



Stability of Solutions for Stochastic Programs with Complete Recourse

Author(s): Werner Römisch and Rüdiger Schultz

Source: *Mathematics of Operations Research*, Aug., 1993, Vol. 18, No. 3 (Aug., 1993), pp. 590-609

Published by: INFORMS

Stable URL: <https://www.jstor.org/stable/3690092>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Mathematics of Operations Research*

JSTOR

STABILITY OF SOLUTIONS FOR STOCHASTIC PROGRAMS WITH COMPLETE RECOURSE

WERNER RÖMISCH AND RÜDIGER SCHULTZ

Quantitative continuity of optimal solution sets to convex stochastic programs with (linear) complete recourse and random right-hand sides is investigated when the underlying probability measure varies in a metric space. The central result asserts that, under a strong-convexity condition for the expected recourse in the unperturbed problem, optimal tenders behave Hölder-continuous with respect to a Wasserstein metric. For linear stochastic programs this carries over to the Hausdorff distance of optimal solution sets. A general sufficient condition for the crucial strong-convexity assumption is given and verified for recourse problems with separable and nonseparable objectives.

1. Introduction. This paper extends the stability analysis of [25] for solutions of certain two-stage stochastic programs. We deal with problems of the type

$$(1.1) \quad P(\mu) \quad \min\{g(x) + Q_\mu(Ax) : x \in C\}, \quad \text{where}$$

$$(1.2) \quad Q_\mu(Ax) = \int_{\mathbb{R}^s} \tilde{Q}(Ax, z) \mu(dz),$$

$$(1.3) \quad \tilde{Q}(Ax, z) = \min_y \{q^T y : Wy = z - Ax, y \geq 0\}.$$

For the data in (1.1)–(1.3) we assume that $g: \mathbb{R}^m \rightarrow \mathbb{R}$ is a convex function, μ is a (Borel) probability measure on \mathbb{R}^s , $A \in L(\mathbb{R}^m, \mathbb{R}^s)$, $z \in \mathbb{R}^s$, $C \subset \mathbb{R}^m$ nonempty, closed, convex, $q \in \mathbb{R}^m$, $W \in L(\mathbb{R}^m, \mathbb{R}^s)$. Throughout, we have the following general assumptions:

(A1) For each $t \in \mathbb{R}^s$, there exists $y \in \mathbb{R}_+^m$ (the nonnegative orthant of \mathbb{R}^m) such that $Wy = t$.

(A2) There exists $u \in \mathbb{R}^s$ such that $W^T u \leq q$.

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

Assumptions (A1)–(A3) imply that the function Q_μ given by (1.2) is defined on \mathbb{R}^s , real-valued and convex (cf. [14], [31]). Two-stage problems of type (1.1) arise as deterministic equivalents of improperly posed convex programs

$$(1.4) \quad \min\{g(x) : x \in C, Ax = z\},$$

where the right-hand side z is random. Given a realization of z , a possible deviation $z - Ax$ is compensated by additional costs $\tilde{Q}(Ax, z)$, whose expectation is added to the objective of (1.4). Accordingly, $Q_\mu(\cdot)$ in (1.2) is called expected recourse, and, since due to (A1) for each deviation $z - Ax$ there exists a compensation y , the case is referred to as complete recourse. For further details on deterministic equivalents to optimization problems with random data, consult [14], [15] and [32].

Received August 21, 1989; revised February 19, 1992.

AMS 1980 subject classification. Primary 90C15, 90C31.

IAOR 1973 subject classification. Main: Programming; Probabilistic. Cross References: Programming; Multiple Criteria.

OR/MS Index 1978 subject classification. Primary: 663 Programming/Stochastic.

Key words. Stochastic programs with recourse, quantitative stability, probability metrics, strong convexity.

In the present paper, we derive quantitative continuity properties for the set-valued mapping ψ assigning to a measure μ in the set $\mathcal{P}(\mathbb{R}^s)$ of all (Borel) probability measures on \mathbb{R}^s the set $\psi(\mu)$ of global minimizers to $P(\mu)$ (cf. (1.1)). To this end, it is necessary to have a deeper insight into the structure of the expected-recourse functional in (1.2).

Our investigations are motivated by recent developments in stochastic programming, such as

- solution procedures for recourse problems that rely on approximating multivariate continuous distributions by simpler (e.g., discrete) ones [3], [11], [17], [32],
- convergence properties for optimal solutions when the true distribution in (1.1) is replaced by parametric or nonparametric estimators [8], [9], [10], [18], [28], [30].

A unified frame for the perturbations of the underlying measures in the above applications is given by the topology of weak convergence of probability measures on the space $\mathcal{P}(\mathbb{R}^s)$ [2] including suitable metrizations [7]. The analysis in [16], [24] is based on this topologization of $\mathcal{P}(\mathbb{R}^s)$ and leads to continuity of the optimal value and upper semicontinuity of the mapping ψ . For linear g and polyhedral C quantitative stability of $P(\mu)$ was studied in [25], [27] where quite general results on the Hölder continuity of the optimal value and on the upper-semicontinuity of ψ were proved. In [25], Hölder continuity of ψ was obtained for problems where Q_μ is locally a C^2 function with certain separability structure (simple recourse) and where g had to fulfill some additional property.

Here, we are aiming at more general problems. The quantitative stability analysis in §2 is based merely on (A1)–(A3). Under strong convexity of the functional Q_μ in (1.2) which results in a growth condition for the objective of the unperturbed problem $P(\mu)$, we show that optimal tenders (i.e., transformations of the optimal solution sets with respect to the matrix A arising in (1.1)–(1.3)) behave Hölder-continuous with rate $1/2$, when equipping $\mathcal{P}(\mathbb{R}^s)$ with a suitable Wasserstein metric [13], [22]. When specifying the function g and the constraint set C in $P(\mu)$ to a linear function and a nonempty polyhedron, respectively, this Hölder estimate for optimal tenders is extended to the mapping ψ . Examples show that this extension is impossible for general convex C and that our rate of Hölder continuity is best possible.

Sufficient conditions for strong convexity of Q_μ , the central assumption for the Hölder estimates in §2, are presented in §3. Provided that Q_μ has locally Lipschitzian gradient we derive a general sufficient condition and apply it to specific recourse models. The essential novelty in this respect is that we are able to treat also recourse models with nonseparable objective. Simple recourse models, the only ones for which, up to now, estimates on perturbed optimal solutions were available [25], are treated in the present paper with the objective to show how they fit into our general theory on strong convexity. At the end of the paper we discuss sufficient conditions on μ to guarantee for existence and local Lipschitz continuity of the gradient of Q_μ .

2. Stability analysis. In this section we investigate stability of optimal solutions to the recourse problem (1.1) (under (A1)–(A3)) when the probability distribution μ is subjected to perturbations. To this end, we select a distance on a (properly chosen) subset of $\mathcal{P}(\mathbb{R}^s)$:

$$M_p(\mathbb{R}^s) := \left\{ \nu \in \mathcal{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \|z\|^p \nu(dz) < \infty \right\} \quad (p \geq 1),$$

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

where $D(\mu, \nu)$ is the set of those measures in $\mathcal{P}(\mathbb{R}^s \times \mathbb{R}^s)$ whose marginal distributions are μ and ν . $\mathbf{W}_p(\mu, \nu)$ is the L_p -Wasserstein metric [7], [13]. It is known that $(M_p(\mathbb{R}^s), \mathbf{W}_p)$ is a metric space [13] and that $\mathbf{W}_p(\mu, \mu_n) \rightarrow 0$ ($n \rightarrow \infty, \mu, \mu_n \in M_p(\mathbb{R}^s)$) holds iff the sequence (μ_n) converges weakly to μ (cf. [2]) and

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^s} \|z\|^p \mu_n(dz) = \int_{\mathbb{R}^s} \|z\|^p \mu(dz)$$

(see [22]). Let us denote $G(x, \nu) := g(x) + Q_\nu(Ax)$ for all $x \in \mathbb{R}^m, \nu \in M_1(\mathbb{R}^s)$. Hence, problem (1.1) becomes

$$(2.1) \quad P(\mu) \quad \min\{G(x, \mu) : x \in C\}.$$

For each $\nu \in M_1(\mathbb{R}^s)$ we denote

$$\varphi(\nu) := \inf\{G(x, \nu) : x \in C\} \quad \text{and} \quad \psi(\nu) := \{x \in C : G(x, \nu) = \varphi(\nu)\}.$$

Our first stability result is a consequence of Berge’s classical stability theory for parametric optimization problems.

PROPOSITION 2.1. *Suppose (A1)–(A3) and assume that $\psi(\mu)$ is nonempty, bounded. Then the set-valued mapping ψ (from $(M_1(\mathbb{R}^s), \mathbf{W}_1)$ into \mathbb{R}^m) is (Berge) upper-semicontinuous at μ and there exist constants $L_0 > 0$ and $\delta_0 > 0$ such that*

$$\psi(\nu) \neq \emptyset, \quad |\varphi(\mu) - \varphi(\nu)| \leq L_0 \mathbf{W}_1(\mu, \nu),$$

whenever $\nu \in M_1(\mathbb{R}^s), \mathbf{W}_1(\mu, \nu) < \delta_0$.

