(y) := \{ x \in C : \bar{A}x = \bar{A}y \}$ we obtain

$$\psi(\nu) = \bigcup_{y \in P(\psi(\nu))} \left\{ C \cap (y + \mathscr{L}^\perp) \right\} = \bigcup_{y \in P(\psi(\nu))} C(y). \tag{2.5}$$

In a third step, select $\delta^* > 0$ with $\delta^* < \delta$ (cf. (2.4)) such that, for some $L_\phi > 0$, $|\phi(\nu) - \phi(\mu)| \leqslant L_\phi W_1(\mu, \nu)$ and $\psi(\nu) \subset \psi(\mu) + B_m$ whenever $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $W_1(\mu, \nu) < \delta^*$ (cf. theorem 2.4 for the specification (2.2)). Let $\nu \in \mathcal{M}_1(\mathbb{R}^s)$ such that $W_1(\mu, \nu) < \delta^*$ and $x \in \psi(\mu)$, $\tilde{x} \in \psi(\nu)$. Then we have

$$\begin{aligned}
|c^\mathsf{T}x - c^\mathsf{T}\tilde{x}| &= |\phi(\mu) - x^\mathsf{T}Hx - Q_\mu(Ax) - \phi(\nu) + \tilde{x}^\mathsf{T}H\tilde{x} + Q_\nu(A\tilde{x})| \\
&\leqslant |\phi(\mu) - \phi(\nu)| + |x^\mathsf{T}Hx - \tilde{x}^\mathsf{T}H\tilde{x}| + |Q_\nu(A\tilde{x}) - Q_\mu(A\tilde{x})| \\
&\quad + |Q_\mu(A\tilde{x}) - Q_\mu(Ax)| \\
&\leqslant L_\phi W_1(\mu, \nu) + L_{\tilde{h}}W_1(\mu, \nu) + |x^\mathsf{T}Hx - \tilde{x}^\mathsf{T}H\tilde{x}| \\
&\quad + |Q_\mu(A\tilde{x}) - Q_\mu(Ax)|
\end{aligned}$$

(the second term in the above estimate is a consequence of the Kantorovich–Rubinstein theorem)

$$\begin{aligned}
&= (L_\phi + L_{\tilde{h}})W_1(\mu, \nu) + |(H^{1/2}x + H^{1/2}\tilde{x})^\mathsf{T}(H^{1/2}x - H^{1/2}\tilde{x})| \\
&\quad + |Q_\mu(A\tilde{x}) - Q_\mu(Ax)|,
\end{aligned}$$

and hence

$$| c^{\mathrm{T}}x - c^{\mathrm{T}}\tilde{x} | \leqslant L_1 W_1(\mu, \nu) + L_2 (\| H^{1/2}x - H^{1/2}\tilde{x} \| + \| Ax - A\tilde{x} \|), \quad (2.6)$$

with suitable constants $L_1 > 0$, $L_2 > 0$.

For the last estimates we used that both x and $\tilde{x}$ are contained in the compact set $\psi(\mu) + B_m$ and that $Q_\mu(\cdot)$, as a convex function, is Lipschitzian on the compact set $A(\psi(\mu) + B_m)$. Now we are able to perform the final step of the proof.

Let $\nu \in \mathcal{M}_1(\mathbb{R}^s)$, $W_1(\mu, \nu) < \delta^*$. By (2.5) we have

$$d_{\mathrm{H}}(\psi(\nu), \psi(\mu)) \leqslant \sup_{\substack{y \in P(\psi(\mu)) \\ \tilde{y} \in P(\psi(\nu))}} d_{\mathrm{H}}(C(y), C(\tilde{y})).$$

By Hoffman's theorem (cf. [28], p. 760) there exists $\hat{L} > 0$ such that we can continue the estimate as follows

$$\leqslant \hat{L} \cdot \sup_{\substack{y \in P(\psi(\mu)) \\ \tilde{y} \in P(\psi(\nu))}} \| \overline{A}y - \overline{A}\tilde{y} \|$$

$$= \hat{L} \cdot \sup_{\substack{x \in \psi(\mu) \\ \tilde{x} \in \psi(\nu)}} \| \overline{A}x - \overline{A}\tilde{x} \|.$$

Using (2.6) and (2.4) we continue this estimate

$$\leqslant \hat{L} \cdot \sup_{\substack{x \in \psi(\mu) \\ \tilde{x} \in \psi(\nu)}} (| c^{\mathrm{T}}x - c^{\mathrm{T}}\tilde{x} | + \| H^{1/2}x - H^{1/2}\tilde{x} \| + \| Ax - A\tilde{x} \|)$$

$$\leqslant \hat{L} \cdot L_1 \cdot W_1(\mu, \nu) + \hat{L}(L_2 + 1) \sup_{\substack{x \in \psi(\mu) \\ \tilde{x} \in \psi(\nu)}} (\| H^{1/2}x - H^{1/2}\tilde{x} \| + \| Ax - A\tilde{x} \|)$$

$$\leqslant \hat{L} \cdot L_1 \cdot W_1(\mu, \nu) + \hat{L}(L_2 + 1)L^{1/2} \cdot \sqrt{2} \cdot W_1(\mu, \nu)^{1/2}$$

$$\leqslant L^* \cdot W_1(\mu, \nu)^{1/2}. \quad \square$$

Remark 2.8

When relating theorem 2.7 to specific instances of (2.2) it is crucial to verify the strong convexity of the functional Q_μ. For the case that Q_μ is differentiable with locally Lipschitzian gradient, general sufficient conditions for strong convexity of Q_μ were derived in [35] (proposition 3.1, lemma 3.2). A sufficient condition for Q_μ to have locally Lipschitzian gradient is that for each non-singular transformation $B \in L(\mathbb{R}^s, \mathbb{R}^s)$ the distribution function of $\mu \circ B$ is locally Lipschitzian (cf. the analysis in [35]). In theorem 3.5 in [35] it is shown that Q_μ is strongly convex on a convex compact set $V \subset \mathbb{R}^s$, if, in addition, $\mu \in \mathcal{M}_1(\mathbb{R}^s)$ has a density Θ such that there exists $r > 0$ with $\Theta(t) \geqslant r$ for all t from an open set $U \supset V$ and either $W \in L(\mathbb{R}^{s+1}, \mathbb{R}^s)$, $q \notin \mathrm{im} \, W^{\mathrm{T}}$ or $W \in L(\mathbb{R}^{2s}, \mathbb{R}^s)$, $W = (I, -I)$, $q^+ + q^- > 0$ (componentwise). The case $W = (I, -I)$ is referred to as simple recourse; here, $I \in L(\mathbb{R}^s, \mathbb{R}^s)$ denotes the identity and $q^+, q^- \in \mathbb{R}^s$ are the subvectors of q which correspond to the splitting of W into I and $-I$.

