Probability metrics and the stability of stochastic programs with recourse*

Michal Houda

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic.

Abstract

In stochastic optimization models the underlying probability measure must be very often replaced by its approximations. This leads to the investigation of the stability of such models with respect to changes in the probability measure. In this context, special attention is paid to recourse problems and the Wasserstein and Kolmogorov metrics.

Keywords: Stochastic programs with recourse, stability, Wasserstein metrics, Kolmogorov metrics.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Designing an optimization model in the stochastic programming, we usually assume that the underlaying probability distribution is known. This is not the case in practical applications: we work with some approximation or estimate, e. g. discretizations, scenario approaches, or empirical estimates.

In such cases (and the others as well) we would query about the quality of such approximation. It is obvious that a small change in probability distribution may cause large changes in the optimal value and/or in the optimal solution set. We speak then about stability analysis of stochastic programming problem with respect to changes in the probability measure. Henceforth, we quantify these changes by

*This research was supported by the Grant Agency of the Czech Republic under Grants 402/01/0539, 402/02/1015 and 402/01/0034.
a suitably chosen distance, a (pseudo-)metrics in a selected space of probability measures (we speak in short about probability metrics).

In Section 2 we present a review of definitions and basic properties of the Wasserstein, Kolmogorov, and some other metrics employed elsewhere in the stability analysis of stochastic programming problem. In the third section, we introduce a recourse optimization model, and present some achieved results: we intend to compare the Wasserstein and Kolmogorov metrics with respect to their application in the stability of stochastic programs.

2 Probability metrics: a review

Let $\mathcal{P}(\Xi)$ be the space of all Borel probability measures defined on some Borel set $\Xi \subset \mathbb{R}^s$. In this section, we introduce a set of probability metrics defined on a suitable subset of $\mathcal{P}(\Xi)$. We focus primarily on the Wasserstein and Kolmogorov metrics which we refer later in the paper. We refer to the book of Rachev [1991] for further details and references.

Choosing an appropriate metrics is still another important task with regard to the stability of stochastic programs: there are examples where the inappropriate selection gives bad results.

2.1 Kolmogorov metrics

The Kolmogorov (Kolmogorov–Smirnov) metrics is defined on $\mathcal{P}(\Xi)$ by

$$K(\mu, \nu) := \sup_{z \in \Xi} |F(z) - G(z)|,$$

Figure 1: Kolmogorov metrics (one-dimensional case).

The Kolmogorov (Kolmogorov–Smirnov) metrics is defined on $\mathcal{P}(\Xi)$ by

$$K(\mu, \nu) := \sup_{z \in \Xi} |F(z) - G(z)|,$$
μ and ν are elements of \( \mathcal{P}(\Xi) \) and \( F, G \) their corresponding distribution functions. A considerable advantage of this metrics is the computational simplicity when it is applied to workaday optimization problems. See Figure 1 for a better idea.

On the other hand, the Kolmogorov metrics is not a good choice when we deal, for example, with approximation of unknown mass points of a discrete distribution, as one can see in Figure 2. For another case we refer to Dupačová, Römisch [1998], where the authors measure changes in the optimal value arising from including an additional scenario to the problem; the Kolmogorov metrics does not reflect (under the cited conditions) the position of the new scenario, compared to the expected behaviour of a suitable metrics.

\[
K(\mu, \nu) = \infty \quad \Rightarrow \quad W_1(\mu, \nu) = \varepsilon
\]

Figure 2: Approximating a degenerated distribution with unknown mass point.