PROOF. Let $V \subset \mathbb{R}^m$ be a bounded open subset of \mathbb{R}^m such that $\psi(\mu) \subset V$. We introduce the notation

$$\begin{aligned} \varphi_V(\nu) &:= \inf\{G(x, \nu) : x \in C \cap \text{cl } V\}, \\ \psi_V(\nu) &:= \{x \in C \cap \text{cl } V : G(x, \nu) = \varphi_V(\nu)\}, \end{aligned}$$

for each $\nu \in M_1(\mathbb{R}^s)$ (cl means “closure”). Since $G(\cdot, \nu) : \mathbb{R}^m \rightarrow \mathbb{R}$ is convex, $\psi_V(\nu)$ is nonempty for all $\nu \in M_1(\mathbb{R}^s)$.

Now let $\nu \in M_1(\mathbb{R}^s), x_0 \in \psi_V(\mu) = \psi(\mu)$ and $x \in \psi_V(\nu)$. Then we have the inequalities

$$\begin{aligned} \varphi_V(\mu) &\leq G(x, \mu) \leq \varphi_V(\nu) + |G(x, \mu) - G(x, \nu)|, \\ \varphi_V(\nu) &\leq G(x_0, \nu) \leq \varphi_V(\mu) + |G(x_0, \nu) - G(x_0, \mu)|. \end{aligned}$$

This implies for each $\nu \in M_1(\mathbb{R}^s)$ the estimate

$$|\varphi_V(\mu) - \varphi_V(\nu)| \leq \sup_{x \in C \cap \text{cl } V} |Q_\mu(Ax) - Q_\nu(Ax)|.$$

(A1), (A2) imply that the marginal function (of the second stage) $h(v) := \inf\{q^T y: Wy = v, y \geq 0\}$ ($v \in \mathbb{R}^s$) is real-valued, piecewise linear and convex (cf. [29]). Consequently, h is globally Lipschitzian with some constant $L_h > 0$. Fixing some $\eta \in D(\mu, \nu)$ we continue:

$$\begin{aligned} |\varphi_V(\mu) - \varphi_V(\nu)| &\leq \sup_{x \in C \cap \text{cl} V} \left| \int_{\mathbb{R}^s} h(z - Ax) \mu(dz) - \int_{\mathbb{R}^s} h(\tilde{z} - Ax) \nu(d\tilde{z}) \right| \\ &\leq \sup_{x \in C \cap \text{cl} V} \int_{\mathbb{R}^s \times \mathbb{R}^s} |h(z - Ax) - h(\tilde{z} - Ax)| \eta(dz \times d\tilde{z}) \\ &\leq L_h \int_{\mathbb{R}^s \times \mathbb{R}^s} \|z - \tilde{z}\| \eta(dz \times d\tilde{z}). \end{aligned}$$

Since $\eta \in D(\mu, \nu)$ was chosen arbitrarily, this implies

$$|\varphi_V(\mu) - \varphi_V(\nu)| \leq L_h \mathbf{W}_1(\mu, \nu).$$

Continuity of $G(\cdot, \mu)$ and the property

$$|G(x, \mu) - G(x, \nu)| \leq L_h \mathbf{W}_1(\mu, \nu)$$

for all $x \in C \cap \text{cl} V$ and all $\nu \in M_1(\mathbb{R}^s)$ imply that ψ_V is (Berge) upper-semicontinuous at μ (cf. Theorem 4.2.1 in [1]). Hence, there exists $\delta_0 > 0$ such that $\psi_V(\nu) \subset V$ whenever $\mathbf{W}_1(\mu, \nu) < \delta_0$. Due to the convexity this implies $\psi(\nu) = \psi_V(\nu)$ and $\varphi(\nu) = \varphi_V(\nu)$ for each $\nu \in M_1(\mathbb{R}^s)$, $\mathbf{W}_1(\mu, \nu) < \delta_0$. \square

Of course, the above upper-semicontinuity of ψ also follows from more general qualitative stability results in [16], [24]. Here, we have recalled it to ask for quantitative properties along this line, which will lead us to the main stability results of the present paper. Our first theorem shows that the set-valued mapping $A\psi(\cdot)$ (which assigns to each element of $M_1(\mathbb{R}^s)$ the set of optimal tenders) behaves Hölder-continuous.

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

$$Q(\lambda z + (1 - \lambda)\tilde{z}) \leq \lambda Q(z) + (1 - \lambda)Q(\tilde{z}) - \kappa \lambda(1 - \lambda) \|z - \tilde{z}\|^2,$$

for all $z, \tilde{z} \in \mathbb{R}^s$, $\lambda \in [0, 1]$ ($\|\cdot\|$ denoting the Euclidean norm on \mathbb{R}^s).

THEOREM 2.2. *Suppose (A1)–(A3) and assume that $\psi(\mu)$ is nonempty, bounded. Let Q_μ be strongly convex on an open convex set U_μ containing $A\psi(\mu)$. Then the set $A\psi(\mu)$ is a singleton (say $A\psi(\mu) = \{Ax_*\}$ for some $x_* \in \psi(\mu)$) and there exist constants $L > 0$ and $\delta > 0$ such that $\psi(\nu) \neq \emptyset$ and*

$$\sup_{x \in \psi(\nu)} \|Ax - Ax_*\| \leq L \mathbf{W}_1(\mu, \nu)^{1/2}$$

whenever $\nu \in M_1(\mathbb{R}^s)$, $\mathbf{W}_1(\mu, \nu) < \delta$.

PROOF. Since $A\psi(\mu)$ is compact, there exists $\varrho > 0$ such that

$$U := \{z \in \mathbb{R}^s: d(z, A\psi(\mu)) \leq \varrho\} \subset U_\mu.$$

We define

$$V := \{x \in \mathbb{R}^m: d(x, \psi(\mu)) \leq \varrho \|A\|^{-1}\}, \text{ where}$$

$$d(x, M) := \inf\{\|x - y\|: y \in M\};$$

then $A(V) \subseteq U$.

Proposition 2.1 implies that there exist positive constants L_0 and δ_0 such that $\emptyset \neq \psi(\nu) \subset V$ and $|\varphi(\mu) - \varphi(\nu)| \leq L_0 \mathbf{W}_1(\mu, \nu)$ whenever $\nu \in M_1(\mathbb{R}^s)$, $\mathbf{W}_1(\mu, \nu) < \delta_0$. Let $x_* \in \psi(\mu)$. Then we obtain for each $x \in C \cap V$,

$$\begin{aligned} G(x_*, \mu) &\leq G\left(\frac{1}{2}(x + x_*), \mu\right) \\ &\leq \frac{1}{2}g(x) + \frac{1}{2}g(x_*) + Q_\mu\left(\frac{1}{2}Ax + \frac{1}{2}Ax_*\right) \\ &\leq \frac{1}{2}G(x, \mu) + \frac{1}{2}G(x_*, \mu) - \frac{\kappa}{4}\|Ax - Ax_*\|^2, \end{aligned}$$

and thus

$$G(x, \mu) \geq G(x_*, \mu) + \frac{\kappa}{2}\|Ax - Ax_*\|^2.$$

In the above estimate we used that V is convex and Q_μ is strongly convex on $A(V)$. A first consequence of the last estimate is that $A\psi(\mu)$ is a singleton. Now, put $\delta := \delta_0$ and let $\nu \in M_1(\mathbb{R}^s)$ such that $\mathbf{W}_1(\mu, \nu) < \delta$. Then we have for each $x \in \psi(\nu)$,

$$\begin{aligned} \|Ax - Ax_*\|^2 &\leq \frac{2}{\kappa}(G(x, \mu) - G(x_*, \mu)) \\ &\leq \frac{2}{\kappa}(|\varphi(\nu) - \varphi(\mu)| + |G(x, \mu) - G(x, \nu)|) \\ &\leq \frac{2}{\kappa}(L_0 \mathbf{W}_1(\mu, \nu) + |Q_\mu(Ax) - Q_\nu(Ax)|). \end{aligned}$$

The second term on the right-hand side can be estimated by repeating the argument in the proof of Proposition 2.1. This completes the proof. \square

The following example shows that the exponent $\frac{1}{2}$ on the right-hand side in the assertion of Theorem 2.2 is optimal.

EXAMPLE 2.3. Let

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

and put $A = 1$ (hence $x \in \mathbb{R}$). Consider (1.1) with $g \equiv 0$, $C := \mathbb{R}$ and μ as the uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$. For $n \in \mathbb{N}$ let $\mu_n \in M_1(\mathbb{R})$ be given by its distribution function F_{μ_n} ((ϵ_n) is a sequence in $(0, \frac{1}{2})$ tending to 0):

$$F_{\mu_n}(t) := \begin{cases} 0, & t < -\frac{1}{2}, \\ t + \frac{1}{2}, & t \in [-\frac{1}{2}, -\epsilon_n) \cup [\epsilon_n, \frac{1}{2}), \\ \frac{1}{2}, & t \in [-\epsilon_n, \epsilon_n), \\ 1, & t \geq \frac{1}{2}. \end{cases}$$

Then it holds that $\psi(\mu) = \{0\}$, $\psi(\mu_n) = [-\epsilon_n, \epsilon_n]$ ($n \in \mathbb{N}$) and, hence, $\sup_{x \in \psi(\mu_n)} |x - 0| = \epsilon_n$.