Remark 2.9

Several examples show that the above result is the best possible in various respects. All the examples are based on the specification

$$\tilde{Q}(x, z) = \min\{ y^+ + y^- : y^+ - y^- = z - Ax, \ y^+ \geqslant 0, \ y^- \geqslant 0\}$$

and on the sequence $\mu_n \in \mathscr{P}(\mathbb{R})$ which converges in W_1 to $\mu_0 \in \mathscr{P}(\mathbb{R})$ – the uniform distribution on $[-1/2, 1/2]$ – and which is given by

$$\mu_n = c_n \mu_{0n} + \varepsilon_n \delta_{(-\varepsilon_n)} + \varepsilon_n \delta_{\varepsilon_n} + c_n \tilde{\mu}_{0n} \quad (n > 4),$$

where $\varepsilon_n = n^{-1/2}$, $c_n = \frac{1}{2} - \varepsilon_n$, μ_{0n} and $\tilde{\mu}_{0n}$ are the uniform distributions on $[-1/2, -\varepsilon_n]$ and $[\varepsilon_n, 1/2]$, respectively, $\delta_{(-\varepsilon_n)}$ and δ_{ε_n} the measures with unit mass at $-\varepsilon_n$ and ε_n, respectively.

Using the formula for W_1 in remark 2.11 below we obtain

$$W_1(\mu_n \mu_0) = n^{-1}.$$

Example 4.5 in [35] (where $g \equiv 0$, $C = \mathbb{R}$, $A = 1$) shows that the exponent $1/2$ on the right-hand side of the estimate is optimal, namely we have in that case $\psi(\mu_n) = [-n^{-1/2}, n^{-1/2}]$, $\psi(\mu_0) = \{0\}$. Thus $d_H(\psi(\mu_n), \psi(\mu_0)) = n^{-1/2}$.

Example 4.6 in [35] (where $g \equiv 0$, $C = \{x \in \mathbb{R}^2 : -x_1 + (x_2)^2 \leqslant 0\}$, $A \in L(\mathbb{R}^2, \mathbb{R}^1)$, $A = (1, 0)$ shows that the result does not hold for a general convex set C. Here, we have $\psi(\mu_0) = \{0\}$, and the points $(n^{-1/2}, n^{-1/4})$ belong to $\psi(\mu_n)$. Thus $d_H(\psi(\mu_n), \psi(\mu_0)) \geqslant n^{-1/4}$.

To see that the result is lost for a general convex function g, let g be a convex function which is zero on $\{x \in \mathbb{R}^2 : -x_1 + (x_2)^2 \leqslant 0\}$ and greater than zero outside this set and put $C = \mathbb{R}^2$, $A \in L(\mathbb{R}^2, \mathbb{R}^1)$, $A = (1, 0)$. Again $\psi(\mu_0) = \{0\}$ and $(n^{-1/2}, n^{-1/4}) \in \psi(\mu_n)$, yielding $d_H(\psi(\mu_n), \psi(\mu_0)) \geqslant n^{-1/4}$.

Remark 2.10

In [33], the stability of recourse models was investigated with respect to the bounded Lipschitz metric β (cf. [9] for its definition). For instance, theorem 3.3 in [33] is a quantitative continuity result for optimal values of stochastic programs with linear and quadratic recourse, respectively, stating that

$$|\phi(\mu) - \phi(\nu)| \leqslant L \cdot \beta(\mu, \nu)^{1-1/p}$$

for some constant $L > 0$ and for all ν close to μ. The main assumption in that result, however, is that the moments

$$M_{2p}(\nu) := \int_{\mathbb{R}^s} \| z \|^{2p} \nu(\mathrm{d}z) \quad (p > 1)$$

are uniformly bounded. Hence, the moment conditions in the present section are weaker and more natural, because they correspond to the assumptions on the existence of finite second moments in the underlying basic theory for recourse models (cf. e.g. [16]). We also refer to the relation between W_1 and β mentioned in remark 4.8 in [35].

Remark 2.11

Explicit formulae for L_p-Wasserstein metrics are known for probability measures on $\mathbb{R}$ (i.e. for $s = 1$) and for multivariate probability distributions ($s > 1$) belonging to special classes (cf. [14,15]) in the case $p = 2$. It is known (e.g. [26]) that, for μ, $\nu \in \mathcal{M}_1(\mathbb{R})$ (where $\mathbb{R}$ is equipped with the natural distance $|\cdot - \cdot|$),

$$W_p(\mu, \nu) = \left(\int_0^1 \left| F_\mu^{-1}(t) - F_\nu^{-1}(t) \right|^p \, dt \right)^{1/p} \quad (p \geqslant 1),$$

$$W_1(\mu, \nu) = \int_{-\infty}^{+\infty} | F_\mu(t) - F_\nu(t) | \, dt,$$

where F_μ and F_ν are the distribution functions of μ and ν, respectively. This has an important consequence for recourse models having a separability structure (e.g. the above mentioned linear simple recourse case and the box-diagonal case for quadratic recourse, cf. e.g. proposition 4.7 in [33]). Since these models only depend on the one-dimensional marginal distributions of the involved measures, we may assume that μ and ν are measures with independent one-dimensional marginal distributions. If we consider the norm

$$\| z \|_p := \left(\sum_{i=1}^s | z_i |^p \right)^{1/p} \quad (p \geqslant 1)$$

on $\mathbb{R}^s$, then we obtain in that situation

$$W_p(\mu, \nu) = \left[\inf \left\{ \sum_{i=1}^s \int_{\mathbb{R}^s \times \mathbb{R}^s} | z_i - \tilde{z}_i |^p \eta(dz, d\tilde{z}) : \eta \in D(\mu, \nu) \right\} \right]^{1/p}.$$

Now, let μ_i, ν_i ($i = 1, \ldots, s$) denote the marginal distributions of μ, ν and let $\eta_i \in D(\mu_i, \nu_i)$ be chosen such that

$$W_p(\mu_i, \nu_i) = \left(\int_{\mathbb{R} \times \mathbb{R}} | z_i - \tilde{z}_i |^p \eta_i(dz_i, d\tilde{z}_i) \right)^{1/p}.$$

This is possible according to proposition 1 in [15]. Hence,

$$W_p(\mu, \nu) = \left(\sum_{i=1}^s \int_{\mathbb{R} \times \mathbb{R}} | z_i - \tilde{z}_i |^p \eta_i(dz_i, d\tilde{z}_i) \right)^{1/p}$$

$$= \left(\sum_{i=1}^s W_p(\mu_i, \nu_i)^p \right)^{1/p} = \left(\sum_{i=1}^s \int_0^1 \left| F_{\mu_i}^{-1}(t) - F_{\nu_i}^{-1}(t) \right|^p \, dt \right)^{1/p}.$$

Hence, for recourse models with separability structure we arrive at stability results with respect to distances which are computable in practical situations.

We end this section with an application to asymptotic properties of statistical estimators in stochastic programs. Let z_1, z_2, ... be independent random varia-

bles (with values in $\mathbb{R}^s$) on some probability space $(\Omega, \mathscr{A}, P)$ with identical distribution $\mu \in \mathscr{P}(\mathbb{R}^s)$. We consider the empirical measures

$$\mu_n(\omega) := \frac{1}{n} \sum_{i=1}^n \delta_{z_i(\omega)} \quad (\omega \in \Omega, \, n \in \mathbb{N}).$$

(δ_z denotes the measure with unit mass at $z \in \mathbb{R}^s$.) Then it is known that, if $\mu \in \mathscr{M}_p(\mathbb{R}^s)$ ($p \geqslant 1$),

$$W_p(\mu, \mu_n(\omega)) \to 0 \quad P\text{-almost surely}$$

(cf. chapter 4.1 in [26]). Convergence rates for the mean of $W_p(\mu, \mu_n(\cdot))$ ($n \in \mathbb{N}$) can be derived from [8,26]. We show how to combine our stability results with speed-of-convergence properties for empirical measures to obtain asymptotic properties of statistical estimators in stochastic programs.