### 2.2 Wasserstein metrics

Consider for \( p \in [1; +\infty) \) the space \( \mathcal{P}_p(\Xi) := \left\{ \nu \in \mathcal{P}(\Xi) : \int_{\Xi} \|\xi\|^p \nu(d\xi) < +\infty \right\} \) where \( \|\cdot\| \) denotes the Euclidean norm. For \( \mu, \nu \in \mathcal{P}_p(\Xi) \), we define the \( p \)-Wasserstein (Kantorovich) metrics by

\[
W_p(\mu, \nu) := \left[ \inf \left\{ \int_{\Xi \times \Xi} \|\xi - \tilde{\xi}\|^p \eta(d\xi \times d\tilde{\xi}) : \eta \in D(\mu, \nu) \right\} \right]^{1/p}
\]

where \( D(\mu, \nu) \) is the set of all probability measures (of \( \mathcal{P}(\Xi \times \Xi) \)), for which \( \mu \) and \( \nu \) are marginal distributions. In one-dimensional case, the value is computable: it can be shown (see Vallander [1973]), that

\[
W_p(\mu, \nu) = \left( \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt \right)^{1/p} \quad \text{and} \quad W_1(\mu, \nu) = \int_{-\infty}^{+\infty} |F(t) - G(t)| dt
\]
This is illustrated clearly in Figure 3. In larger dimensions, the main difficulty is the complicated or even impossible computation. But also in the one-dimensional space of probability measures, we have to take care of such cases as illustrated in Figure 4, where the functions $F, G$ take forms of $F(z) := \varepsilon/(1 - z)$ on $(-\infty; 0)$ for $\varepsilon \in (0; 1)$, $G(z) := 0$ on the same interval, $F(z) = G(z)$ (arbitrary) on $[0; +\infty)$.

![Figure 3: Wasserstein metrics (one-dimensional case).](image)

![Figure 4: Kolmogorov and Wasserstein metrics applied to a function with heavy tail.](image)

2.3 Distances having $\zeta$-structure

The theory below is introduced by Zolotarev [1983] and developed in Rachev [1991]. Let $\mathcal{F}$ be some class of measurable functions $f : \Xi \to \mathbb{R}$; consider a suitable class $\mathcal{P}_{\mathcal{F}} \subset \mathcal{P}(\Xi)$ and $\mu, \nu \in \mathcal{P}_{\mathcal{F}}$. We denote as Zolotarev’s pseudometrics (or distance having $\zeta$-structure) on $\mathcal{P}_{\mathcal{F}}$ the value

$$
\bar{d}_{\mathcal{F}}(\mu, \nu) := \sup_{f \in \mathcal{F}} \left| \int_{\Xi} f(\xi) \mu(d\xi) - \int_{\Xi} f(\xi) \nu(d\xi) \right|
$$
where $\mathcal{F}$ and $\mathcal{P}_\mathcal{F}$ are chosen such that the integrals always have a sense. If $d_{\mathcal{F}}$ is finite and the class $\mathcal{F}$ is rich enough to ensure that $d_{\mathcal{F}}(\mu, \nu) = 0$ implies $\mu = \nu$, then $d_{\mathcal{F}}$ is a metrics.

The class $\mathcal{F}$ is chosen in a way that corresponds to the shape of functions with which we are dealing with the considered problem. For example, a large class of optimization problems is connected with Lipschitz functions. Denote the first order Lipschitz constant by

$$L_1(f) := \inf \{ L : |f(\xi) - f(\tilde{\xi})| \leq L\|\xi - \tilde{\xi}\| \forall \xi, \tilde{\xi} \in \Xi \}$$

and consider the class $\mathcal{F}_1 := \{ f : \Xi \to \mathbb{R} : L_1(f) \leq 1 \}$. The resulting distance having $\zeta$-structure defined on $\mathcal{P}_{\mathcal{F}_1} := \mathcal{P}_1(\Xi)$ is the Wasserstein metrics (due to Kantorovich–Rubinstein theorem, see RACHEV [1991]).

Sometimes we have to work with functions that grow faster than linearly. Therefore we define a Lipschitz constant of order $p \in [1; +\infty)$ by

$$L_p(f) := \inf \{ L : |f(\xi) - f(\tilde{\xi})| \leq L\|\xi - \tilde{\xi}\| \max\{1, \|\xi\|^{p-1}, \|\tilde{\xi}\|^{p-1}\} \forall \xi, \tilde{\xi} \in \Xi \},$$

the corresponding class $\mathcal{F}$ is then $\mathcal{F}_p := \{ f : \Xi \to \mathbb{R} : L_p(f) \leq 1 \}$ and the resulting metrics, called the $p$-Fortet–Mourier metrics, is defined by $\zeta_p(\mu, \nu) := d_{\mathcal{F}_p}(\mu, \nu)$. It is obvious that it coincides with the Wasserstein metrics for the case of $p = 1$.

