Mathematical Programming 67 (1994) 99–108

Quantitative stability in stochastic programming

Alexander Shapiro

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA

Received 4 February 1991; revised manuscript received 25 February 1994

Abstract

In this paper we study stability of optimal solutions of stochastic programming problems with fixed recourse. An upper bound for the rate of convergence is given in terms of the objective functions of the associated deterministic problems. As an example it is shown how it can be applied to derivation of the Law of Iterated Logarithm for the optimal solutions. It is also shown that in the case of simple recourse this upper bound implies upper Lipschitz continuity of the optimal solutions with respect to the Kolmogorov–Smirnov distance between the corresponding cumulative probability distribution functions.

Keywords: Stochastic programming with recourse; Quantitative stability; Lipschitz continuity; Law of Iterated Logarithm; Kolmogorov–Smirnov distance

1. Introduction

In this paper we study stability of optimal solutions of stochastic programs with respect to perturbations of the underlying probability measure. This problem has been discussed in a number of recent publications (e.g., [4, 5, 10, 14, 18]). In the context of stochastic programming with fixed recourse, Hölder continuity of the set of optimal solutions was established in [13] and [14] with respect to various metrics on the corresponding space of probability measures. It was also argued in these papers that, for the chosen metrics, the exponent $\frac{1}{2}$ in the Hölder continuity result is optimal.

In this article we give a quantitative measure of stability in terms of the objective functions of the associated deterministic optimization problems rather than the underlying probability measures. This has an advantage of a direct analysis of the problem and is not governed by a particular choice of the metric. The case of stochastic programming with fixed recourse will be discussed in detail. It will be shown, as an example, how the obtained upper bound for the rate of convergence of the optimal solutions can be applied to derive the Law of

SSDI 0025-5610(94)00025-O

Iterated Logarithm for those optimal solutions. This upper bound becomes especially convenient when the recourse is simple. We show that in this case the set of optimal solutions is upper Lipschitz continuous with respect to the Kolmogorov–Smirnov distance between the corresponding cumulative probability distribution functions. Eventually this indicates that non-Lipschitzian behavior of the optimal solutions observed in [13, Example 4.6] and [14, Example 2.3] is a property of the chosen metrics rather than the associated optimization problems. Again it immediately follows from our results, for example, that the Law of Iterated Logarithm for the distribution functions carries over to the optimal solutions of the corresponding stochastic programs with simple recourse.

The following notation and terminology will be used throughout the paper. The scalar product of two vectors $x, y \in \mathbb{R}^m$ is denoted by $x \cdot y$. For a set $S \subset \mathbb{R}^m$, $\operatorname{dist}(x, S)$ denotes the distance from x to S with respect to the Euclidean norm $\|x\| = (x \cdot x)^{1/2}$. For a matrix A, the associated (operator) norm is defined as

$$\|A\| = \sup_{\|x\| \leqslant 1} \|Ax\|.$$

The Kolmogorov–Smirnov distance between two cumulative distribution functions $F(t)$ and $G(t)$, $t \in \mathbb{R}$, is given by

$$\|F - G\|_\infty = \sup_{t \in \mathbb{R}} |F(t) - G(t)|.$$

By A^{T} we denote the transpose of a matrix A. For two vectors $x, y \in \mathbb{R}^m$ the inequality $x \leqslant y$ is understood componentwise. We use the notion of the generalized gradient of Clarke [3]. That is, for a locally Lipschitz function $f: \mathbb{R}^m \to \mathbb{R}$, $\partial f(x)$ denotes the convex hull of all limits of the form $\lim_{n \to \infty} \nabla f(x_n)$, where $x_n \to x$ and $x_n \in \mathbb{R}^m \setminus E$ with E being the set of those points where f is not differentiable. Note that the set E has Lebesgue measure zero and that $\partial f(x)$ remains the same if E is replaced by any larger set of Lebesgue measure zero. Note also that if the function $f(x)$ is convex, then the generalized gradient coincides with the subdifferential in the sense of convex analysis [11].