COROLLARY 2.12

Let the assumptions of theorem 2.7 be fulfilled and let $\mu \in \mathscr{M}_p(\mathbb{R}^s)$ for each $p \geqslant 1$. Let C be bounded. Then $E[d_H(\psi(\mu), \psi(\mu_n(\omega)))] = O(n^{-1/2K})$ for each $K > \max\{2, s\}$.

Proof

Let $a_n(\omega) := d_H(\psi(\mu), \psi(\mu_n(\omega)))$ and put

$$\Omega_n := \{\omega : W_1(\mu, \mu_n(\omega)) < \delta^*\} \quad (n \in \mathbb{N}),$$

where δ^* and (later on) L^* are chosen as in theorem 2.7. Then

$$E[a_n(\omega)] = \int_{\Omega_n} a_n(\omega) P(d\omega) + \int_{\bar{\Omega}_n} a_n(\omega) P(d\omega)$$

$$\leqslant L^* \cdot E[W_1(\mu, \mu_n(\omega))^{1/2}]$$

$$+ \operatorname{diam} C \cdot P(\{\omega : W_1(\mu, \mu_n(\omega)) \geqslant \delta^*\})$$

$$\leqslant L^* (E[W_1(\mu, \mu_n(\omega))])^{1/2} + \operatorname{diam} C \cdot \frac{1}{\delta^*} E[W_1(\mu, \mu_n(\omega))],$$

where we used Hölder's and Chebyshev's inequalities, and diam C denotes the diameter of C. Hence we obtain

$$E[a_n(\omega)] = O((E[W_1(\mu, \mu_n(\omega))])^{1/2})$$

and it remains to study the convergence rate of $E[W_1(\mu, \mu_n(\omega))]$. Theorem 6 in [26] reads in our specific situation (by putting there $h \equiv 1$, $\alpha := 1$, $a := 0$): If $\mu \in \mathscr{M}_r(\mathbb{R}^s)$ for some $r > 1$ and if there exist numbers $k > 2$ and $K_1 > 0$ such that $N(\mu, \varepsilon, \varepsilon^{k/(k-2)}) \leqslant K_1 \varepsilon^{-k}$ for each $\varepsilon \in (0, 1]$, then

$$E[W_1(\mu, \mu_n(\omega))] = O(n^{-(1-1/r)/k}).$$

(Here $N(\mu, \varepsilon, \delta)$ is the minimal number of sets of diameter at most 2ε which cover $\mathbb{R}^s$ except for a set A with $\mu(A) \leqslant \delta$.) Since $\mu \in \mathscr{M}_r(\mathbb{R}^s)$ for each $r > 1$, the above hypothesis holds if $k > \max\{2, s\}$ (proposition 3.4 in [8]). Hence, we have $E[W_1(\mu, \mu_n(\omega))] = O(n^{-1/K})$ if $K > \max\{2, s\}$ and the proof is complete. $\qquad \square$

3. Stability of stochastic programs with probabilistic constraints

The problem we shall consider in this section is

$$\mathbf{P}(\mu) \qquad \min\{ f(x) \colon x \in C, \ \mu\big(H_j(x)\big) \geqslant p_j, \ j = 1, \ldots, d \},$$

where f is a mapping from $\mathbb{R}^m$ into $\mathbb{R}$, C is a nonempty closed subset of $\mathbb{R}^m$, H_j is a set-valued mapping from $\mathbb{R}^m$ into $\mathbb{R}^s$ having closed graph ($j = 1, \ldots, d$), $p_j \in (0, 1)$ ($j = 1, \ldots, d$) and $\mu \in \mathscr{P}(\mathbb{R}^s)$.

In the following, we make use of the notations $p := (p_1, \ldots, p_d)^T \in \mathbb{R}^d$ and $C_v(\nu) := \{ x \in C \colon \nu(H_j(x)) \geqslant v_j, \ j = 1, \ldots, d \}$ for each $v \in \mathbb{R}^d$ and $\nu \in \mathscr{P}(\mathbb{R}^s)$. Hence problem $\mathbf{P}(\mu)$ becomes $\min\{ f(x) \colon x \in C_p(\mu) \}$.

Our approach to study stability of $\mathbf{P}(\mu)$ is to identify a suitable topology (metric) on $\mathscr{P}(\mathbb{R}^s)$ such that the mapping which assigns to each $\nu \in \mathscr{P}(\mathbb{R}^s)$ the set $C_p(\nu)$ enjoys certain continuity properties at the unperturbed measure μ. It is known (cf. the discussion in [17,33,34]) that continuity properties of that mapping (especially lower semicontinuity) with respect to the topology of weak convergence on $\mathscr{P}(\mathbb{R}^s)$ can be achieved only at measures μ with certain smoothness properties.

Extending the ideas of our earlier work (cf. [36,33,34]) we consider the following distance on $\mathscr{P}(\mathbb{R}^s)$:

$$\alpha(\mu, \nu) := \max_{j=1,\ldots,d} \sup\{ |\mu(B) - \nu(B)| \colon B \in \mathscr{B}_j \},$$

where each $\mathscr{B}_j$ is a class of Borel subsets of $\mathbb{R}^s$ such that $\{ H_j(x) \colon x \in \mathbb{R}^m \} \subset \mathscr{B}_j$ ($j = 1, \ldots, d$) and that α forms a metric. The latter holds if, for some $j \in \{1, \ldots, d\}$, $\mathscr{B}_j$ is a determining class, i.e. it has the property that if any two measures agree on $\mathscr{B}_j$ then they must agree on the whole of $\mathscr{B}(\mathbb{R}^s)$. Following e.g. [13] we call this distance "discrepancy". A first step in our stability analysis of $\mathbf{P}(\mu)$ now is

PROPOSITION 3.1

The multifunction $\nu \mapsto C_p(\nu)$ from $(\mathscr{P}(\mathbb{R}^s), \alpha)$ into $\mathbb{R}^m$ has a closed graph.

Proof

Let $G := \{ (\nu, x) \colon \nu \in \mathscr{P}(\mathbb{R}^s), \ x \in C_p(\nu) \}$ and fix (ν, x) with $\nu \in \mathscr{P}(\mathbb{R}^s)$, $x \in \mathbb{R}^m$, $(\nu_n, x_n) \in G$ ($n \in \mathbb{N}$) such that $\alpha(\nu_n, \nu) \to 0$ and $x_n \to x$. We have to show that this implies $x \in C_p(\nu)$.