On the other hand, we have (cf. e.g., [22, p. 653])

$$W_1(\mu, \mu_n) = \int_{-\infty}^{\infty} |F_\mu(t) - F_{\mu_n}(t)| dt = \int_{-\epsilon_n}^{\epsilon_n} |t| dt = \epsilon_n^2. \quad \square$$

For the special case that g is linear and C is polyhedral, we even have Hölder stability of optimal sets. This is stated in our next result which extends Theorem 4.4 in [25].

THEOREM 2.4. *Let, in addition to the assumptions of Theorem 2.2, G have the shape*

$$G(x, \nu) := c^T x + Q_\nu(Ax) \quad (c \in \mathbb{R}^m)$$

and C be polyhedral. Then there exist constants $L^* > 0$ and $\delta^* > 0$ such that

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

whenever $\nu \in M_1(\mathbb{R}^s)$, $W_1(\mu, \nu) < \delta^*$. (Here d_H is the Hausdorff distance on subsets of \mathbb{R}^m .)

PROOF. We consider $\bar{A} \in L(\mathbb{R}^m, \mathbb{R}^{s+1})$, $\bar{A}^T = (A^T, c)$, and define $\mathcal{L} := \text{im } \bar{A}^T$. Let P denote the orthogonal projection from \mathbb{R}^m onto \mathcal{L} . Then there holds for each $x \in \mathbb{R}^m$, $\nu \in M_1(\mathbb{R}^s)$, $G(x, \nu) = G(Px, \nu)$. This implies

$$P(\psi(\nu)) = \text{argmin}\{c^T y + Q_\nu(Ay) : y \in P(C)\} =: \psi_{\mathcal{L}}(\nu),$$

and, hence,

$$\Psi(\nu) = \bigcup_{y \in \psi_{\mathcal{L}}(\nu)} C(y),$$

where

$$\begin{aligned} C(y) &:= \{x \in C : x - y \in \mathcal{L}^\perp\} = \{x \in C : x - y \in \ker \bar{A}\} \\ &= \{x \in C : \bar{A}x = \bar{A}y\}. \end{aligned}$$

Therefore

$$d_H(\psi(\mu), \psi(\nu)) \leq \sup_{\substack{y \in \psi_{\mathcal{L}}(\mu), \\ \tilde{y} \in \psi_{\mathcal{L}}(\nu)}} d_H(C(y), C(\tilde{y})).$$

By Hoffman's Theorem [23, p. 760], there exists $\hat{L} > 0$ such that

$$d_H(C(y), C(\tilde{y})) \leq \hat{L} \|\bar{A}y - \bar{A}\tilde{y}\|$$

for all $y, \tilde{y} \in \mathbb{R}^m$ with $C(y) \neq \emptyset$ and $C(\tilde{y}) \neq \emptyset$.

Let $L > 0$ and $\delta > 0$ be chosen such that

$$\sup_{\substack{x \in \psi(\mu), \\ \bar{x} \in \psi(\nu)}} \|Ax - A\bar{x}\| \leq L\mathbf{W}_1(\mu, \nu)^{1/2}$$

whenever $\mathbf{W}_1(\mu, \nu) < \delta$ (according to Theorem 2.2). Let $\nu \in M_1(\mathbb{R}^s)$ with $\mathbf{W}_1(\mu, \nu) < \delta$, $x \in \psi(\mu)$ and $\bar{x} \in \Psi(\nu)$. Then we have

$$\begin{aligned} |c^T x - c^T \bar{x}| &= |\varphi(\mu) - Q_\mu(Ax) - \varphi(\nu) + Q_\nu(A\bar{x})| \\ &\leq |\varphi(\mu) - \varphi(\nu)| + |Q_\nu(A\bar{x}) - Q_\mu(A\bar{x})| + |Q_\mu(A\bar{x}) - Q_\mu(Ax)| \\ &\leq L_h \mathbf{W}_1(\mu, \nu) + L_h \mathbf{W}_1(\mu, \nu) + L_\mu \|A\bar{x} - Ax\|. \end{aligned}$$

Here, the first estimate in the last row follows from Proposition 2.1 and the second estimate is obtained as in the proof of Proposition 2.1. The constant $L_\mu > 0$ is a Lipschitz modulus for $Q_\mu(\cdot)$ with respect to a suitable compact set $\tilde{K} \subset \mathbb{R}^s$ which contains all the points $A\bar{x}$ for $\bar{x} \in \psi(\nu)$, $\nu \in M_1(\mathbb{R}^s)$, $\mathbf{W}_1(\mu, \nu) < \delta$. By strong convexity of $Q_\mu(\cdot)$ and Theorem 2.2, such a set \tilde{K} indeed exists.

Put $\delta^* := \delta$ and let $\nu \in M_1(\mathbb{R}^s)$ such that $\mathbf{W}_1(\mu, \nu) < \delta^*$. The above estimates now yield

$$\begin{aligned} d_H(\psi(\mu), \psi(\nu)) &\leq \hat{L} \sup_{\substack{y \in \psi_\mu(\mu), \\ \bar{y} \in \psi_\mu(\nu)}} \|\bar{A}y - \bar{A}\bar{y}\| = \hat{L} \sup_{\substack{x \in \psi(\mu), \\ \bar{x} \in \psi(\nu)}} \|\bar{A}x - \bar{A}\bar{x}\| \\ &\leq \hat{L} \sup_{\substack{x \in \psi(\mu), \\ \bar{x} \in \psi(\nu)}} \{\|Ax - A\bar{x}\| + |c^T x - c^T \bar{x}|\} \\ &\leq 2\hat{L}L_h \mathbf{W}_1(\mu, \nu) + \hat{L}(1 + L_\mu) \sup_{\substack{x \in \psi(\mu), \\ \bar{x} \in \psi(\nu)}} \|Ax - A\bar{x}\| \\ &\leq L^* \mathbf{W}_1(\mu, \nu)^{1/2} \end{aligned}$$

where $L^* := 2\hat{L}L_h\delta^{1/2} + \hat{L}L(1 + L_\mu)$. \square

The next example demonstrates that the assertion of Theorem 2.4 becomes wrong when the constraint set C is no longer polyhedral.

EXAMPLE 2.5. Modify Example 2.3 by putting $A \in L(\mathbb{R}^2, \mathbb{R}^1)$, $A := (1, 0)$ (hence $x \in \mathbb{R}^2$) and $C := \{x \in \mathbb{R}^2: -x_1 + (x_2)^2 \leq 0\}$ and leaving the remainder unchanged. Then $\psi(\mu) = \{0\}$, and the points $(\epsilon_n, \epsilon_n^{1/2})$ belong to $\psi(\mu_n)$ for each $n \in \mathbb{N}$. Therefore, $d_H(\psi(\mu), \psi(\mu_n)) \geq \epsilon_n^{1/2}$; on the other hand, $\mathbf{W}_1(\mu, \mu_n) = \epsilon_n^2$.

REMARK 2.6. In [25] quantitative stability of recourse models was investigated with respect to the bounded Lipschitz metric β (see [7] for the definition of β), which metrizes the topology of weak convergence. Between β and the Wasserstein metric \mathbf{W}_1 , we have the following interrelation: There exists a constant $C > 0$ such that

$$\beta(\mu, \nu) \leq \mathbf{W}_1(\mu, \nu) \leq C(1 + m_p(\mu) + m_p(\nu))\beta(\mu, \nu)^{1-1/p}$$

for all $\mu, \nu \in M_p(\mathbb{R}^s)$ and each $p > 1$ ($m_p(\mu)$ denotes $(\int_{\mathbb{R}^s} \|z\|^p \mu(dz))^{1/p}$). Here, the first inequality is a consequence of the Kantorovich-Rubinstein Theorem and the

second one can be proved using Theorem 2.1 in [27]. The L_2 -Wasserstein metric is of special interest since for $W_2(\mu, \nu)$ there exist explicit formulae when μ and ν belong to certain classes of probability measures (cf. e.g., [12], [13], [22]). In [12], an explicit formula for $W_2(\mu, \nu)$ containing only expectations and covariance matrices of μ and ν has been established for a family of elliptically contoured distributions. Especially, the formula holds when μ and ν are Gaussian measures on \mathbb{R}^s . In view of $W_1(\mu, \nu) \leq W_2(\mu, \nu)$, therefore, from Theorems 2.2 and 2.4 stability properties for optimal solutions to recourse problems with finite-dimensional distribution parameters result when μ and ν belong to the above-mentioned classes of measures.