When constructing a scenario tree within a discretization process, one discusses the problem how to minimize the chosen metrics. In the case of the $p$-Fortet–Mourier metrics with $p > 1$, the problem could be transformed to the minimization of the 1-Fortet–Mourier metrics, i.e. the 1-Wasserstein metrics. For details see PFLUG [2001].

In early papers concerning the stability with respect to the probability measures, a more complicated distance was considered, but which also holds the $\zeta$-structure. It deals with bounded Lipschitz functions defining “bounded Lipschitz” constant by

$$\|f\|_{BL} := \sup_{\xi \in \Xi} \|f(\xi)\| + \sup_{\substack{\xi, \tilde{\xi} \in \Xi \\ \xi \neq \tilde{\xi}}} \frac{|f(\xi) - f(\tilde{\xi})|}{\|\xi - \tilde{\xi}\|}$$

This leads to the class $\mathcal{F}_{BL} := \{ f : \Xi \to \mathbb{R} : \|f\|_{BL} \leq 1 \}$ and the metrics $\beta := d_{\mathcal{F}_{BL}}$ called the bounded Lipschitz $\beta$-metrics. Employing of this metrics concerning the stability of stochastic programs has later been replaced by the Wasserstein metrics (see Section 3).
3 Stability results

3.1 Problem formulation

Consider a general decision model of the form

\[
\inf_{x \in \mathbb{R}^n} \int_{\Xi} f_0(x; \xi) \mu(d\xi) \quad \text{subject to} \quad \mu\{\xi \in \Xi : x \in X(\xi)\} \geq p_0
\]

where \(\mu \in \mathcal{P}(\Xi), f_0 : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}, X : \mathbb{R}^s \Rightarrow \mathbb{R}^n\) (both measurable), and \(p_0 \in [0; 1]\) is a prescribed probability level. Stochastic programs with recourse, as well as with probabilistic constraints, fit into (1). Denote \(\varphi(\mu)\) the optimal value and \(\psi(\mu)\) the optimal solution set of the problem (1).

Let now replace the original distribution \(\mu\) with another one, denoted \(\nu \in \mathcal{P}(\Xi)\) (e.g., empirical one, scenario set, etc.). The stability theory tries to answer the question how the optimal value and optimal solution set have changed. In detail, we want to find conditions under which \(\varphi(\cdot)\) and \(\psi(\cdot)\) are (semi-)continuous at \(\mu\) and its neighbourhood if possible, and further try to quantify this continuity, e.g. by some Lipschitz property of the form \(|\varphi(\mu) - \varphi(\nu)| \leq L \cdot d(\mu, \nu), d\) being a suitably chosen probability metrics. In the case of empirical estimates this task is, of course, adjusted to the probability sense.

3.2 Stability for recourse problems

First consider a problem with the complete fixed (linear) recourse (see e.g. Prékopa [1985]): minimize

\[
\begin{align*}
c^T x + \int_{\Xi} h_1(x; \xi) \mu(d\xi) & \quad \text{subject to} \quad x \in X_0, \\
\end{align*}
\]

where \(X_0 \subset \mathbb{R}^n\) is nonempty convex polyhedron, \(c \in \mathbb{R}^n\), and the linear recourse takes the form of

\[
h_1(x; \xi) := \min_{y \in \mathbb{R}^m} \{q^T y : Wy = a - Tx, y \geq 0\}
\]

where \(\xi = (q, a, T)\).

Let assume

(A1) \(\{y \in \mathbb{R}^m : Wy = b, b \geq 0\} \neq \emptyset\) for all \(b \in \mathbb{R}^r\), and

(A2) \(\{u \in \mathbb{R}^r : W^T u \leq q\} \neq \emptyset\).