Since C is closed, we have $x \in C$, and it remains to prove that $\nu(H_j(x)) \geqslant p_j$. Since the multifunctions H_j ($j = 1, \ldots, d$) have closed graphs, there holds for each $j = 1, \ldots, d$

$$H_j(x) = \bigcap_{\delta > 0} \text{cl}\left(\left\{z : z \in H_j(y), \|x - y\| \leqslant \delta\right\}\right)$$

and, hence,

$$\nu(H_j(x)) = \inf_{\delta > 0} \nu\left(\text{cl}\left\{z : z \in H_j(y), \|x - y\| \leqslant \delta\right\}\right).$$

Let $\varepsilon > 0$ be chosen arbitrarily. Then there exists $\delta_0 > 0$ such that for each $j = 1, \ldots, d$ we have

$$\nu(H_j(x)) \geqslant \nu\left(\text{cl}\left\{z : z \in H_j(y), \|x - y\| \leqslant \delta_0\right\}\right) - \varepsilon/2.$$

Furthermore, there exists $n_0 \in \mathbb{N}$ such that $\alpha(\nu, \nu_n) \leqslant \varepsilon/2$ and $\|x - x_n\| \leqslant \delta_0$ for all $n \geqslant n_0$. Then we obtain for each $j = 1, \ldots, d$ and $n \geqslant n_0$

$$\nu(H_j(x))$$
$$\geqslant \nu(H_j(x)) - \nu(H_j(x_n)) - |\nu(H_j(x_n)) - \nu_n(H_j(x_n))| + \nu_n(H_j(x_n))$$
$$\geqslant -\varepsilon/2 - \alpha(\nu, \nu_n) + p_j \geqslant p_j - \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this yields the assertion. □

Now we state the central stability result of this section.

THEOREM 3.2

Let f be Lipschitz continuous on bounded subsets of $\mathbb{R}^m$ and assume that M is a CLM set for $\mathbf{P}(\mu)$ with respect to Q, where $Q \subset \mathbb{R}^m$ is bounded (i.e. $M = \psi_Q(\mu)$). Assume that, for each $x_0 \in \psi_Q(\mu)$, the function

$$(x, v) \mapsto \inf\left\{\|x - y\| : y \in C_v(\mu)\right\}$$

is Lipschitz continuous on some neighbourhood of (x_0, p). Then the multifunction ψ_Q (from $(\mathscr{P}(\mathbb{R}^s), \alpha)$ into $\mathbb{R}^m$) is upper semicontinuous at μ, and there exist constants $L > 0$, $\delta > 0$ such that

$$\varnothing \neq \psi_Q(\nu) \subset Q$$

and

$$|\phi_Q(\mu) - \phi_Q(\nu)| \leqslant L\alpha(\mu, \nu),$$

whenever $\alpha(\mu, \nu) < \delta$.

Proof

Again we show that the assertion follows from lemma A.2 in the appendix. In view of the assumptions and of proposition 3.1 it remains to verify condition (c)

of lemma A.2. In our terminology condition (c) reads as follows: For each $x_0 \in \psi_Q(\mu)$ there exist a neighbourhood $U = U(x_0)$, positive reals δ_0 and λ such that the following holds for all $\nu \in \mathscr{P}(\mathbb{R}^s)$ with $\alpha(\mu, \nu) < \delta_0$;

$$x \in C_p(\mu) \cap U \text{ implies } d\big(x, C_p(\nu)\big) \leqslant \lambda \cdot \alpha(\mu, \nu)$$

and

$$x \in C_p(\nu) \cap U \text{ implies } d\big(x, C_p(\mu)\big) \leqslant \lambda \cdot \alpha(\mu, \nu).$$

(Here $d(x, A) := \inf\{ \|x - y\| : y \in A \}$ for each subset A of $\mathbb{R}^m$.) So let $x_0 \in \psi_Q(\mu)$ be fixed.

Our last assumption (in theorem 3.2) is equivalent to the existence of neighbourhoods U of x_0, V of p and of a constant $\lambda > 0$ such that $d(x, C_v(\mu))$ is finite and

$$d\big(x, C_{v_1}(\mu)\big) \leqslant d\big(x, C_{v_2}(\mu)\big) + \lambda \| v_1 - v_2 \|_\infty,$$

for all $x \in U$ and all $v, v_1, v_2 \in V$. Here, $\| \cdot \|_\infty$ denotes the max-norm on $\mathbb{R}^d$. This holds, since $d(\cdot, C_v(\mu))$ is Lipschitzian with modulus 1 for each $v \in V$.

Let $\delta_0 > 0$ such that $\{ v : \| v - p \| \leqslant \delta_0 \} \subset V$, and let $\nu \in \mathscr{P}(\mathbb{R}^s)$ such that $\alpha(\mu, \nu) \leqslant \delta_0$. Now, observe that by definition of the distance α the following inclusions hold

$$C_{p + \alpha(\mu, \nu)}(\mu) \subset C_p(\nu) \subset C_{p - \alpha(\mu, \nu)}(\mu).$$

Let $x \in U$, set $v_1 := p + \alpha(\mu, \nu)$, $v_2 := p$ and $v_1 := p$, $v_2 := p - \alpha(\mu, \nu)$, respectively. Then we obtain

$$d\big(x, C_p(\nu)\big) \leqslant d\big(x, C_{p + \alpha(\mu, \nu)}(\mu)\big) \leqslant d\big(x, C_p(\mu)\big) + \lambda \cdot \alpha(\mu, \nu)$$

and

$$d\big(x, C_p(\mu)\big) \leqslant d\big(x, C_{p - \alpha(\mu, \nu)}(\mu)\big) + \lambda \cdot \alpha(\mu, \nu).$$

For $x \in C_p(\mu) \cap U$ and $x \in C_p(\nu) \cap U$, respectively, these estimates now yield the desired implications. $\square$

Remark 3.3

Recall that a class $\mathscr{B}$ of Borel subsets of $\mathbb{R}^s$ is a μ-uniformity class if $\sup\{ |\mu(B) - \mu_n(B)| : B \in \mathscr{B} \} \to 0$ holds for every sequence $\{\mu_n\}$ converging weakly to μ [2]. Necessary and sufficient conditions for $\mathscr{B}$ to be a μ-uniformity class may be found in [2] and in the recent paper [21]. Especially, we mention Range Rao's result (theorem 2.11 in [2]) that the class $\mathscr{B}_C := \{ B \subset \mathbb{R}^s : B \text{ convex and Borel} \}$ is a μ-uniformity class iff $\mu(\partial B) = 0$, ∂B denoting the topological boundary of B, for all $B \in \mathscr{B}_C$. If each class $\mathscr{B}_j$ ($j = 1, \ldots, d$) is a μ-uniformity class, weak convergence of $\{\mu_n\}$ to μ implies $\alpha(\mu_n, \mu) \to 0$. In this case, theorem 3.2 represents a stability result for $\mathbf{P}(\mu)$ with respect to the topology of weak convergence.

Remark 3.4

The last (and most essential) assumption in theorem 3.2 reads in the terminology of [32]: Assume that the multifunction $v \mapsto C_v(\mu)$ is pseudo-Lipschitzian at each $(x_0, p) \in \psi_Q(\mu) \times \{p\}$ (see theorem 2.3 in [32]). Rockafellar's paper contains a detailed study of pseudo-Lipschitzian multifunctions including a variety of sufficient conditions for a multifunction to have this property. It is well-known that constraint qualifications of mathematical programming play an important role in this context.