When applying Theorems 2.2 and 2.4 to specific instances of (1.1), it is decisive to resort to general and verifiable sufficient conditions for the strong convexity of the expected recourse function Q_μ . For the case that Q_μ has locally Lipschitzian gradient, such conditions are developed in the following section (Theorems 3.1 and 3.4). Roughly speaking they consist of an interplay between algebraical and analytical properties of W, q (cf. (1.3)) and the probability distribution μ , respectively.

3. Strong convexity and $C^{1,1}$ properties. When studying strong convexity of Q_μ we may neglect the functional dependence of Q_μ on the measure μ , and only properties of Q_μ as a function of Ax (cf. (1.2)) are interesting. Hence, we simplify our notation by writing $Q(x)$ instead of $Q_\mu(Ax)$, with the consequence that now x is a variable in \mathbb{R}^s (instead of \mathbb{R}^m). Thus, we are here interested in the strong convexity of the functional Q which is given by

$$Q(x) = \int_{\mathbb{R}^s} \tilde{Q}(x, z) \mu(dz) \quad \text{and}$$

$$\tilde{Q}(x, z) = \min\{q^T y : Wy = z - x, y \geq 0\}.$$

In addition to our general assumptions (A1)–(A3) let us impose:

(A4) The probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$ is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^s .

Then, Q is continuously differentiable on \mathbb{R}^s , and the gradient $\nabla Q(x)$ has the form (cf. [14], [31]):

$$(3.1) \quad \nabla Q(x) = \sum_{i=1}^l d_i f_i(x), \quad x \in \mathbb{R}^s,$$

where for $i = 1, \dots, l$,

$$d_i := -(B_i^{-1})^T q_{B_i},$$

$$f_i(x) := \mu(\{z \in \mathbb{R}^s : B_i^{-1}z \geq B_i^{-1}x\}),$$

$B_i \in L(\mathbb{R}^s, \mathbb{R}^s)$ are nonsingular matrices consisting of columns of W which occur as components of optimal basis solutions in the course of the computation of $\tilde{Q}(x, z)$. An essential point for further considerations is the property

$$\bigcup_{i=1}^l \{t \in \mathbb{R}^s : B_i^{-1}t \geq 0\} = \mathbb{R}^s,$$

and $q_{B_i} \in \mathbb{R}^s$ are the vectors formed by those components of q that correspond to the columns of W which form B_i .

Without loss of generality we assume that the basis matrices B_i ($i = 1, \dots, l$) are selected in such a way that the above covering of \mathbb{R}^s is minimal, i.e., the sets $\{t \in \mathbb{R}^s: B_i^{-1}t \geq 0, B_j^{-1}t \geq 0\}$ are for all pairs (i, j) with $i, j \in \{1, \dots, l\}, i \neq j$ subsets of hyperplanes in \mathbb{R}^s . By (A4) we then have

$$\mu(\{z \in \mathbb{R}^s: B_i^{-1}z \geq B_i^{-1}x\} \cap \{z \in \mathbb{R}^s: B_j^{-1}z \geq B_j^{-1}x\}) = 0$$

for all $i \neq j$ and all $x \in \mathbb{R}^s$. Hence

$$(3.2) \quad \sum_{i=1}^l f_i(x) = 1 \quad \text{for all } x \in \mathbb{R}^s.$$

Obviously, $f_i(x)$ ($i = 1, \dots, l, x \in \mathbb{R}^s$) may be written as

$$f_i(x) = \mu(x + \mathcal{K}_i),$$

where $\mathcal{K}_i \subset \mathbb{R}^s$ ($i = 1, \dots, l$) is a simplicial cone (i.e., the conical hull of s linearly independent vectors in \mathbb{R}^s) and $x + \mathcal{K}_i$ is understood as the Minkowski sum.

We assume

(A5) For any simplicial cone $\mathcal{K} \subset \mathbb{R}^s$, the function $x \mapsto \mu(x + \mathcal{K})$ is locally Lipschitzian on \mathbb{R}^s .

Then, (A1)–(A5) imply that ∇Q is locally Lipschitzian on \mathbb{R}^s , in other words, Q is a $C^{1,1}$ functional. We will discuss (A5) at the end of the present section.

To formulate our general sufficient condition for strong convexity of Q we need a specific notion of directional differentiability for locally Lipschitzian mappings which goes back to Kummer [19]. Given $f: \mathbb{R}^s \rightarrow \mathbb{R}^s$ locally Lipschitzian on \mathbb{R}^s and $x, y \in \mathbb{R}^s$, the generalized directional derivative $\Delta f(x; u)$ is defined as the set consisting of all points $z \in \mathbb{R}^s$ which are a limit of points

$$z^k = (\lambda_k)^{-1}(f(x^k + \lambda_k u) - f(x^k)),$$

where $x^k \rightarrow x$ and $\lambda_k \downarrow 0$ ($k \rightarrow \infty$). (Definition 1.4 in [19]).

To the generalized Jacobian $\partial f(x)$ in the sense of Clarke [6] we have the relation ((P4) in [19]):

$$\Delta f(x; u) \subset \{z \in \mathbb{R}^s: z = Mu, M \in \partial f(x)\}.$$

This inclusion can be strict (Example 2.2 in [19]).

THEOREM 3.1. *Suppose (A1)–(A5), let $V \subset \mathbb{R}^s$ convex, compact and*

$$\emptyset = \ker D \cap (\text{lin}\{\mathbf{1}\})^\perp \cap \bigtimes_{i=1}^l \Delta f_i(x; u)$$

for all $x \in V, u \in \mathbb{R}^s, \|u\| = 1$. Then Q is strongly convex on V .

Here, $\ker D$ denotes the kernel of the linear mapping given by the matrix $D \in L(\mathbb{R}^l, \mathbb{R}^s)$ whose columns are the vectors d_1, \dots, d_l (cf. (3.1)), $(\text{lin}\{\mathbf{1}\})^\perp$ is the orthogonal complement of the linear subspace spanned by the vector $\mathbf{1} \in \mathbb{R}^l$ whose components are identically 1 and $\bigtimes_{i=1}^l \Delta f_i(x; u)$ is the cartesian product.

PROOF. Consider $\Delta(\nabla Q)(x; u) \subset \mathbb{R}^s$, the generalized directional derivative in the sense of Kummer for the mapping ∇Q , and assume that we have already established

$$(3.3) \quad \inf\{\langle w, u \rangle : w \in \Delta(\nabla Q)(x; u)\} > 0$$

for all $x \in V$, $u \in \mathbb{R}^s$, $\|u\| = 1$, ($\langle w, u \rangle := w^T u$). Then it is possible to conclude strong convexity of Q on the set V by showing that there exists $\kappa > 0$ such that

$$Q(x) - Q(\tilde{x}) - \langle \nabla Q(\tilde{x}), x - \tilde{x} \rangle \geq \kappa \|x - \tilde{x}\|^2$$

for all $x, \tilde{x} \in V$.

Indeed, since the mapping $\Delta(\nabla Q)(\cdot, \cdot)$ is locally bounded and closed ((P2) in [19]), relation (3.3) implies that for any $x^0 \in V$ there exist $\varrho > 0$ and $c > 0$ such that

$$\inf\{\langle w, u \rangle : w \in \Delta(\nabla Q)(\tilde{x}; u)\} \geq c$$

for all $\tilde{x} \in V$, $\|\tilde{x} - x^0\| \leq \varrho$, $u \in \mathbb{R}^s$, $\|u\| = 1$. Since V is compact, this yields that there is some $c_0 > 0$ such that

$$\inf\{\langle w, u \rangle : w \in \Delta(\nabla Q)(x; u)\} \geq c_0$$

for all $x \in V$, $u \in \mathbb{R}^s$, $\|u\| = 1$. Using the Taylor formula in [19, Theorem 3.2], we obtain that for any $x, \tilde{x} \in V$ there exist

$$\gamma \in (0, 1) \quad \text{and} \quad \tilde{w} \in \Delta(\nabla Q)(x + \gamma(\tilde{x} - x); \tilde{x} - x)$$

such that

$$Q(\tilde{x}) - Q(x) - \langle \nabla Q(x), \tilde{x} - x \rangle = \frac{1}{2} \langle \tilde{w}, \tilde{x} - x \rangle$$

and, hence,

$$\begin{aligned} & Q(\tilde{x}) - Q(x) - \langle \nabla Q(x), \tilde{x} - x \rangle \\ & \geq \frac{1}{2} \|\tilde{x} - x\|^2 \inf \left\{ \left\langle w, \frac{\tilde{x} - x}{\|\tilde{x} - x\|} \right\rangle : w \in \Delta(\nabla Q) \left(x + \gamma(\tilde{x} - x); \frac{\tilde{x} - x}{\|\tilde{x} - x\|} \right) \right\} \\ & \geq \frac{1}{2} c_0 \|\tilde{x} - x\|^2 \end{aligned}$$

which yields the desired strong convexity.