The first results with the \(\beta\)-metrics applied to a study of stability of the problem (2) appeared in Römisch, Wakolbinger [1987] and were later extended in
Römisch, Schultz [1991a], as a special case of a general stability theorem. They have proved a Hölder continuity of optimal value and optimal solution sets — but under some rather unpleasant conditions, for example, the original distribution \( \mu \) must have finite the generalized moment \( M_{2p}(\mu) \) (let emphasize the order of \( 2p \)).

The generalized moment is defined by
\[
M_p(\nu) := (\int_{\Xi} (L(\|\xi\|) \cdot \|\xi\|)^p \nu(d\xi))^{1/p}
\]
for \( p \in [1; +\infty) \) and \( \nu \in \mathcal{P}(\Xi) \), where \( L : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) is some continuous increasing function.

Later Römisch, Schultz [1991b] have weakened significantly this assumption, and got (better) Lipschitz estimates with the Wasserstein metrics. The case of the complete linear recourse is studied in Römisch, Schultz [1993], where the next theorem is proved:

**Theorem 1** Consider the problem (2) with (A1), (A2) and assume (A3): \( q \) or \((a,T)\) are non-stochastic.

Let \( \mu \in \mathcal{P}_1(\Xi) \) and assume that \( \psi(\mu) \) is a nonempty and bounded set. Then \( \psi \) is (Berge) upper semi-continuous at \( \mu \) with respect to \((\mathcal{P}_1(\Xi), W_1)\), and there exist constants \( \delta_1 > 0 \), \( L_1 > 0 \) such that for any \( \nu \in \mathcal{P}_1(\Xi) \) for which \( W_1(\mu, \nu) < \delta_1 \) we have that \( \psi(\nu) \neq \emptyset \) and
\[
|\varphi(\mu) - \varphi(\nu)| \leq L_1 W_1(\mu, \nu).
\]

For a definition of Berge’s upper semi-continuity, see e.g. Bank et al. [1982]. If we do not take into account assumption (A3), we would get a similar proposition with \( \mathcal{P}_2(\Xi) \) and the metrics \( W_2 \) (see Römisch, Schultz [1991b] again).

In Houda [2001], a possible nonlinear generalization of Theorem 1 has appeared. The task is done by a gentle modification of the assumptions and the proof of Theorem 1.

**Theorem 2** Consider a program of minimizing \( \int_{\Xi} [c(x; \xi) + h(x; \xi)] \mu(d\xi) \) subject to \( x \in X_0 \), where \( X_0 \subset \mathbb{R}^n \) is a compact set, the functions \( c \) and \( h \) are uniformly continuous on \( \mathbb{R}^n \times \mathbb{R}^s \), and the functions \( c(x; \cdot), h(x; \cdot) \) are Lipschitz for all \( x \in X_0 \) with constants \( L_c, L_h \); let denote \( L := L_c + L_h \). Then \( \psi \) is (Berge) upper semi-continuous at \( \mu \) with respect to \((\mathcal{P}_1(\Xi), W_1)\), and for any \( \nu \in \mathcal{P}_1(\Xi) \) we have that \( \psi(\nu) \neq \emptyset \) and
\[
|\varphi(\mu) - \varphi(\nu)| \leq LW_1(\mu, \nu).
\]

The important assumption of the theorem is the Lipschitz continuity of the second stage function \( h \) with respect to the (stochastic) variable \( \xi \). Of course, the program with complete fixed linear recourse with (A1)–(A3) falls into this frame.
1. Denote \( c_\mu(x) := \int_\Xi [c(x; \xi) + h(x; \xi)] \mu(d\xi) \); \( c_\mu \) is continuous on a compact set \( X_0 \), thus \( \psi(\nu) \) is nonempty set for all \( (\nu \in \mathcal{P}_1) \).