We shall make use of the following two results which give sufficient conditions for a multifunction to be pseudo-Lipschitzian.

LEMMA 3.5

Let Γ be a multifunction from $\mathbb{R}^d$ into $\mathbb{R}^m$ having closed convex graph. Then Γ is pseudo-Lipschitzian at each pair $(x_0, v_0) \in \Gamma(v_0) \times \mathbb{R}^d$ with $v_0 \in \text{int}(\text{dom } \Gamma)$ where dom $\Gamma := \{v \in \mathbb{R}^d : \Gamma(v) \neq \varnothing\}$.

In the literature, this result is known as the Robinson–Ursescu theorem (cf. e.g. [29]).

The next result is a particular case of theorem 3.2 in [32].

LEMMA 3.6

Let $\Gamma(v) := \{x \in C : G(v, x) \leqslant 0\}$ $(v \in \mathbb{R}^d)$, where $C \subset \mathbb{R}^m$ is closed and $G: \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^l$ is locally Lipschitzian. Let $v_0 \in \mathbb{R}^d$, $x_0 \in \Gamma(v_0)$ and assume that the only $y \in \mathbb{R}^l$ such that $y_i \geqslant 0$, $y_i G_i(v_0, x_0) = 0$, $i = 1, \ldots, l$, and $\{-A^T y : A \in \pi_2 \, \partial G(v_0, x_0)\} \cap N_C(x_0) \neq \varnothing$, is $y = 0$. Then Γ is pseudo-Lipschitzian at (x_0, v_0).

(Here, $\partial G(v_0, x_0)$ denotes the Clarke generalized Jacobian, cf. [6], containing $l \times (d + m)$ matrices, and $\pi_2 \, \partial G(v_0, x_0)$ denotes the set of all $l \times m$ matrices A such that there exists an $l \times d$ matrix $\tilde{A}$ with $(\tilde{A}, A) \in \partial G(v_0, x_0)$. $N_C(x_0)$ denotes the Clarke normal cone to C at x_0.)

Next, we study stability of specific probabilistic constrained programs by revealing conditions on the measure μ and on the multifunctions H_j $(j = 1, \ldots, d)$ to have the mapping $v \mapsto C_v(\mu)$ pseudo-Lipschitzian at some pair (x_0, p).

First, we consider a situation where f, C, H_j $(j = 1, \ldots, d)$ and μ have certain convexity properties. Following [7] we introduce the notation $M_r^\lambda(a, b)$, for all $r \in \mathbb{R} \setminus \{0\}$, $\lambda \in [0, 1]$, $a, b \geqslant 0$, defined by

$$M_r^\lambda(a, b) := \begin{cases} \left(\lambda a^r + (1 - \lambda) b^r\right)^{1/r} & \text{if } ab > 0, \\ 0 & \text{if } ab = 0. \end{cases}$$

This definition can be extended for $r = 0$, $r = -\infty$ in the following natural way

$$M_0^\lambda(a, b) := \lim_{r \to 0} M_r^\lambda(a, b) = a^\lambda b^{1-\lambda} \quad \text{if } ab > 0,$$

$$M_{-\infty}^\lambda(a, b) := \lim_{r \to -\infty} M_r^\lambda(a, b) = \min\{a, b\}.$$

A Borel probability measure μ on $\mathbb{R}^s$ is called r-convex, $r \in [-\infty, +\infty)$ (cf. [5,7,25]), if for each $\lambda \in [0, 1]$

$$\mu(\lambda B_1 + (1 - \lambda) B_2) \geqslant M_r^\lambda(\mu(B_1), \mu(B_2))$$

holds for all Borel subsets B_1, B_2 of $\mathbb{R}^s$ such that $\lambda B_1 + (1 - \lambda) B_2$ is Borel. For $r = 0$ and $r = -\infty$, μ is also called logarithmic concave and quasi-concave, respectively [24]. Since $M_r^\lambda(a, b)$ is increasing in r, when all other variables are fixed, the sets of all r-convex measures are decreasing if r is increasing. It is known (cf. [5,7,27]) that $\mu \in \mathscr{P}(\mathbb{R}^s)$ is r-convex ($r \in [-\infty, 1/s]$) iff μ has a density Θ_μ such that

$$\Theta_\mu(\lambda x + (1 - \lambda)\tilde{x}) \geqslant M_{r/(1-rs)}^\lambda(\Theta_\mu(x), \Theta_\mu(\tilde{x}))$$

for all $\lambda \in [0, 1]$, $x, \tilde{x} \in \mathbb{R}^s$. We note that a number of multivariate distributions are r-convex for some $r \in (-\infty, 0]$, e.g. the (non-degenerate) multivariate normal distribution (cf. [5,24]).

COROLLARY 3.7

Assume that in **P(μ)** f is convex, C is convex, H_j ($j = 1, \ldots, d$) have convex graphs and that μ is r-convex with $r \in (-\infty, 0]$. Suppose there exists $\bar{x} \in C$ such that $\mu(H_j(\bar{x})) > p_j$, $j = 1, \ldots, d$ (Slater condition). Let $\psi(\mu)$ be nonempty and $Q \subset \mathbb{R}^m$ be open, bounded such that $\psi(\mu) \subset Q$.

Then ψ_Q is upper semicontinuous at μ. Furthermore, there exist constants $L > 0$, $\delta > 0$ such that $\psi_Q(\nu)$ is a nonempty set of local minimizers of **P(ν)** and

$$|\phi(\mu) - \phi_Q(\nu)| \leqslant L\alpha(\mu, \nu) \quad \text{whenever } \alpha(\mu, \nu) < \delta.$$

(Here, the classes $\mathscr{B}_j$ ($j = 1, \ldots, d$) in the definition of α may be chosen such that $\mathscr{B}_j \subseteq \mathscr{B}_C$.)

Proof

We apply theorem 3.2 noting that $\psi(\mu) = \psi_Q(\mu)$. It remains to show that the mapping $v \mapsto C_v(\mu)$ is pseudo-Lipschitzian at each pair $(x_0, p) \in \psi(\mu) \times \{p\}$. We even show that this is true for each (x_0, p) with $x_0 \in C_p(\mu)$.