It remains to verify (3.3). Let $x \in \mathbb{R}^s$, $u \in \mathbb{R}^s$, $w \in \Delta(\nabla Q)(x; u)$. Then

$$\langle w, u \rangle = \lim_{k \rightarrow \infty} \lambda_k^{-2} \langle \nabla Q(x^k + \lambda_k u) - \nabla Q(x^k), \lambda_k u \rangle$$

for some sequences $x^k \rightarrow x$ and $\lambda_k \downarrow 0$. By monotonicity of ∇Q the above limit is nonnegative and, therefore, the same is true for the infimum in (3.3). Assume that, for some $x \in V$, $u \in \mathbb{R}^s$, $\|u\| = 1$, the infimum in (3.3) were zero. Denote by $E_{\nabla Q} \subset \mathbb{R}^s$ the set of those points where ∇Q is (Frechet-) differentiable and consider

$$\partial_*(\nabla Q)(x) = \{M \in L(\mathbb{R}^s, \mathbb{R}^s) : \exists x^k \rightarrow x, x^k \in E_{\nabla Q}, \nabla^2 Q(x^k) \rightarrow M\}.$$

By (P4) and (P5) in [19] we have

$$(3.4) \quad \partial_*(\nabla Q)(x)u \subset \Delta(\nabla Q)(x; u) \subset \partial^2 Q(x)u$$

where $\partial^2 Q(x)$ is the convex hull of $\partial_*(\nabla Q)(x)$. Next, one confirms that

$$(3.5) \quad \inf\{\langle Mu, u \rangle : M \in \partial_*(\nabla Q)(x)\} = 0.$$

Indeed, the first inclusion in (3.4) and our assumption that the infimum in (3.3) were zero imply that the infimum in (3.5) is nonnegative. If the latter infimum were positive, then, by $\partial^2 Q(x) = \text{conv } \partial_*(\nabla Q)(x)$, the same is true for

$$\inf\{\langle Mu, u \rangle : M \in \partial^2 Q(x)\}.$$

Together with the second inclusion in (3.4) this would imply that the infimum in (3.3) is positive, contradicting our assumption that this infimum were zero. Hence, (3.5) is verified and, by compactness of $\partial_*(\nabla Q)(x)$ (recall that ∇Q is locally Lipschitzian) there exists $M \in \partial_*(\nabla Q)(x)$ such that $\langle Mu, u \rangle = 0$. By a standard argument (representing u by an orthonormalized basis of eigenvectors of M in \mathbb{R}^s), this yields $Mu = 0$.

Now, there exist $x^k \in E_{\nabla Q}$, $x^k \rightarrow x$ with $\nabla^2 Q(x^k)u \rightarrow Mu = 0$. For any $k \in \mathbb{N} \setminus \{0\}$, we have some $\lambda_k \in (0, k^{-1}]$ such that

$$\|\lambda_k^{-1}(\nabla Q(x^k + \lambda_k u) - \nabla Q(x^k)) - \nabla^2 Q(x^k)u\| \leq k^{-1}.$$

In view of $\nabla^2 Q(x^k)u \rightarrow 0$, it follows that

$$\lambda_k^{-1}(\nabla Q(x^k + \lambda_k u) - \nabla Q(x^k)) \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Formula (3.1) yields

$$(3.6) \quad \sum_{i=1}^l d_i \lambda_k^{-1} (f_i(x^k + \lambda_k u) - f_i(x^k)) \rightarrow 0.$$

For k sufficiently large, now $x^k + \lambda_k u$ and x^k belong to the neighbourhood of x where all the f_i ($i = 1, \dots, l$) are Lipschitz (w.l.o.g. with joint module $L > 0$). Hence

$$|\lambda_k^{-1}(f_i(x^k + \lambda_k u) - f_i(x^k))| \leq L \quad \text{for } i = 1, \dots, l.$$

Successively, we pick subsequences of (x^k) and (λ_k) , again denoted by $(x^k), (\lambda_k)$, such that

$$\lambda_k^{-1}(f_i(x^k + \lambda_k u) - f_i(x^k)) \rightarrow f_i^* \quad (i = 1, \dots, l; k \rightarrow \infty)$$

for certain $f_1^*, \dots, f_l^* \in \mathbb{R}$.

In view of the properties of $(x^k), (\lambda_k)$, we have $f_i^* \in \Delta f_i(x; u)$ for $i = 1, \dots, l$. According to (3.6), it holds that $\sum_{i=1}^l d_i f_i^* = 0$, hence $f^* := (f_1^*, \dots, f_l^*)^T \in \ker D$. Furthermore, (3.2) yields

$$\sum_{i=1}^l f_i(x^k + \lambda_k u) - \sum_{i=1}^l f_i(x^k) = 1 - 1 = 0 \quad \text{for all } k,$$

and, therefore, $\sum_{i=1}^l f_i^* = 0$, implying $f^* \in (\text{lin}\{\mathbf{1}\})^\perp$. Hence,

$$f^* \in \ker D \cap (\text{lin}\{\mathbf{1}\})^\perp \cap \bigtimes_{i=1}^l \Delta f_i(x; u)$$

in contradiction to the assumption in our theorem. Hence, the infimum in (3.3) is not zero, and (3.3) is verified. \square

Before applying Theorem 3.1 let us consider some “exceptional” situation where the functional Q is linear (hence, necessarily not strongly convex). Assume (A1)–(A4) and split the second-stage cost vector q (see the definition of \tilde{Q} above) into a direct sum

$$(3.7) \quad q = q^1 + q^2$$

where $q^1 \in \ker W$ and $q^2 \in (\ker W)^\perp = \text{im } W^T$. Then, there exists $u \in \mathbb{R}^s$ such that $W^T u = q^2$, implying $B_i^T u = q_{B_i}^2$ for $i = 1, \dots, l$, and, therefore, $u = (B_i^{-1})^T q_{B_i}^2$ for $i = 1, \dots, l$.

According to formula (3.1), we have

$$\begin{aligned} \nabla Q(x) &= \sum_{i=1}^l - (B_i^{-1})^T q_{B_i}^1 f_i(x) + \sum_{i=1}^l - (B_i^{-1})^T q_{B_i}^2 f_i(x) \\ &= - \sum_{i=1}^l (B_i^{-1})^T q_{B_i}^1 f_i(x) - u \sum_{i=1}^l f_i(x) \\ &= - \sum_{i=1}^l (B_i^{-1})^T q_{B_i}^1 f_i(x) - u \quad (\text{in view of (3.2)}). \end{aligned}$$

If $q \in \text{im } W^T$, then $q^1 = 0$ and $\nabla Q(x) = -u$ for all $x \in \mathbb{R}^s$, hence Q is a linear function independently on the choice of μ . To exclude this case from further considerations we always will assume that $q \notin \text{im } W^T$. Then in (3.7) we have $q^1 \neq 0$ and, by the above representation of ∇Q , we may neglect the portion q^2 in our considerations on strong convexity of Q .

LEMMA 3.2. *Let $\mu \in \mathcal{P}(\mathbb{R}^s)$ fulfill (A3)–(A5) and consider some function $f: \mathbb{R}^s \rightarrow \mathbb{R}$ given by $f(x) := \mu(x + \mathcal{K})$, where $\mathcal{K} \subset \mathbb{R}^s$ is a simplicial cone. Suppose that for some density Θ of μ and for some $x^0 \in \mathbb{R}^s$ there exist $R > 0$ and $r > 0$ such that*

$$\Theta(x) \geq r \quad \text{for all } x \in \mathbb{R}^s, \|x - x^0\| \leq R.$$

Then there exists a constant $c > 0$ such that for all $u \in \mathcal{K}$

$$\sup\{w : w \in \Delta f(x^0; u)\} \leq -c\|u\|.$$

PROOF. Let $B \in L(\mathbb{R}^s, \mathbb{R}^s)$ nonsingular, such that $\mathcal{K} = \{t \in \mathbb{R}^s : Bt \geq 0\}$. Consider $u \in \mathcal{K} \setminus \{0\}$ (for $u = 0$ the assertion trivially holds). Then we have

$$(3.8) \quad Bu \geq 0 \quad \text{and} \quad Bu \neq 0.$$

Without loss of generality, let

$$(3.9) \quad [Bu]_1 = \|Bu\|_\infty := \max_{i=1, \dots, s} |[Bu]_i|$$

where $[Bu]_i$ denotes the i th component of Bu . Let $w \in \Delta f(x^0; u)$. Then there exist sequences $(x^k), (\lambda_k)$ with $x^k \rightarrow x^0, \lambda_k \downarrow 0$ and

$$w = \lim_{k \rightarrow \infty} \lambda_k^{-1} (f(x^k + \lambda_k u) - f(x^k)).$$

Take some $\varrho > 0$ such that

$$(3.10) \quad B^{-1}(\{y \in \mathbb{R}^s : \|Bx^0 - y\|_\infty < 2\varrho\}) \subset \{x \in \mathbb{R}^s : \|x - x^0\| \leq R\}$$

and then fix some $k_0 \in \mathbb{N}$ sufficiently large such that

$$(3.11) \quad \|Bx^0 - Bx^k\|_\infty \leq \frac{1}{2}\varrho \quad \text{and} \quad \lambda_k \|Bu\|_\infty \leq \frac{1}{2}\varrho \quad \text{for } k \geq k_0.$$