2. Variational principle and stability of stochastic programs with recourse

Let $(\Omega, \mathscr{F})$ be a measurable space, i.e. $\mathscr{F}$ is a σ-algebra of subsets of Ω. Consider a probability measure P on $(\Omega, \mathscr{F})$, a real valued function $g(x, \omega)$ on $\mathbb{R}^m \times \Omega$ and a set $S \subset \mathbb{R}^m$. Suppose that the corresponding expected value function

$$f(x) = \mathbb{E}_P\{g(x, \omega)\} = \int_\Omega g(x, \omega) P(\mathrm{d}\omega)$$

exists for all $x \in S$ and consider the set M of minimizers of $f(x)$ over S. Now let $\hat{P}$ be another probability measure on $(\Omega, \mathscr{F})$ and let $\hat{x}$ be a minimizer over S of the function

$$\psi(x) = \int_\Omega g(x, \omega) \hat{P}(\mathrm{d}\omega).$$

The measure $\hat{P}$ is viewed as an approximation or an estimator of the "true" measure P and the problems of minimization of $f(x)$ and $\psi(x)$ over S are referred to as the true and the approximating programs, respectively.

We use the following variational principle which gives an upper bound for the distance $\operatorname{dist}(\hat{x}, M)$ in terms of the expected value functions $f(x)$ and $\psi(x)$ rather than the corresponding probability measures P and $\hat{P}$. It will be assumed throughout the paper that the optimal set M is nonempty.

Assumption A (Second-order growth condition)**.** There exist a positive constant α and a neighborhood V of the optimal set M such that

$$f(x) \geqslant \inf_{x\in S} f(x) + \alpha[\operatorname{dist}(x, M)]^2 \tag{2.1}$$

for all $x \in S \cap V$.

The following variational principle is given in [16, Lemma 4.1] (see also [17, Section 2]). Suppose that assumption A holds and that $\hat{x} \in V$. Then

$$\operatorname{dist}(\hat{x}, M) \leqslant \alpha^{-1} k(\delta), \tag{2.2}$$

where $\delta(x) = \psi(x) - f(x)$ and

$$k(\delta) = \sup\left\{\frac{|\delta(x) - \delta(y)|}{\|x-y\|}: x\in M,\ y\in S\cap V,\ x\neq y\right\}. \tag{2.3}$$

In the subsequent analysis we deal with functions which are locally Lipschitz. Of course, in case the difference function $\delta(x)$ is Lipschitz continuous in the neighborhood V, the constant $k(\delta)$ is less than or equal to the corresponding Lipschitz constant of $\delta(x)$ and hence is finite. This constant depends on $\delta(x)$ alone and can be estimated further by employing the generalized gradient of $\delta(x)$.

Proposition 2.1. *Suppose that assumption A holds, that $\hat{x} \in V$, that V is convex and that $\delta(x)$ is locally Lipschitz. Then*

$$\operatorname{dist}(\hat{x}, M) \leqslant \alpha^{-1} \kappa(\delta), \tag{2.4}$$

where

$$\kappa(\delta) = \sup\{\|z\|: z\in \partial\delta(x),\ x\in V\}. \tag{2.5}$$

Proof. By the Mean Value Theorem for locally Lipschitz functions [3, p. 41] there is a point x^*, on the segment joining x and y, and $z \in \partial\delta(x^*)$ such that

$$\delta(x) - \delta(y) = z\cdot(x-y).$$

This implies that $k(\delta) \leqslant \kappa(\delta)$ and hence (2.4) follows from (2.2). □

Note that it follows from the definition of the generalized gradient that $\kappa(\delta)$ can be written in the following equivalent form

$$\kappa(\delta) = \sup\{\|\nabla\delta(x)\|: x \in V \setminus E\}, \tag{2.6}$$

with E being a set of Lebesgue measure zero containing the set where $\nabla\delta(x)$ fails to exist. The corresponding upper bound (2.4) will be especially convenient for our purposes.