To this end, we consider the multifunction

$$\Gamma(v) := \left\{ x \in C : \left[\mu(H_j(x)) \right]^r \leqslant v_j, \ j = 1, \ldots, d \right\} \quad (v \in \mathbb{R}^d),$$

where we assume without loss of generality that $r < 0$. We consider the extended-real-valued function g_j defined on $\mathbb{R}^m$ such that

$$g_j(x) := \left[\mu(H_j(x)) \right]^r.$$

Applying the fact that H_j has convex graph and μ is r-convex, we obtain for all $x_1,\ x_2 \in \mathbb{R}^m$ and $\lambda \in [0, 1]$

$$g_j(\lambda x_1 + (1 - \lambda)x_2) \leqslant \left[\mu\left(\lambda H_j(x_1) + (1 - \lambda)H_j(x_2)\right)\right]^r$$

$$\leqslant \lambda g_j(x_1) + (1 - \lambda)g_j(x_2).$$

Hence, each g_j ($j = 1, \ldots, d$) is convex, and the multifunction Γ from $\mathbb{R}^d$ into $\mathbb{R}^m$ has closed convex graph. We note that $C_v(\mu) = \Gamma(v^r)$, where $v^r := (v_1^r, \ldots, v_d^r)^T$, $v_j \in (0, 1)$, $j = 1, \ldots, d$. Moreover, since $g_j(\bar{x}) < p_j^r$ for each $j = 1, \ldots, d$, the point $(p_1^r, \ldots, p_d^r)^T$ belongs to the interior of dom Γ. Now, lemma 3.5 applies, and we conclude that Γ is pseudo-Lipschitzian at $(x_0,\ p^r)$ where $x_0 \in \Gamma(p^r) = C_p(\mu)$. This means that, for each $x_0 \in C_p(\mu)$, the mapping $(x,\ v) \mapsto d(x,\ \Gamma(v))$ is Lipschitz continuous on some neighbourhood of $(x_0,\ v_0)$ where $v_0 := p^r$. Since the mapping $(x,\ v) \mapsto (x,\ v^r)$ is Lipschitz continuous on some neighbourhood of (x_0, p), the same holds true for the composition of both mappings. Hence, $(x,\ v) \mapsto d(x,\ \Gamma(v^r)) = d(x,\ C_v(\mu))$ is Lipschitz continuous on some neighbourhood of $(x_0,\ p)$. The assertion now follows from theorem 3.2. $\square$

Remark 3.8

If the set C describing the deterministic constraints of $\mathbf{P}(\mu)$ is even compact, Q may be chosen such that $C \subset Q$ and corollary 3.7 then yields upper semicontinuity of ψ and (local) Lipschitz continuity of ϕ at μ. Corollary 3.7 extends corollary 2.1 of [34] and theorem 6 of [41]. It applies to a number of practical models which are known from the literature (see chapter 3 of [34]).

Secondly, we consider a particular stochastic program given by

$$\tilde{\mathbf{P}}(\mu) \qquad \min\{f(x)\colon x \in C,\ \mu(H(x)) \geqslant p\},$$

where f and C are convex, $p \in (0, 1)$, $s := m + 1$,

$$H(x) := \left\{z \in \mathbb{R}^{m+1}\colon \sum_{i=1}^{m} z_i x_i \geqslant z_{m+1}\right\},$$

and $\mu \in \mathscr{P}(\mathbb{R}^{m+1})$ is the multivariate normal distribution with mean $a \in \mathbb{R}^{m+1}$ and covariance matrix $S \in L(\mathbb{R}^{m+1}, \mathbb{R}^{m+1})$.

The following result is well-known [23] (see also [42]).

PROPOSITION 3.9

If $p \in [1/2, 1)$, the constraint set $C_p(\mu) = \{x \in C\colon \mu(H(x)) \geqslant p\}$ is convex and can be expressed as

$$C_p(\mu) = \left\{x \in C\colon \Phi^{-1}(p)\sigma(x) - l(x) \leqslant 0\right\},$$

where Φ is the standard normal distribution function, $\sigma^2(x)$ is a positive semidefinite quadratic form and $l(x)$ is affine linear in x.

COROLLARY 3.10

Consider $\tilde{\mathbf{P}}(\mu)$ under the above assumptions. Let $p \in (1/2, 1)$ and assume that the optimal set $\psi(\mu)$ of $\tilde{\mathbf{P}}(\mu)$ is nonempty and bounded (let $Q \subset \mathbb{R}^m$ be open, bounded such that $\psi(\mu) \subset Q$). Assume that there exists $\tilde{x} \in C$ such that $\mu(H(\tilde{x})) \in (p, 1)$.

Then ψ_Q is upper semicontinuous at μ and there exist constants $L > 0$, $\delta > 0$ such that $\psi_Q(\nu)$ is a nonempty set of local minimizers of $\tilde{\mathbf{P}}(\nu)$ and

$$|\phi(\mu) - \phi_Q(\nu)| \leqslant L\alpha_H(\mu, \nu) \quad \text{whenever } \alpha_H(\mu, \nu) < \delta.$$

(Here, $\alpha_H(\mu, \nu) := \sup\{|\mu(B) - \nu(B)|: B \text{ is a half-space in } \mathbb{R}^{m+1}\}$.)

Proof

We again apply theorem 3.2 by using lemma 3.6 to show that the mapping $\nu \mapsto C_\nu(\mu)$ is pseudo-Lipschitzian at each (x_0, p) where $x_0 \in C_p(\mu)$.

Let $g: \mathbb{R} \to \mathbb{R}$ be a locally Lipschitzian function such that $g(\nu) \geqslant 0$ for all $\nu \in \mathbb{R}$ and $g(\nu) = \Phi^{-1}(\nu)$ for all ν in some neighbourhood U of p. We consider

$$G(\nu, x) := g(\nu)\sigma(x) - l(x),$$

with σ and l as in proposition 3.9, and the multifunction

$$\Gamma(\nu) := \{x \in C: G(\nu, x) \leqslant 0\} \quad (\nu \in \mathbb{R}).$$

Note that, according to our construction, $\Gamma(\nu)$ coincides with $C_\nu(\mu)$ for $\nu \in U$.

Since G is locally Lipschitzian, we may apply lemma 3.6. Since g is continuously differentiable at p and $G(\nu, \cdot)$ is convex for each $\nu \in \mathbb{R}$, we obtain

$$\partial G(p, x_0) = (g'(p)\sigma(x_0), g(p) \partial\sigma(x_0) - \nabla_x l(x_0)),$$

where $\partial\sigma(x_0)$ is the subgradient of the convex function σ at x_0 and $\nabla_x l(x_0)$ is the gradient of l at x_0.

Now, let $x_0 \in C_p(\mu) = \Gamma(p)$. Then the constraint qualification in lemma 3.6 is satisfied if, in case $G(p, x_0) = 0$, there holds

$$-(g(p) \partial\sigma(x_0) - \nabla_x l(x_0)) \cap N_C(x_0) = \varnothing.$$

Assume $G(p, x_0) = 0$ and let $x \in \partial\sigma(x_0)$. We have to show that

$$-(g(p)x - \nabla_x l(x_0)) \notin N_C(x_0).$$

By assumption there exists $\tilde{x} \in C$ such that $\mu(H(\tilde{x})) \in (p, 1)$. This implies $G(p, \tilde{x}) < 0$, and for the directional derivative

$$\lim_{t \to +0} \frac{1}{t}(G(p, x_0 + t(\tilde{x} - x_0)) - G(p, x_0)) < 0.$$

Finally,

$$\langle -(g(p)x - \nabla_x l(x_0)), \tilde{x} - x_0 \rangle < 0,$$

and the proof is complete. □

Remark 3.11

In corollaries 3.7 and 3.10 we obtain stability with respect to a discrepancy $\alpha(\mu, \nu; \mathscr{B}) := \sup\{|\mu(B) - \nu(B)|: B \in \mathscr{B}\}$ where $\mathscr{B}$ is a subset of $\mathscr{B}_C$. Hence, remark 3.3 applies and both results assert stability with respect to the topology of weak convergence if $\mathscr{B}_C$ is a μ-uniformity class. The latter property is fulfilled in case of corollary 3.7, since a measure μ which is r-convex for some $r \in (-\infty, 0]$ also has a density. The same is true in case of corollary 3.10 if the covariance matrix S in $\tilde{\mathbf{P}}(\mu)$ is non-singular.