With the transformation $\tau := Bt$ we have for all $x \in \mathbb{R}^s$

$$\begin{aligned} f(x) &= \int_{\{t \in \mathbb{R}^s : Bt \geq Bx\}} \Theta(t) dt \\ &= \int_{\{\tau \in \mathbb{R}^s : \tau \geq Bx\}} \Theta(B^{-1}\tau) |\det B^{-1}| d\tau. \end{aligned}$$

Hence

$$\begin{aligned} f(x^k + \lambda_k u) - f(x^k) &= \int_{\{\tau \in \mathbb{R}^s : \tau \geq Bx^k + \lambda_k Bu\}} \Theta(B^{-1}\tau) |\det B^{-1}| d\tau \\ &\quad - \int_{\{\tau \in \mathbb{R}^s : \tau \geq Bx^k\}} \Theta(B^{-1}\tau) |\det B^{-1}| d\tau \\ &\leq - \int_{[Bx^k]_1}^{[Bx^k + \lambda_k Bu]_1} \int_{[Bx^k]_2}^\infty \dots \int_{[Bx^k]_s}^\infty \Theta(B^{-1}\tau) |\det B^{-1}| d\tau \\ &\quad \text{(this estimate holds in view of (3.8) and (3.9))} \\ &\leq - \int_{[Bx^k]_1}^{[Bx^k]_1 + \lambda_k \|Bu\|_\infty} \int_{[Bx^k]_2}^{[Bx^k]_2 + \varrho/2} \dots \\ &\quad \dots \int_{[Bx^k]_s}^{[Bx^k]_s + \varrho/2} \Theta(B^{-1}\tau) |\det B^{-1}| d\tau \\ &\leq -\lambda_k \|Bu\|_\infty \left(\frac{1}{2}\varrho\right)^{s-1} r |\det B^{-1}| \end{aligned}$$

(this estimate holds in view of (3.10) and (3.11)). Due to the equivalence of norms in \mathbb{R}^s , there exists $c_1 > 0$ such that

$$c_1 \|u\| \leq \|Bu\|_\infty \quad \text{for all } u \in \mathbb{R}^s,$$

and we finally obtain

$$\lambda_k^{-1}(f(x^k + \lambda_k u) - f(x^k)) \leq -\|u\|c_1(\frac{1}{2}\varrho)^{s-1}r|\det B^{-1}|,$$

i.e., $w \leq -c\|u\|$ with $c := c_1(\frac{1}{2}\varrho)^{s-1}r|\det B^{-1}|$. \square

LEMMA 3.3. *Let $W \in L(\mathbb{R}^{s+1}, \mathbb{R}^s)$ fulfill (A1) and $q \in \ker W \setminus \{0\}$. Then $\ker D$ is spanned by the vector whose components are the squares of the components of q .* \square

The proof of the above lemma can be bound in [26] (Lemma 3.4). It consists of a rather long algebraic argumentation using specific properties of W as a complete-recourse matrix with dimension $s \times (s + 1)$ (cf. Lemma 13 and Theorem 14 in [14, p. 52]).

Although the general sufficient condition in Theorem 3.1 can, formally, be checked for any given recourse model, one, nevertheless, is interested in direct relations between problem data and the desired strong convexity. Such relations are established in the next theorem, which also covers simple recourse, where, due to the inherent separability, of course, a direct approach is possible too.

THEOREM 3.4. *Let either $W \in L(\mathbb{R}^{s+1}, \mathbb{R}^s)$, $q \notin \text{im } W^T$ and (A1)–(A5) or $W \in L(\mathbb{R}^{2s}, \mathbb{R}^s)$, $W = (H, -H)$ with some nonsingular matrix $H \in L(\mathbb{R}^s, \mathbb{R}^s)$, $q = (q^+, q^-)^T$ with $q^+, q^- \in \mathbb{R}^s$, $q^+ + q^- > 0$ (componentwise) and (A3)–(A5). Let further $V \subset \mathbb{R}^s$ convex, compact and suppose there exist an open set $U \supset V$ and a constant $r > 0$ such that for some density Θ of μ*

$$\Theta(t) \geq r \text{ for all } t \in U.$$

Then Q is strongly convex on V .

PROOF. Let us denote

$$\mathcal{A}(x, u) := \ker D \cap (\text{lin}\{\mathbf{1}\})^\perp \cap \bigtimes_{i=1}^l \Delta f_i(x; u).$$

First, we show that

$$(3.12) \quad 0 \notin \mathcal{A}(x, u) \text{ for all } x \in V, u \in \mathbb{R}^s, \|u\| = 1.$$

Indeed, if $0 \in \mathcal{A}(x, u)$ for some $x \in V$, $u \in \mathbb{R}^s$, $\|u\| = 1$, then there exists $i \in \{1, \dots, l\}$ such that $u \in \mathcal{K}_i$ (cf. (3.1)). Our positivity condition for the density and Lemma 3.2 now provide that $0 \notin \Delta f_i(x; u)$, contradicting $0 \in \mathcal{A}(x, u)$. We continue with the case $W \in L(\mathbb{R}^{s+1}, \mathbb{R}^s)$ and split q into a direct sum as in (3.7). Since $q \notin \text{im } W^T$, we have $q^1 \neq 0$ and, as stated above, the portion q^2 can be neglected. Lemma 3.3 (applied to q^1) yields

$$\ker D \cap (\text{lin}\{\mathbf{1}\})^\perp = \{0\}.$$

Hence, $\mathcal{A}(x, u) \subseteq \{0\}$ for all $x \in V$, $u \in \mathbb{R}^s$, $\|u\| = 1$. By (3.12), this yields $\mathcal{A}(x, u) = \emptyset$ for all $x \in V$, $u \in \mathbb{R}^s$, $\|u\| = 1$, and the assumption of Theorem 3.1 is verified.

For the second part of our assertion we first reduce the proof to “simple recourse”, i.e., to the case $H = I$ with $I \in L(\mathbb{R}^s, \mathbb{R}^s)$ denoting the identity matrix. Denote $W_j := (I, -I)$, then $W = HW_j$ and for any basis matrix B_i arising in (3.1) we obtain the representation $B_i = HB_{i,j}$, where $B_{i,j}$ is the basis matrix formed by the corresponding columns of W_j . This yields $d_i = (H^{-1})^T d_{i,j}$ where $d_{i,j}$ again corresponds to

W_I , and for the matrix D_I with columns $d_{i,I}$ ($i = 1, \dots, l$) we obtain $\ker D_I = \ker D$. For the functions f_i in (3.1) we obtain

$$f_i(x) = \mu \circ H(H^{-1}x + \mathcal{K}_{i,I}) \quad \text{where } \mathcal{K}_{i,I} = H^{-1}(\mathcal{K}_i).$$

The transformation formula for densities together with our positivity condition for the density of μ yield that there is a density of $\mu \circ H$ satisfying the positivity condition on a neighbourhood of $H^{-1}(V)$. Hence, we may confine ourselves to simple recourse. In a first step, we will show that for each $x \in V$

$$(3.13) \quad \begin{aligned} \emptyset &= \mathcal{A}(x, u_j) \quad \text{for all } u_j \text{ in the canonical basis of } \mathbb{R}^s, \text{ implies} \\ \emptyset &= \mathcal{A}(x, u) \quad \text{for all } u \in \mathbb{R}^s, \|u\| = 1. \end{aligned}$$

Assume that there exists $v \in \mathbb{R}^l$ such that

$$v \in \mathcal{A}(x, u) \quad \text{for some } x \in V, u \in \mathbb{R}^s, \|u\| = 1.$$

Then $u = \sum_{j=1}^s \lambda_j u_j$ with suitable $\lambda_j \in \mathbb{R}$, and, by property (P1) in [19], $v \in \bigcup_{j=1}^s \lambda_j \mathcal{A}(x, u_j)$. Therefore $v \in \lambda_{j^*} \mathcal{A}(x, u_{j^*})$ for some $j^* \in \{1, \dots, s\}$. In case $\lambda_{j^*} = 0$ this would provide $v = 0$, and $0 \in \mathcal{A}(x, u)$ in contradiction to (3.12). If $\lambda_{j^*} \neq 0$, then $(\lambda_{j^*})^{-1}v \in \mathcal{A}(x, u_{j^*})$ in contradiction to (3.13).

In the remainder, let u_j be the canonical basis vector whose j th component equals 1. When investigating $\ker D$ we may confine ourselves to $q \in \ker W \setminus \{0\}$, i.e., we may assume $q^+ = q^-$, and the assumption $q^+ + q^- > 0$ turns into $q^+ > 0$. For simple recourse, the basis matrices B_i in (3.1) are diagonal with nonzero entries in $\{-1, 1\}$. Furthermore $l = 2^s$, and the cones \mathcal{K}_i in (3.1) are the orthants in \mathbb{R}^s . Now there exists a subset $I \subset \{1, \dots, 2^s\}$ with cardinality 2^{s-1} such that $u_j \in \mathcal{K}_i$ for all $i \in I$.