Assumption A can be ensured by various forms of second-order sufficient conditions (see, e.g., [1], [9, p. 205] and references therein). Standard second-order sufficient conditions imply that the set M consists of a number of locally unique optimal solutions. In case $f(x)$ is differentiable, S is convex and M is closed, a sufficient condition ensuring assumption A is that the inequality

$$f(x) \geqslant f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) + \alpha\|x - \bar{x}\|^2 \tag{2.7}$$

holds for all $x \in S \cap V$ and $\bar{x} \in M$ such that $\text{dist}(x, M) = \|x - \bar{x}\|$ (i.e. $\bar{x}$ is the orthogonal projection of x onto the set M). Indeed, by the first-order necessary conditions we have that $\nabla f(\bar{x}) \cdot (x - \bar{x}) \geqslant 0$ for all $x \in S$, and hence (2.7) implies (2.1). Condition (2.7) is closely related to the concept of strong convexity. It is said that $f(x)$ is strongly convex on a convex set $V \subset \mathbb{R}^m$ if

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y) - \beta\|x - y\|^2$$

for all $x, y \in V$, all $t \in [0, 1]$ and some $\beta > 0$. If $f(x)$ is differentiable, then it is strongly convex on V if and only if

$$f(x) - f(y) \geqslant \nabla f(y) \cdot (x - y) + \beta\|x - y\|^2$$

for all $x, y \in V$ and some $\beta > 0$. If $f(x)$ is strongly convex on a convex neighborhood V of M and the set S is convex, then M is a singleton and assumption A follows. Strong convexity of recourse functionals in stochastic programming is discussed in detail in [14, Section 3].

In some, rather specific situations, condition (2.1) of assumption A can be modified to the more stringent condition

$$f(x) \geqslant \inf_{x \in S} f(x) + \beta \,\text{dist}(x, M), \tag{2.8}$$

for all $x \in S \cap V$ and some $\beta > 0$. This condition, called weak sharp minima, is discussed in [2] where additional references can be found. It is not difficult to see that if (2.8) holds and $\hat{x} \in V$, then $\text{dist}(\hat{x}, M) = 0$ for any function $\psi(x)$ such that the corresponding constant $k(\delta)$ is less than β.

We discuss now an application of the upper bound (2.4) to stochastic programming problems with fixed (linear) recourse. In that case the function $g(x, \omega)$ is given in the form

$$g(x, \omega) = c \cdot x + Q(x, \omega), \tag{2.9}$$

where

$$Q(x, \omega) = \inf\{q \cdot y: Wy = h(\omega) - Ax, y \geqslant 0\}. \tag{2.10}$$

For the sake of simplicity we consider here the situation where only vector $h(\omega)$ is stochastic while matrices W, A and vector q are deterministic.

Consider the function

$$G(t) = \inf\{q \cdot y\colon\ Wy = t,\ y \geqslant 0\}, \tag{2.11}$$

where by the definition $G(t)$ is $+\infty$ if the system $Wy = t$, $y \geqslant 0$, has no solutions. Clearly $Q(x, \omega)$ can be represented then in the form

$$Q(x, \omega) = G(h(\omega) - Ax).$$

Some basic properties of $G(t)$ are discussed in [19]. By the arguments of duality we have that

$$G(t) = \sup_{\gamma \in \Gamma} t \cdot \gamma, \tag{2.12}$$

where Γ is a convex polyhedral set given by

$$\Gamma = \{\gamma\colon\ W^{\mathrm{T}}\gamma \leqslant q\}. \tag{2.13}$$

We assume subsequently that:

(i) For every t the system $Wy = t$, $y \geqslant 0$, has a solution, i.e. the recourse is complete.

(ii) The set Γ is nonempty.

(iii) For all x the expectations $\mathbb{E}_P\{Q(x, \omega)\}$ and $\mathbb{E}_{\hat{P}}\{Q(x, \omega)\}$ are finite.

Note that condition (i) is equivalent to the condition that the system $W^{\mathrm{T}}\gamma \leqslant 0$ has only one solution $\gamma = 0$ or, equivalently, that the set Γ is bounded for every q. Conditions (i) and (ii) imply that the function $G(t)$ is finite valued.