We also note that under certain assumptions on μ there exist estimates for the discrepancy $\alpha(\mu, \nu; \mathscr{B}_C)$ (cf. [22]).

Remark 3.12

Let $\mu \in \mathscr{P}(\mathbb{R}^s)$ and $\mu_n(\omega)$ be the empirical measure based upon n independent $\mathbb{R}^s$-valued random variables (on some probability space $(\Omega, \mathscr{A}, \mathbf{P})$) with common distribution μ. From corollaries 3.7 and 3.10 we then obtain for each $\varepsilon > 0$

$$\mathbf{P}\left(\left\{\omega \in \Omega: |\phi(\mu) - \phi_Q(\mu_n(\omega))| \geqslant \varepsilon\right\}\right)$$
$$\leqslant \mathbf{P}\left(\left\{\omega: \alpha(\mu, \mu_n(\omega); \mathscr{B}) \geqslant \delta\right\}\right) + \mathbf{P}\left(\left\{\omega: \varepsilon \leqslant L\alpha(\mu, \mu_n(\omega); \mathscr{B})\right\}\right)$$
$$\leqslant 2\mathbf{P}\left(\left\{\omega: \alpha(\mu, \mu_n(\omega); \mathscr{B}) \geqslant \min\{\delta, \varepsilon/L\}\right\}\right).$$

Estimates for $\mathbf{P}(\{\omega: \alpha(\mu, \mu_n(\omega): \mathscr{B}) \geqslant \lambda\})$ $(\lambda > 0)$ are known if $\mathscr{B}$ is a so-called Vapnik–Červonenkis (VC) class of subsets of $\mathbb{R}^s$. If $\mathscr{B}$ is a VC class, then for all $\lambda > 0$ and $n \geqslant 2$ there exists $K(\lambda) > 0$ such that

$$\mathbf{P}\left(\left\{\omega: \alpha(\mu, \mu_n(\omega); \mathscr{B}) \geqslant \lambda\right\}\right) \leqslant K(\lambda)p(n)\exp(-2\lambda^2 n),$$

where the function p grows at most polynomially in n. For this inequality, further details on VC classes and about empirical measures on general sample spaces we refer to chapter 26 in [39], where it is also stated that the set of all half-spaces of a Euclidean space is a VC class. Hence, the above remarks apply to the situation in corollary 3.10 and lead to convergence rates for optimal values when the original measure μ is estimated by empirical ones.

Remark 3.13

In [34], the counterpart of theorem 3.2 for the case $d = 1$ was applied to nonconvex problems where

$$H(x) := \{z \in \mathbb{R}^s: Ax \geqslant z\}, \quad A \in L(\mathbb{R}^m, \mathbb{R}^s).$$

We mention that using the methodology of the present section it is possible to extend these results to the case of several probabilistic constraints.

Appendix

We collect relevant prerequisites about the stability of abstract parametric programs

$$\mathbf{P}(t) \qquad \min\{G(x, t): x \in C(t)\} \quad (t \in T).$$

Here, (T, d) is a metric space, G a real-valued function on $\mathbb{R}^m \times T$ and C a multifunction from T into $\mathbb{R}^m$. Given $Q \subset \mathbb{R}^m$ we set for $t \in T$

$$C_Q(t) = C(t) \cap \text{cl } Q,$$

$$\phi_Q(t) = \inf \{ G(x, t) \colon x \in C_Q(t) \},$$

$$\psi_Q(t) = \{ x \in C_Q(t) \colon G(x, t) = \phi_Q(t) \}.$$

DEFINITION A.1 [30,20]

Given $\mathbf{P}(t_0)$ for fixed $t = t_0$ we call a nonempty subset M of $\mathbb{R}^m$ a complete local minimizing set (CLM set) for $\mathbf{P}(t_0)$ with respect to Q, if Q is an open subset of $\mathbb{R}^m$ such that $M \subset Q$ and $M = \psi_Q(t_0)$.

Examples of CLM sets are the set of global minimizers and strict local minimizers.

The following quantitative stability results for the abstract problem $\mathbf{P}(t)$ go back to results for the standard non-linear programming problem with differentiable data and (in-)equality constraints obtained by Alt [1]. In the above setting, the stability analysis is presented in Klatte [20]. A proof of lemma A.3 can also be found in [35].

LEMMA A.2

Consider the parametric program $\mathbf{P}(t)$, fix some $t_0 \in T$.
(a) Assume that M is a CLM set for $\mathbf{P}(t_0)$ with respect to Q, where $Q \subset \mathbb{R}^m$ is bounded (i.e. $M = \psi_Q(t_0)$).
(b) Let the multifunction C be closed-valued and closed at t_0.
(c) Suppose that for each pair $(x_0, t_0) \in \psi_Q(t_0) \times \{t_0\}$ there exist a neighbourhood $U = U(x_0)$, positive reals δ_0 and λ such that

$$C(t) \cap U \subset C(t_0) + \lambda \cdot d(t, t_0) \cdot B_m$$

and

$$C(t_0) \cap U \subset C(t) + \lambda \cdot d(t, t_0) \cdot B_m,$$

for all $t \in T$ such that $d(t, t_0) < \delta_0$. (B_m denotes the closed unit ball in $\mathbb{R}^m$.)
(d) Suppose there are real numbers $L_G > 0$, $\delta_G > 0$ such that $G(\cdot, t)$ is continuous at $\text{cl} Q$ and

$$|G(x, t_0) - G(\tilde{x}, t)| \leqslant L_G(\| x - \tilde{x}\| + d(t, t_0)),$$

for all $x, \tilde{x} \in \text{cl } Q$ and each $t \in T$ with $d(t, t_0) < \delta_G$.
Then we have
(i) The multifunction ψ_Q is upper semicontinuous at t_0, i.e. for each $\varepsilon > 0$ there exists $\delta' > 0$ such that $\psi(t) \subset \psi(t_0) + \varepsilon \cdot B_m$ whenever $t \in T$, $d(t, t_0) < \delta'$.
(ii) There exist constants $L > 0$ and $\delta > 0$ such that $\varnothing \neq \psi_Q(t) \subset Q$ and $|\phi_Q(t) - \phi_Q(t_0)| \leqslant L \cdot d(t, t_0)$ whenever $t \in T$, $d(t, t_0) < \delta$.

If we specify $\mathbf{P}(t)$ to the situation where $C(t) = C$ for all $t \in T$, with some nonempty, closed, convex set $C \subset \mathbb{R}^m$; and if we denote by ϕ and ψ the global optimal value and the set of global minimizers, respectively, then the following holds.