2. Let \( \nu \in \mathcal{P}_1 \), \( x_\mu \in \psi(\mu) \) and \( x_\nu \in \psi(\nu) \). The two inequalities are valid:

\[
\varphi(\mu) \leq c_\mu(x_\nu) \leq \varphi(\nu) + |c_\mu(x_\nu) - c_\nu(x_\nu)| \quad (a)
\]
\[
\varphi(\nu) \leq c_\nu(x_\mu) \leq \varphi(\mu) + |c_\mu(x_\mu) - c_\nu(x_\mu)| \quad (b)
\]

The left inequality in (a) is due to the fact that \( \varphi(\mu) \) is the optimal value of the problem, the right one to the equality \( c_\nu(x_\nu) = \varphi(\nu) \) and a trivial fact that \( a + b \leq |a + b| \). Arguments for (b) are similar.

3. From (a) and (b) we get

\[-|c_\mu(x_\mu) - c_\nu(x_\mu)| \leq \varphi(\mu) - \varphi(\nu) \leq |c_\mu(x_\nu) - c_\nu(x_\nu)|\]

and so

\[|\varphi(\mu) - \varphi(\nu)| \leq \sup_{x \in X_0} |c_\mu(x) - c_\nu(x)|\]

4. Functions \( c \) and \( h \) are Lipschitz in \( \xi \), thus \( c_0 := c + h \) have the same property and for \( x \in X_0 \) one has

\[
\left| \int_\Xi c_0(x, \xi) \mu(d\xi) - \int_\Xi c_0(x, \xi) \nu(d\xi) \right| \leq \int_{\Xi \times \Xi} |c_0(x, \xi) - c_0(x, \xi')| \eta(d\xi \times d\xi')
\]
\[
\leq L \int_{\Xi \times \Xi} \|\xi - \xi'\| \eta(d\xi \times d\xi')
\]

for arbitrary \( \eta \in D(\mu, \nu) \). So,

\[|c_\mu(x) - c_\nu(x)| \leq L \inf_{\eta \in D(\mu, \nu)} \int_{\Xi \times \Xi} \|\xi - \xi'\| \eta(d\xi \times d\xi') = LW_1(\mu, \nu)\]

and finally

\[|\varphi(\mu) - \varphi(\nu)| \leq \sup_{x \in X_0} |c_\mu(x, \xi) - c_\nu(x)| \leq LW_1(\mu, \nu)\]

5. The continuity of \( \varphi(\cdot) \) at \( \mu \) follows from the previous point. Furthermore, the continuity of \( c_\mu(\cdot) \) follows from the theorem continuity assumptions. The set \( X_0 \) is compact, so Proposition 4.2.1 from BANK ET AL. [1982] leads to the upper semi-continuity of the mapping \( \psi(\cdot) \) at \( \mu \). This completes the proof.

Now let us consider a problem with the following special structure: minimize

\[
\sum_{i=1}^{s} \int_{\Xi_i} \left( c_i(x; \xi_i) + h_i(x; \xi_i) \right) \mu_i(d\xi_i) \quad \text{subject to} \quad x \in X_0; \quad (3)
\]
for $i = 1, \ldots, n$, all $\mu_i$ are one-dimensional probability measures with support $\Xi_i \subset \mathbb{R}$, $X_0 \subset \mathbb{R}^n$ is compact, functions $c_i$ and $h_i$ are uniformly continuous on $\mathbb{R}^n \times \mathbb{R}$ and Lipschitz in $\xi_i$ with constants $L^c_i$ and $L^h_i$ for all $x \in X_0$. KAŇKOVÁ [1994] introduces a notion of the generalized simple recourse problem as a special case of (3).