Various properties of $G(t)$ can be derived from the representation (2.12) via convex analysis [11]. In particular we have that $G(t)$ is the support function of the set Γ and that $G(t)$ is a piecewise linear, convex, positively homogeneous function. Its subdifferential is given by

$$\partial G(t) = \operatorname{argmax}\{t \cdot \gamma\colon\ \gamma \in \Gamma\}. \tag{2.14}$$

It follows that the expected value of $Q(x, \omega)$ is a convex function of x and that the subdifferential can be taken inside the expected value, that is

$$\partial \int_{\Omega} Q(x, \omega) P(\mathrm{d}\omega) = -A^{\mathrm{T}} \int_{\Omega} \partial G(h(\omega) - Ax) P(\mathrm{d}\omega). \tag{2.15}$$

A discussion of interchangeability of the subdifferential and integral operators can be found in [8, Section 8.3.3] and [12]. (See also [7, Theorem 12] and [20, Theorem 1.29] for an application of the interchangeability formula to stochastic programming.)

Note that the function

$$v(x) = \mathbb{E}_P\{Q(x, \omega)\} \tag{2.16}$$

is differentiable at x if and only if the corresponding subdifferential, given in the left hand side of (2.15), is a singleton. By (2.15) this subdifferential is a singleton if and only if $\partial G(h(\omega) - Ax)$ is a singleton for P-almost every ω. In the last case we have

$$\nabla v(x) = -A^{\mathrm{T}} \mathbb{E}_P\{\nabla G(h(\omega) - Ax)\}. \tag{2.17}$$

Since a convex function is differentiable everywhere except possibly on a set of Lebesgue measure zero, we have that formula (2.17) holds for almost every x (with respect to Lebesgue measure).

The set Γ, given in (2.13), is a convex, compact, polyhedral set and hence has a finite number of extreme points. Let $\gamma_1, \ldots, \gamma_k$, be the extreme points of Γ. With every extreme point γ_i is associated the normal cone, denoted by $\bar{C}_i$, to Γ at γ_i. Every $\bar{C}_i$, $i=1, \ldots, k$, is a convex, closed, polyhedral cone with a nonempty interior denoted by C_i. Note that $\bar{C}_i$ is the set of those t such that the linear function $l(\gamma) = t \cdot \gamma$ attains its maximum over Γ at the extreme point γ_i and that C_i represents those points t where γ_i is the unique maximizer of $l(\gamma)$ over Γ. It follows that $G(\cdot)$ is differentiable at a point t if and only if $t \in C_i$ for some $i=1, \ldots, k$, in which case $\nabla G(t) = \gamma_i$. We also have that the set formed by nondifferentiable points of $G(\cdot)$ is given by the union of the sets $\bar{C}_i \setminus C_i$, $i=1, \ldots, k$.

From the above discussion and Proposition 2.1 we obtain the following result for stochastic programs with fixed recourse.

Theorem 2.1. *Suppose that assumption A holds, that the neighborhood V is convex and that* $\hat{x} \in V$. *Then*

$$\operatorname{dist}(\hat{x}, M) \leqslant \alpha^{-1} \|A\| \sup_{\chi \in U^*} \left\| \sum_{i=1}^{k} (\hat{p}_i(\chi) - p_i(\chi)) \gamma_i \right\|, \tag{2.18}$$

where

$$p_i(\chi) = P\{h(\omega) - \chi \in C_i\}, \tag{2.19}$$

$$\hat{p}_i(\chi) = \hat{P}\{h(\omega) - \chi \in C_i\} \tag{2.20}$$

and U^* *is the set of those* $\chi = Ax$, $x \in V$, *such that*

$$P\{h(\omega) - \chi \in \bar{C}_i \setminus C_i\} = \hat{P}\{h(\omega) - \chi \in \bar{C}_i \setminus C_i\} = 0, \quad i=1, \ldots, k. \tag{2.21}$$

Proof. Consider the constant $\kappa = \kappa(\delta)$ defined in (2.6) with V being a neighborhood of M specified in assumption A. Denote by $V(P)$ the set of such $x \in V$ that the function G is differentiable at $t = h(\omega) - Ax$ for P-almost every ω. We can write then the set U^* in the form

$$U^* = \{\chi : \chi = Ax, \ x \in V(\hat{P}) \cap V(P)\}. \tag{2.22}$$

It follows that κ can be written in the form

$$\kappa = \|A\| \sup_{\chi \in U^*} \left\| \int_{\Omega} \nabla G(h(\omega) - \chi)(\hat{P} - P)(\mathrm{d}\omega) \right\|. \tag{2.23}$$