Consider $[d_i]_j$, the j th components of the vectors d_i ($i = 1, \dots, l$) which form the matrix D . We have

$$[d_i]_j = -[q^+]_j \quad \text{for } i \in I \quad \text{and} \quad [d_i]_j = [q^+]_j \quad \text{for } i \notin I.$$

If there were $v \in \ker D \cap (\text{lin}\{1\})^\perp$ we would obtain in light of $[q^+]_j > 0$,

$$(3.14) \quad \sum_{i \in I} [v]_i = \sum_{i \notin I} [v]_i = 0.$$

Since $u_j \in \mathcal{K}_i$ for all $i \in I$, we have by our positivity condition for the density and Lemma 3.2 that there exists $c > 0$ such that

$$\sup\{w : w \in \Delta f_i(x; u_j)\} \leq -c$$

for all $x \in V$ and all $i \in I$. Together with (3.14) this yields $\mathcal{A}(x, u_j) = \emptyset$ for all $x \in V$. □

The above theorem raises the question whether already the positivity condition for the density together with $q \notin \text{im } W^T$ will yield us strong convexity of Q for an arbitrary complete-recourse matrix W . The following example shows that the answer is negative.

EXAMPLE 3.5. Let

$$W = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 1 & 1 & 1 & -1 \end{pmatrix} \quad \text{and} \quad q = (1, -1, 1, 1)^T.$$

Then (A1) is fulfilled and (A2) holds with $u = (0, -1)^T$. We choose an arbitrary $\mu \in \mathcal{P}(\mathbb{R}^2)$ fulfilling (A3)–(A5). Since $q \in \ker W \setminus \{0\}$, we have $q \notin \text{im } W^T$. One calculates

$$D = \begin{pmatrix} -2 & -2 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

which yields together with (3.1) and (3.2)

$$\nabla Q(x) = \begin{pmatrix} \sum_{i=1}^4 [d_i]_1 \cdot f_i(x) \\ 1 \end{pmatrix}.$$

Hence ∇Q is not strongly monotone, and, therefore, Q is not strongly convex on any convex set $V \subset \mathbb{R}^s$ containing elements x, \tilde{x} such that $[x]_2 \neq [\tilde{x}]_2$.

Using Theorems 3.1 and 3.4 it is in principle possible to check in Theorem 2.2 whether Q_μ satisfies the hypothesis to be strongly convex on an open neighbourhood V_μ of (the bounded set) $A(\psi(\mu))$. Of course, in general $A(\psi(\mu))$ is hardly available and one cannot benefit from restricting considerations to V_μ . Nevertheless, there are relevant cases where the verification is possible. For instance, if in $P(\mu)$ the constraint set C is compact and the density of μ is positively bounded below on an open set containing $A(C)$ or if the density of μ is positively bounded below on any compact subset of \mathbb{R}^s (as, e.g., for the nondegenerated normal distribution).

REMARK 3.6. Our analysis on strong convexity of Q_μ readily extends to probabilistic measures $\mu \in \mathcal{P}(\mathbb{R}^s)$ with representation

$$\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$$

where $\alpha \in [0, 1)$, $\mu_1 \in \mathcal{P}(\mathbb{R}^s)$ is arbitrary and $\mu_2 \in \mathcal{P}(\mathbb{R}^s)$ fulfills (A3)–(A5) and the hypotheses in Theorem 3.1 and 3.4, respectively. Indeed, we have

$$Q_\mu = \alpha Q_{\mu_1} + (1 - \alpha)Q_{\mu_2}$$

which is strongly convex as the sum of the convex function αQ_{μ_1} and the strongly convex function $(1 - \alpha)Q_{\mu_2}$.

Let us now discuss assumption (A5). We have

PROPOSITION 3.7. (A5) is fulfilled if and only if, for any nonsingular matrix $B \in L(\mathbb{R}^s, \mathbb{R}^s)$, the distribution function $F_{\mu \circ B}$ of $\mu \circ B$ is locally Lipschitzian.

PROOF. Let $\mathcal{K} \subset \mathbb{R}^s$ be a simplicial cone. Then there exists a nonsingular matrix $B \in L(\mathbb{R}^s, \mathbb{R}^s)$ such that

$$\mathcal{K} = \{t \in \mathbb{R}^s : Bt \geq 0\}.$$

Hence, for all $x \in \mathbb{R}^s$,

$$\begin{aligned} \mu(x + \mathcal{K}) &= \mu(\{t \in \mathbb{R}^s : -Bt \leq -Bx\}) \\ &= (\mu \circ (-B^{-1}))(\{t \in \mathbb{R}^s : t \leq -Bx\}) \\ &= F_{\mu \circ (-B^{-1})}(-Bx). \end{aligned}$$

This verifies the asserted equivalence. \square

PROPOSITION 3.8. *Let $\mu \in \mathcal{P}(\mathbb{R}^s)$ such that for all one-dimensional marginal distributions of μ there exist densities which are locally bounded on \mathbb{R} . Then the distribution function F_μ of μ is locally Lipschitzian on \mathbb{R}^s .*

PROOF. The assumption implies that all the one-dimensional marginal distribution functions F_{μ_i} ($i = 1, \dots, s$) are locally Lipschitzian on \mathbb{R} . To have a simpler notation we continue the proof for $s = 2$. For higher dimensions, an analogous argument applies. Let $(x_1, x_2), (\tilde{x}_1, \tilde{x}_2) \in \mathbb{R}^2$,

$$\begin{aligned} &|F_\mu(x_1, x_2) - F_\mu(\tilde{x}_1, \tilde{x}_2)| \\ &\leq |F_\mu(x_1, x_2) - F_\mu(\tilde{x}_1, x_2)| + |F_\mu(\tilde{x}_1, x_2) - F_\mu(\tilde{x}_1, \tilde{x}_2)| \\ &\leq \mu(\{(\xi_1, \xi_2) : \xi_1 \in (\min\{x_1, \tilde{x}_1\}, \max\{x_1, \tilde{x}_1\}], \xi_2 \leq x_2\}) \\ &\quad + \mu(\{(\xi_1, \xi_2) : \xi_1 \leq \tilde{x}_1, \xi_2 \in (\min\{x_2, \tilde{x}_2\}, \max\{x_2, \tilde{x}_2\})\}) \\ &\leq \mu([\min\{x_1, \tilde{x}_1\}, \max\{x_1, \tilde{x}_1\}] \times \mathbb{R}) \\ &\quad + \mu(\mathbb{R} \times [\min\{x_2, \tilde{x}_2\}, \max\{x_2, \tilde{x}_2\}]) \\ &\leq |F_{\mu_1}(x_1) - F_{\mu_1}(\tilde{x}_1)| + |F_{\mu_2}(x_2) - F_{\mu_2}(\tilde{x}_2)|. \end{aligned}$$

Hence, F_μ is locally Lipschitzian. \square

Combining the “if” part of Proposition 3.7 and Proposition 3.8 one obtains the sufficient condition of Wang for Q_μ to have locally Lipschitzian gradient ([30, Theorem 2.8]). It is clear that the local Lipschitz property in (A5) is needed only for those simplicial cones which are determined by the basis matrices arising in representation (3.1). Sometimes this observation can lead to simplifications. For instance, in the case of simple recourse it can be shown that ∇Q is locally Lipschitzian if $\mu \in \mathcal{P}(\mathbb{R}^s)$ has a locally Lipschitzian distribution function.

We mention that in general (local) boundedness of the marginal densities is not implied by (local) boundedness of the density of the distribution itself (cf., e.g., Example 2.5 in [26]). Furthermore, if a distribution function is locally Lipschitzian then this property is in general not preserved under (nonsingular) linear transformations, as can be seen by the following example, which we present without the rather long but straightforward calculation in the background.

EXAMPLE 3.9. Consider $\mu \in \mathcal{P}(\mathbb{R}^2)$ with the density

$$\Theta(t_1, t_2) = \begin{cases} -2t_2 \log t_1 & \text{if } t_1 \in (0, 1], t_2 \in [0, 1], \\ 0 & \text{else,} \end{cases}$$

and the matrix

$$B = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Then the one-dimensional marginal distributions of $\mu \circ B$ have bounded densities.

On the other hand, we have for $(\mu \circ B) \circ B^{-1} = \mu$

$$\Theta_1(t_1) = \begin{cases} -\log t_1 & \text{if } t_1 \in (0, 1], \\ 0 & \text{else,} \end{cases}$$

which is not bounded on neighbourhoods of $t_1 = 0$.