LEMMA A.3

Let $t_0 \in T$ be fixed and suppose:
(a) There exists a bounded open set $Q \subset \mathbb{R}^m$ such that $\psi(t_0) \subset Q$.
(b) There exist $x_0 \in \psi(t_0)$ and $\alpha: \mathbb{R}^m \to \mathbb{R}_+$ such that $\alpha(0) = 0$ and $G(x, t_0) \geqslant G(x_0, t_0) + \alpha(x - x_0)$ for all $x \in C \cap \mathrm{cl}\, Q$.
(c) There exist a neighbourhood $U = U(t_0)$ and a constant $L > 0$ such that $G(\cdot, t)$ is a convex function for each $t \in U$ and $|G(x, t) - G(x, t_0)| \leqslant L \cdot d(t, t_0)$ for all $x \in \mathrm{cl}\, V$ and all $t \in U$.

Then there exists a constant $\delta > 0$ such that for all $t \in T$ with $d(t, t_0) < \delta$ we have

$$\psi(t) \neq \varnothing$$

and

$$\sup_{x \in \psi(t)} \alpha(x - x_0) \leqslant 2L \cdot d(t, t_0).$$

Acknowledgement

References

[1] W. Alt, Lipschitzian perturbations of infinite optimization problems, in: *Mathematical Programming with Data Perturbations II*, ed. A.V. Fiacco (M. Dekker, New York, 1983) pp. 7–21.

[2] R.N. Bhattacharya and R. Ranga Rao, *Normal Approximation and Asymptotic Expansions* (Wiley, New York, 1976).

[3] P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968).

[4] J.R. Birge and R.J.-B. Wets, Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse, Math. Progr. Study 27 (1986) 54–102.

[5] C. Borell, Convex set functions in d-space, Periodica Math. Hung. 6 (1975) 111–136.

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

[7] S. Dancs and B. Uhrin, On a class of integral inequalities and their measure-theoretic consequences, J. Math. Anal. Appl. 74 (1980) 388–400.

[8] R.M. Dudley, The speed of mean Glivenko–Cantelli convergence, Ann. Math. Statist. 40 (1969) 40–50.

[9] R.M. Dudley, *Probabilities and Metrics*, Lecture Notes Series No. 45, Aarhus Universitet, Aarhus, Denmark (1976).

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

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

[12] Y. Ermoliev and R.J.-B. Wets (eds.), *Numerical Techniques for Stochastic Optimization* (Springer, Berlin, 1988).

[13] P. Gaenssler and W. Stute, Empirical processes: A survey of results for independent and identically distributed random variables, Ann. Prob. 7 (1979) 193–243.

[14] M. Gelbrich, On a formula for the L^2-Wasserstein metric between measures on Euclidean and Hilbert spaces, Math. Nachr. 147 (1990) 185–203.

[15] C.R. Givens and R.M. Shortt, A class of Wasserstein metrics for probability distributions, Michigan Math. J. 31 (1984) 231–240.

[16] P. Kall, *Stochastic Linear Programming* (Springer, Berlin, 1976).

[17] P. Kall, On approximations and stability in stochastic programming, in: *Parametric Optimization and Related Topics*, eds. J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička (Akademie-Verlag, Berlin, 1987) pp. 387–407.

[18] V. Kaňková, Estimates in stochastic programming-chance constrained case, Problems Control Inf. Theory, to appear.

[19] A.J. King and R.J.-B. Wets, Epi-consistency of convex stochastic programs, Research Report, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York (1989).

[20] D. Klatte, A note on quantitative stability results in nonlinear optimization, in: *Proc. 19. Jahrestagung Mathematische Optimierung*, ed. K. Lommatzsch (Seminarbericht Nr. 90, Humboldt-Universität Berlin, Sektion Mathematik, 1987) pp. 77–86.

[21] R. Lucchetti, G. Salinetti and R.J.-B. Wets, Uniform convergence of probability measures: Topological criteria, Technical Report, University of California, Davis (1989).

[22] H. Niederreiter and J.M. Wills, Diskrepanz und Distanz von Massen bezüglich konvexer und Jordanscher Mengen, Math. Zeits. 144 (1975) 125–134.

[23] C. van de Panne and W. Popp, Minimum cost cattle feed under probabilistic protein constraints, Manag. Sci. 9 (1963) 405–430.

[24] A. Prékopa. Logarithmic concave measures with applications to stochastic programming, Acta Sci. Math. 32 (1971) 301–316.

[25] A. Prékopa, Logarithmic concave measures and related topics, in: *Stochastic Programming*, ed. M.A.H. Dempster (Academic Press, London, 1980) pp. 63–82.

[26] S.T. Rachev, The Monge–Kantorovich mass transference problem and its stochastic applications, Theory Prob. Appl. 29 (1984) 647–676.

[27] Y. Rinott, On convexity of measures, Ann. Prob. 4 (1976) 1020–1026.

[28] S.M. Robinson, Stability theory for systems of inequalities, Part I: Linear systems, SIAM J. Num. Anal. 12 (1975) 754–769.

[29] S.M. Robinson, Regularity and stability for convex multivalued functions, Math. Oper. Res. 1 (1976) 130–143.

[30] S.M. Robinson, Local epi-continuity and local optimization, Math. Progr. 37 (1987) 208–223.

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

[32] R.T. Rockafellar, Lipschitzian properties of multifunctions, Nonlinear Anal., Theory, Meth. Appl. 9 (1985) 867–885.

[33] W. Römisch and R. Schultz, Distribution sensitivity in stochastic programming, Math. Progr. (1991) to appear.

[34] W. Römisch and R. Schultz, Distribution sensitivity for certain classes of chance constrained models – with application to power dispatch, J. Optim. Theory Appl., to appear.

[35] W. Römisch and R. Schultz, Stability of solutions for stochastic programs with complete recourse having $C^{1,1}$ data, Manuskript, Institut für Operations Research, Universität Zürich, Switzerland (1989).

[36] W. Römisch and A. Wakolbinger, Obtaining convergence rates for approximations in stochastic programming, in: *Parametric Optimization and Related Topics*, eds. J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička (Akademie-Verlag, Berlin, 1987) pp. 327–343.

[37] G. Salinetti, Approximations for chance-constrained programming problems, Stochastics 10 (1983) 157–179.

[38] A. Shapiro, Gâteaux, Fréchet and Hadamard directional differentiability of functional solutions in stochastic programming, Mimeograph Series No. 395, Operations Research and Statistics, Technion, Haifa, Israel (1988).

[39] G.R. Shorack and J.A. Wellner, *Empirical Processes with Applications to Statistics* (Wiley, New York, 1986).

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

[41] J. Wang, Continuity of feasible solution sets of probabilistic constrained programs, J. Optim. Theory Appl. 63 (1989) 79–89.

[42] R.J.-B. Wets, Stochastic programming: solution techniques and approximation schemes, in: *Mathematical Programming – The State of the Art 1982*, eds. A Bachem, M. Grötschel and B. Korte (Springer, Berlin, 1983) pp. 566–603.

[43] R.J.-B. Wets, Stochastic programming, in: *Handbooks in Operations Research and Management Science. Vol. 1: Optimization*, eds. G.L. Nemhauser, A.H.G. Rinnooy Kan and M.J. Todd (North-Holland, 1989) pp. 573–629.