Note that U^* is the set of those $\chi = Ax$, $x \in V$, where the integral in (2.23) makes sense and that $V^* = V(P) \cap V(\hat{P})$ is the set of those $x \in V$ where both functions $f(x)$ and $\psi(x)$ are differentiable. Note also that the expected value functions here are convex and hence are locally Lipschitz. Consequently the set $V \setminus V^*$, where $f(x)$ or $\psi(x)$ are not differentiable, has Lebesgue measure zero. We have that if $h-\chi\in C_i$, then $\nabla G(h-\chi)=\gamma_i$. Therefore the inequality (2.18) follows from the inequality (2.4) of Proposition 2.1. □

Note that, because of (2.21), for every $\chi\in U^*$ the probabilities $\hat{p}_i(\chi)$ and $p_i(\chi)$ remain the same if the cone C_i in (2.19) and (2.20) is replaced by its topological closure $\bar{C}_i$.

As an example let us outline how the Law of Iterated Logarithm for stochastic programs with fixed recourse can be easily derived from the upper bound (2.18) and some known results for distribution functions. Let $h_1, \ldots, h_n$ be a random sample of independent, identically distributed random vectors with a common probability measure (distribution) P on the space of vectors h. Consider the corresponding sample (empirical) measure $\hat{P}_n = n^{-1}\sum_{i=1}^{n}\Delta(h_i)$, where $\Delta(h)$ denotes a measure of mass one at the point h, and the associated expected value function $\psi_n(x)$. Let $\hat{x}_n$ be a minimizer of $\psi_n(x)$ over the set S. Suppose that with probability one $\operatorname{dist}(\hat{x}_n, M)\to 0$ as $n\to\infty$. It is not difficult to show that such consistency property of $\hat{x}_n$ holds under mild regularity conditions (cf. [5]). Suppose further that assumption A (for the corresponding true program) is satisfied. Then for n large enough with probability one $\hat{x}_n\in V$ and hence the upper bound (2.18) can be applied. In order to establish the Law of Iterated Logarithm for $\operatorname{dist}(\hat{x}_n, M)$ it will be sufficient now to show that for every $i=1, \ldots, k$,

$$\sup_{\chi}|\hat{P}_n\{h\in\chi+\bar{C}_i\}-P\{h\in\chi+\bar{C}_i\}|=O((\log\log n/n)^{1/2}) \tag{2.24}$$

with probability one. Suppose for the moment that $C_i=\{t\colon t_i\leqslant 0,\ i=1, \ldots, r\}$, with $t_1, \ldots, t_r$, being the first r components of the vector t. In this case the probabilities in (2.24) represent the respective (marginal) distribution functions for the random vector formed by the first r components of the random vector h. Consequently (2.24) follows from the Law of Iterated Logarithm for distribution functions (see, e.g., [15, Section 2.1] for a discussion of the Law of Iterated Logarithm for distribution functions).

In general the polyhedral cones C_i are defined by a finite number of linear constraints, that is $C_i=\{t\colon a_{ij}\cdot t\leqslant 0,\ j=1, \ldots, r_i\}$. Suppose that for every $i=1, \ldots, k$, the respective vectors a_{ij}, $j=1, \ldots, r_i$, are *linearly independent*. Then the general case can be reduced to the particular case above by a suitable linear transformation for every $i=1, \ldots, k$. Therefore we obtain that under this additional assumption of linear independence the Law of Iterated Logarithm follows. That is, $\hat{x}_n$ converges to M at a rate of $O((\log\log n/n)^{1/2})$ almost surely.