Finally, we present distributions which satisfy (A5). Following [4], [20], [21] we say that $\mu \in \mathcal{P}(\mathbb{R}^s)$ belongs to the class \mathcal{M}_r ($r \in [-\infty, 0]$) if for all $\lambda \in [0, 1]$ and all Borel subsets C_1, C_2 of \mathbb{R}^s such that $\lambda C_1 + (1 - \lambda)C_2$ is Borel,

$$\mu(\lambda C_1 + (1 - \lambda)C_2) \geq M_r(\mu(C_1), \mu(C_2); \lambda).$$

Here, we denote by $M_r(a, b; \lambda)$ the r th mean of the nonnegative numbers a, b with weights $\lambda, 1 - \lambda$, defined as

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

Measures belonging to $\mathcal{M}_r(\mathcal{M}_{-\infty}, \mathcal{M}_0)$ are called ‘‘convex of order r ’’ (‘‘quasi-concave’’, ‘‘logarithmic concave’’) (see [21]). It is known that μ belongs to \mathcal{M}_r ($r \in [-\infty, 0]$) if μ has a density Θ_μ and $\log \Theta_\mu$ is concave ($r = 0$, [20]), $\Theta_\mu^{1/(1/r-s)}$ is convex ($r < 0$, [4]) (see also [5]).

PROPOSITION 3.10. *Assume that, for some $r \leq 0$, $\mu \in \mathcal{M}_r$, and that the support of μ is the whole of \mathbb{R}^s . Then (A5) holds.*

PROOF. First we observe that for each nonsingular $B \in L(\mathbb{R}^s, \mathbb{R}^s)$ we have $\mu \circ B \in \mathcal{M}_r$ and $\text{supp}(\mu \circ B) = \mathbb{R}^s$. Hence, it remains to show that for every measure $\mu \in \mathcal{M}_r$ having the property $\text{supp } \mu = \mathbb{R}^s$, its distribution function F_μ is locally Lipschitzian. Since $\mathcal{M}_0 \subset \mathcal{M}_r$, for every $r < 0$, we may assume $r < 0$. We have that $F'_\mu: \mathbb{R}^s \rightarrow (0, +\infty)$ is convex. Let $K \subset \mathbb{R}^s$ be a compact subset, $L(K)$ be a Lipschitz constant for F'_μ on K and L_1 be a Lipschitz constant of the mapping $\xi \mapsto \xi^{1/r}$ on $F'_\mu(K)$. Then we have for all $z, \bar{z} \in K$

$$|F_\mu(z) - F_\mu(\bar{z})| \leq L_1 |F'_\mu(z) - F'_\mu(\bar{z})| \leq L_1 L(K) \|z - \bar{z}\|. \quad \square$$

Special instances of measures $\mu \in \mathcal{P}(\mathbb{R}^s)$ satisfying the hypotheses of Proposition 3.10 are the (nondegenerate) multivariate normal distribution and the multivariate t -distribution (cf. [4, p. 113]).

Acknowledgements. We wish to thank K. Altmann (Humboldt-Universität Berlin), P. Kall (Universität Zürich), B. Kummer (Humboldt-Universität Berlin) and A. Wakolbinger (Johannes Kepler Universität Linz) for beneficial discussions during the work of this paper. We extend our gratitude to the referees for their valuable suggestions. A first draft of this paper was written while the authors were guests of the Institut für Operations Research der Universität Zürich. The second author was

supported by a grant of the Eidgenössische Stipendienkommission für Ausländische Studierende. We should like to thank the mentioned institutions for their hospitality and support.

References

- [1] Bank, B., Guddat, J., Klatte, D., Kummer, B. and Tammer, K. (1982). *Non-Linear Parametric Optimization*. Akademie-Verlag, Berlin.
- [2] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- [3] Birge, J. and Wets, R. J. B. (1986). Designing Approximation Schemes for Stochastic Optimization Problems, In Particular for Stochastic Programs with Recourse. *Math. Programming Study* **27** 54–102.
- [4] Borell, C. (1975). Convex Set Functions in d -space. *Periodica Mathematica Hungarica* **6** 111–136.
- [5] Brascamp, H. J. and Lieb, E. H. (1976). On Extensions of the Brunn-Minkowski and Prékopa-Leindler Theorems, Including Inequalities for Log Concave Functions, and with an Application to the Diffusion Equation. *J. Functional Anal.* **22** 366–389.
- [6] Clarke, F. H. (1983). *Optimization and Nonsmooth Analysis*. Wiley, New York.
- [7] Dudley, R. M. (1976). *Probabilities and Metrics*. Lecture Notes Series No. 45, Aarhus Universitet, Denmark.
- [8] Dupačová, J. (1984). Stability in Stochastic Programming with Recourse-Estimated Parameters. *Math. Programming* **28** 72–83.
- [9] _____ (1987). Stochastic Programming with Incomplete Information: A Survey of Results on Postoptimization and Sensitivity Analysis. *Optimization* **18** 507–532.
- [10] _____ and Wets, R. J. B. (1988). Asymptotic Behaviour of Statistical Estimators and of Optimal Solutions of Stochastic Optimization Problems. *Ann. Statistics* **16** 1517–1549.
- [11] Frauendorfer, K. (1988). Solving SLP Recourse Problems with Arbitrary Multivariate Distributions—The Dependent Case. *Math. Oper. Res.* **13** 377–394.
- [12] Gelbrich, M. (1990). On a Formula for the L^2 -Wasserstein Metric Between Measures on Euclidean and Hilbert Spaces. *Mathematische Nachrichten* **147** 185–203.
- [13] Givens, C. R. and Shortt, R. M. (1984). A Class of Wasserstein Metrics for Probability Distributions. *Michigan Math. J.* **31** 231–240.
- [14] Kall, P. (1976). *Stochastic Linear Programming*. Springer-Verlag, Berlin.
- [15] _____ (1982). Stochastic Programming. *European J. Oper. Res.* **10** 125–130.
- [16] _____ (1987). On Approximations and Stability in Stochastic Programming, *Parametric Optimization and Related Topics* (J. Guddat, H. Th. Jongen, B. Kummer and F. Nožička, Eds.). Akademie-Verlag, Berlin, 387–407.
- [17] _____, Ruszczyński, A. and Frauendorfer, K. (1988). Approximation Techniques in Stochastic Programming, *Numerical Techniques for Stochastic Optimization* (Y. Ermoliev and R. J. B. Wets, Eds.). Springer-Verlag, Berlin, 33–64.
- [18] King, A. J. and Wets, R. J. B. (1991). Epi-consistency of Convex Stochastic Programs. *Stochastics and Stochastics Reports* **34** 83–92; Erratum **37** 259–260.
- [19] Kummer, B. (1991). Lipschitzian Inverse Functions, Directional Derivatives and Applications in $C^{1,1}$ Optimization. *J. Optimization Theory and Applications* **70** 561–582.
- [20] Prékopa, A. (1971). Logarithmic Concave Measures with Applications to Stochastic Programming. *Acta Scientiarum Mathematicarum* **32** 301–316.
- [21] _____ (1980). Logarithmic Concave Measures and Related Topics, *Stochastic Programming* (M. A. H. Dempster, Ed.). Academic Press, London, 63–82.
- [22] Rachev, S. T. (1984). The Monge-Kantorovich Mass Transference Problem and Its Stochastic Applications. *Theory of Probability and Its Applications* **29** 647–676.
- [23] Robinson, S. M. (1975). Stability Theory for Systems of Inequalities, Part I: Linear Systems. *SIAM J. Numerical Anal.* **12** 754–769.
- [24] _____ and Wets, R. J. B. (1987). Stability in Two-stage Stochastic Programming. *SIAM J. Control and Optimization* **25** 1409–1416.
- [25] Römis, W. and Schultz, R. (1991). Distribution Sensitivity in Stochastic Programming. *Math. Programming* **50** 197–226.
- [26] _____ and _____ (1989). Stability of Solutions for Stochastic Programs with Complete Recourse Having $C^{1,1}$ Data. Manuskript, Institut für Operations Research der Universität Zürich, Switzerland.
- [27] _____ and Wakolbinger, A. (1987). Obtaining Convergence Rates for Approximations in Stochastic Programming. *Parametric Optimization and Related Topics* (J. Guddat, H. Th. Jongen, B. Kummer and F. Nožička, Eds.), Akademie-Verlag, Berlin, 327–343.

- [28] Vogel, S. (1988). Stability Results for Stochastic Programming Problems. *Optimization* **19** 269–288.
- [29] Walkup, D. W. and Wets, R. J. B. (1969). Lifting Projections of Convex Polyhedra. *Pacific J. Math.* **28** 465–475.
- [30] Wang, J. (1985). Distribution Sensitivity Analysis for Stochastic Programs with Complete Recourse. *Math. Programming* **31** 286–297.
- [31] Wets, R. J. B. (1974). Stochastic Programs with Fixed Recourse: The Equivalent Deterministic Program. *SIAM Review* **16** 309–339.
- [32] _____ (1989). *Stochastic Programming, Handbooks in Operations Research and Management Science*. Vol. 1, *Optimization* (G. L. Nemhauser, A. H. G. Rinnooy Kan and M. J. Todd, Eds.), North-Holland, Amsterdam, 573–629.

W. Römisch and R. Schultz: Humboldt-Universität Berlin, Fachbereich Mathematik, PSF 1297, D-O-1086 Berlin, Germany