Denote $\mu = (\mu_1, \ldots, \mu_n)$, $\nu = (\nu_1, \ldots, \nu_n)$, and $\Xi = \Xi_1 \times \cdots \times \Xi_s$. Then we can state two theorems with the Wasserstein and Kolmogorov metrics:

**Theorem 3** For any $\nu \in \mathcal{P}_1(\Xi)$, for the problem (3) we have that

$$|\varphi(\mu) - \varphi(\nu)| \leq \sum_{i=1}^{s} L_i W_1(\mu_i, \nu_i).$$

**Theorem 4** For $\mu \in \mathcal{P}_1(\Xi)$ satisfying

1. $\mu_i$ are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$ with $f_i(\cdot)$ as density functions;
2. $\Xi_i$ are closed (non-degenerated) intervals on $\mathbb{R}$;
3. there exist real constants $\vartheta_i > 0$ such that $\vartheta_i \leq f_i(\xi_i)$, $\xi_i \in \Xi_i$;

and for any $\nu \in \mathcal{P}_1(\Xi)$, for the problem (3) we have that

$$|\varphi(\mu) - \varphi(\nu)| \leq \sum_{i=1}^{s} \frac{1}{\vartheta_i} L_i K(\mu_i, \nu_i).$$

Theorem 3 follows from Theorem 2 (see HOUDA [2001]), the assertion of Theorem 4 is introduced and proved in KAŇKOVÁ [1994]. Both theorems give us possibility to explore the stability with respect to the metrics which can approximate the difference $|\varphi(\mu) - \varphi(\nu)|$ in better way. A numerical illustration of this possibility is given in HOUDA [2001]: the introduced example deals with linear recourse, but with the underlying random variables on the left-hand side of constraints. This case does not usually appear in the theory nor in the applications.

The recent stability results base upon probability distances having $\zeta$-structure as distances closer to the problem. Thus the Fortet–Mourier metrics takes place; it may have different asymptotic properties compared to the Wasserstein distance. This question is discussed in RACHEV, RÖMISCH [2000] and the results applied to the optimal scenario reduction problem in DUPAČOVÁ, GRÖVE-KUSKA, RÖMISCH [2000].
3.3 Stability of the optimal solution sets

We have already got in touch with the qualitative stability of the optimal solutions set $\psi(\cdot)$ in Theorem 2: we have proved the upper semi-continuity of the functional $\psi$. In order to quantify this behaviour, we need locally unique solutions, i.e. some kind of the growth condition.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called strongly convex on a convex set $V \subset \mathbb{R}^n$ with parameter $\sigma > 0$, if for any $x_1, x_2 \in V$ and any $\lambda \in [0; 1]$ we have

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - \frac{1}{2}\sigma\lambda(1 - \lambda)\|x_1 - x_2\|^2$$

This property is equivalent to the strong monotonicity of the gradient $\nabla f$ on the set $V$ (if it exists), as stated in Dentcheva et al. [1995] or Rockafellar, Wets [1997] (section 12.H). A mapping $F : \mathbb{R}^n \Rightarrow \mathbb{R}^n$ is strongly monotone if there is a constant $\sigma > 0$ such that

$$\langle y_1 - y_2, x_1 - x_2 \rangle \geq \sigma\|x_1 - x_2\|^2$$

whenever $y_1 \in F(x_1)$ and $y_2 \in F(x_2)$. In our case of $F := \nabla f$, this is further equivalent to the positive definiteness of the Hessian $\nabla^2 f$ on $V$, i.e. $\langle \nabla^2 f(x)h, h \rangle \geq \sigma\|h\|^2$ for all $x \in V$ and $h \in \mathbb{R}^n$.

**Theorem 5** Let the assumptions of Theorem 2 be satisfied, let $X_0$ be a convex set. Let for all $\xi \in \Xi$ the function $h(\cdot; \xi)$ be convex, and the function $c(\cdot; \xi)$ strongly convex on $\mathbb{R}^n$ with the parameter $\sigma > 0$. Then

$$\|\psi(\mu) - \psi(\nu)\|^2 \leq \frac{8}{\sigma}LW_1(\mu, \nu).$$

Due to the strong convexity property, we can consider $\psi(\nu)$ as a (unique) point of $\mathbb{R}^n$ and use the Euclidean norm (still having in mind that $\psi(\nu)$ is in fact the set with a single element).