3. Stability of optimal solutions of two-stage programs with simple recourse

The constant $\kappa=\kappa(\delta)$ and the associated upper bound take especially simple form in the case of *simple* recourse which we shall discuss in this section. That is, let $W=[I_m, -I_m]$, where I_m denotes the $m\times m$ identity matrix, and let $q=(q^+, q^-)$ be partitioned accordingly. Then

$$\Gamma=\{\gamma\colon -q^-\leqslant\gamma\leqslant q^+\} \tag{3.1}$$

and

$$G(t)=\sum_{i=1}^{m} G_i(t_i), \tag{3.2}$$

where

$$G_i(t_i)=\max\{q_i^+ t_i,\ -q_i^- t_i\} \tag{3.3}$$

with t_i, q_i^+ and q_i^- being components of the respective vectors. We assume that $q^+ + q^- \geqslant 0$, which is equivalent to the condition that the set Γ is nonempty.

The functions $G_i(t_i)$ are differentiable at all $t_i \neq 0$ and the corresponding derivatives are

$$G_i'(t_i)=\begin{cases} q_i^+, & \text{if } t_i>0,\\ -q_i^-, & \text{if } t_i<0.\end{cases}$$

Consider the cumulative distribution functions $F_i(h_i)$ and $\hat{F}_i(h_i)$ of the ith component $h_i(\omega)$ of the random vector $h(\omega)$ corresponding to the probability measures P and $\hat{P}$, respectively. We have then that

$$\int_{-\infty}^{+\infty} G_i'(h_i-\chi_i)\mathrm{d}F_i(h_i)=q_i^+ - (q_i^+ + q_i^-)F_i(\chi_i), \tag{3.4}$$

provided $F_i(h_i)$ is continuous at $h_i=\chi_i$. This leads to the following result.

Theorem 3.1. *Consider the case of two-stage stochastic programming with simple recourse. Suppose that assumption A holds and let $\hat{x}$ be an optimal solution of the approximating program such that $\hat{x}\in V$. Then*

$$\mathrm{dist}(\hat{x}, M) \leqslant \alpha^{-1}\|A\| \sum_{i=1}^{m} (q_i^+ + q_i^-)\|\hat{F}_i - F_i\|_\infty. \tag{3.5}$$

Proof. For an $i\in\{1, \dots, m\}$ consider the cumulative distribution function $F_i(h_i)$ and the function $q(\chi_i, h_i)=G_i(h_i-\chi_i)$. If $F_i(h_i)$ is continuous at a point $h_i=\chi_i$, then $q(\cdot, h_i)$ is differentiable at χ_i for almost every h_i with respect to the distribution F_i, and formula (3.4) holds. Since F_i is a monotonically nondecreasing function, it has at most a countable set of discontinuous points. Consequently, assuming the nontrivial case that $(Ax)_i$ is not identically zero, we have that the set of those x for which F_i is discontinuous at $h_i=(Ax)_i$ has Lebesgue measure zero. Similar arguments and a formula similar to (3.4) apply to $\hat{F}_i$ as well. It follows that

$$\kappa \leqslant \|A\| \sup_{\chi\in U} \sum_{i=1}^{m} (q_i^+ + q_i^-)\,|\hat{F}_i(\chi_i) - F_i(\chi_i)|, \tag{3.6}$$

where U is an appropriate subset of $\mathbb{R}^m$. By replacing U with the larger set $\mathbb{R}^m$ we obtain from (3.6) that

$$\kappa \leqslant \|A\| \sum_{i=1}^{m} (q_i^+ + q_i^-)\|\hat{F}_i - F_i\|_\infty.$$

The above inequality together with Proposition 2.1 imply (3.5). □

The upper bound given by the right hand side of (3.5) is especially convenient since it is presented in terms of the Kolmogorov–Smirnov distance for which classical results are available. (Note that we do not assume here that the distribution functions F_i or $\hat{F}_i$ are continuous.) For example, let $\hat{F}_{in}$, $i=1, \ldots, m$, be sample (empirical) distribution functions corresponding to respective samples of size n. Consider the associated objective function $\psi_n(x)$ of the approximating program and let $\hat{x}_n$ be a minimizer of $\psi_n(x)$ over the set S. Suppose that with probability one $\text{dist}(\hat{x}_n, M) \to 0$ as $n \to \infty$. By the Law of Iterated Logarithm (see, e.g., [15, p. 62]) we have that with probability one

$$\limsup_{n \to \infty} (2n/\log \log n)^{1/2} \|\hat{F}_{in} - F_i\|_\infty \leqslant 1.$$

(Note that the above lim sup is equal to one almost surely if F_i is continuous.) We obtain then that, under assumption A, the distance $\text{dist}(\hat{x}_n, M)$ is of order $O((\log \log n/n)^{1/2})$ with probability one. Under certain regularity conditions, in particular if F_i is sufficiently smooth, the Law of Iterated Logarithm also holds for kernel-type estimators of F_i (see [21]). Via upper bound (3.5) this leads to the same rate of convergence of the corresponding minimizers (cf. [6]).