**Proof** Using the triangular inequality and the property of argmin of strong convex functions from Kaňková, Lachout [1992], we get

$$\|\psi(\mu) - \psi(\nu)\|^2 \leq \frac{8}{\sigma}|\varphi(\mu) - \varphi(\nu)|$$

Finally investigating a stability theorem (in our case it is Theorem 2), we get the result.
Our next question concerns conditions which are sufficient and/or necessary for the strong convexity of a function. A straightforward way is to show that $f$ is strongly convex if and only if the function $f(\cdot) - \frac{1}{2}|\cdot|^2$ is convex: the class of such functions could not be very large.

In the case of the complete fixed linear recourse, let us consider the two additional assumptions:

(A2*) $\text{int}\{u \in \mathbb{R}^r : W^T u \leq q\} \neq \emptyset$, and

(A4*) there exists an open convex set $U \subset \Xi$, constants $\vartheta > 0$, $\delta > 0$, and a density $\overline{f}(\cdot)$ of $\mu$ such that $\vartheta \leq \overline{f}(\xi)$ for all $\xi \in \Xi$ with $\text{dist}(\xi, U) \leq \delta$.

**Theorem 6** Assuming (A1), (A2*), (A3), and (A4*) the expected value function $h_\mu$, defined by $h_\mu(x) := \int_{\Xi} h_1(x; \xi) \mu(d\xi)$, is strongly convex.

The theorem is proved and details are given in Schultz [1995]. The deep analysis concerning the strong convexity in this linear case is presented in Römisch, Schultz [1993].

Shapiro [1994] uses the so-called second-order growth condition (SOGC) of the function $h_\mu : \mathbb{R}^n \to \mathbb{R}$: there exists a constant $\alpha > 0$ and a neighbourhood $V$ of the set $M := \text{argmin}_{x \in X_0} h_\mu(x)$ such that for all $x \in X_0 \cap V$ one has

$$h_\mu(x) \geq \inf_{x' \in X_0} h_\mu(x') + \alpha \left[ \text{dist}(x, M) \right]^2$$

Under (SOGC), Shapiro gives an upper Lipschitz bound for the value of the distance $\text{dist}(x, M)$, where $x_\nu \in \text{argmin}_{x \in X_0} h_\nu(x)$, so in terms of functions $h_\mu$ and $h_\nu$ (defined as in Theorem 6). If $h_\mu$ is strongly convex on a convex neighbourhood $V$ of $M$, and $X_0$ is convex, then $M$ is singleton and (SOGC) holds.

**Acknowledgements.**

The author wish to thank Vlasta Kaňková and Karel Sladký (Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic), for insightfull reading the manuscript and many helpful comments.

**References**

Dentcheva, D., Römisch, W., and Schultz, R.: Strong convexity and directional
derivatives of marginal values in two-stage stochastic programming. In: *Stochastic
Programming – Numerical Techniques and Engineering Applications* (K. Marti
and P. Kall, eds.), Lecture Notes in Economics and Mathematical Systems 423,

Dupačová, J., Gröve-Kuska N., and Römisch, W.: Scenario reduction in stochastic
programming: An approach using probability metrics. Preprint 00-9, Institut für
Mathematik, Humboldt-Universität, Berlin, 2000. Revised version to appear in
*Mathematical Programming*.

Dupačová, J., and Römisch, W.: Quantitative stability for scenario-based stochas-
tic programs. In: *Proceedings of the Prague Stochastics ’98* (M. Hušková, P. Lachout,
and J. Á. Víšek, eds.), Union of Czech Mathematicians and Physicists,

Houda, M.: Stability and estimates in stochastic programming problems (Special
cases). Diploma Thesis (supervisor V. Kaňková), Charles University, Faculty of

Kaňková, V. and Lachout, P.: Convergence rate of empirical estimates in stochastic

Kaňková, V.: On the stability in stochastic programming — generalized simple

Pflug, G. Ch.: Scenario tree generation for multiperiod financial optimization by


Chichester, 1991.

Rachev, S. T. and Römisch, W.: Quantitative stability in stochastic program-
ming: the method of probability metrics. Preprint 00-22, Institut für Mathematik,
of Operations Research*.