Finally let us make the following remarks. It follows from (3.4) that here the Hessian matrix $\nabla^2 f(x)$ is equal to $A^{\mathrm{T}}DA$, where D is a diagonal matrix with diagonal elements $(q_i^+ + q_i^-) f_i(\chi_i)$, $i=1, \ldots, m$, $\chi = Ax$, provided the densities $f_i(\chi_i) = F_i'(\chi_i)$ do exist. Clearly this Hessian matrix is positive definite if and only if all diagonal elements of the matrix D are positive and A has full column rank. In the last case assumption A will follow provided the set S is convex.

References

[1] J.F. Bonnans and A.D. Ioffe, "Second-order sufficiency and quadratic growth for non isolated minima," preprint, 1993.

[2] J.V. Burke and M.C. Ferris, "Weak sharp minima in mathematical programming," Computer Sciences Technical Report 1050, University of Wisconsin-Madison, 1991.

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

[4] J. Dupačová, "Stability and sensitivity analysis for stochastic programming," *Annals of Operations Research* 27 (1990) 115–142.

[5] J. Dupačová and R.J.B. Wets "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems," *The Annals of Statistics* 16 (1988) 1517–1549.

[6] N. Gröwe and W. Römisch, "A stochastic programming model for optimal power dispatch: stability and numerical treatment," in K. Marti ed., *Stochastic Optimization: Numerical Methods and Technical Applications*, Lecture Notes in Economics and Mathematical Systems (Springer, Berlin, 1992) pp. 111–139.

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

[8] A.D. Ioffe and V.M. Tihomirov, *Theory of Extremal Problems* (North-Holland Publishing Company, Amsterdam, 1979).

[9] S.M. Robinson, "Generalized equations and their solutions, Part II: applications to nonlinear programming," *Mathematical Programming Study* 19 (1982) 200–221.

[10] S.M. Robinson and R.J.B. Wets, "Stability in two-stage stochastic programming," *SIAM Journal on Control and Optimization* 25 (1987) 1409–1416.

[11] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[12] R.T. Rockafellar and R.J.B. Wets, "On the interchange of subdifferentiation and conditional expectation for convex functionals," *Stochastics* 7 (1982) 173–182.

[13] W. Römisch and R. Schultz, "Distribution sensitivity in stochastic programming," *Mathematical Programming* 50 (1991) 197–226.

[14] W. Römisch and R. Schultz, "Stability of solutions for stochastic programs with complete recourse," *Mathematics of Operations Research* 18 (1993) 590–609.

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

[16] A. Shapiro, "Asymptotic analysis of stochastic programs," *Annals of Operations Research* 30 (1991) 169–186.

[17] A. Shapiro, "Perturbation analysis of optimization problems in Banach spaces," *Numerical Functional Analysis and Optimization* 13 (1992) 97–116.

[18] J. Wang, "Distribution sensitivity analysis for stochastic programs with complete recourse," *Mathematical Programming* 31 (1985) 286–297.

[19] R.J.B. Wets, "Stochastic programs with fixed recourse: The equivalent deterministic program," *SIAM Review* 16 (1974) 309–339.

[20] R.J.B. Wets, "Stochastic programming: Solution techniques and approximation schemes," in A. Bachem, A. Grötschel and B. Korte, ed., *Mathematical Programming. The State of the Art* (Springer, Berlin, 1983) pp. 566–603.

[21] B.B. Winter, "Convergence rate of perturbed empirical distribution functions," *Journal of Applied Probability* 16 (1979) 163–173.