ASYMPTOTIC DISTRIBUTIONS FOR SOLUTIONS IN STOCHASTIC OPTIMIZATION AND GENERALIZED *M*-ESTIMATION

Alan J. King

July 1988
WP-88-58

IIASA
International Institute
for Applied Systems Analysis

ASYMPTOTIC DISTRIBUTIONS FOR SOLUTIONS IN STOCHASTIC OPTIMIZATION AND GENERALIZED *M*-ESTIMATION

Alan J. King

July 1988
WP-88-58

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

New techniques of local sensitivity analysis in nonsmooth optimization are applied to the problem of determining the asymptotic distribution (generally non-normal) for solutions in stochastic optimization, and generalized M-estimation – a reformulation of the traditional maximum likelihood problem that allows the introduction of hard constraints.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

ASYMPTOTIC DISTRIBUTIONS FOR SOLUTIONS IN STOCHASTIC OPTIMIZATION AND GENERALIZED *M*-ESTIMATION

Alan J. King[*]

Abstract. New techniques of local sensitivity analysis in nonsmooth optimization are applied to the problem of determining the asymptotic distribution (generally non-normal) for soutions in stochastic optimization, and generalized M-estimation — a reformulation of the traditional maximum likelihood problem that allows the introduction of hard constraints.

Keywords: stochastic programs, generalized equations, asymptotic distribution, contingent-derivative, strong monotonicity.

[*] International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

1. Introduction

Many problem formulations in statistics and stochastic optimization generate estimates from data by selecting a "best" or "optimal" point $\underset{\sim}{x}^\nu = x^\nu(\xi_1, \ldots, \xi_\nu)$, frequently by choosing $\underset{\sim}{x}^\nu$ to solve a *generalized equation* in the form

$$\text{(1.1)} \qquad \text{Choose } x \in \mathbb{R}^n \text{ such that } 0 \in \tfrac{1}{\nu} \sum_{i=1}^{\nu} g(x, \xi_i) + N(x),$$

where g is a function, $\{\xi_i\}$ an i.i.d. sequence of random variables, and N a set-valued mapping. In stochastic programming, for example, this equation represents the first-order necessary conditions for the optimization problem

$$\text{minimize } \tfrac{1}{\nu}\Sigma f(x, \xi_i) \text{ over all } x \in X \subset \mathbb{R}^n,$$

setting $\Gamma(x, \xi) = \nabla f(x, \xi)$ and $N(x) = N_X(x)$ — the normal cone to X at x in the sense of nonsmooth analysis. In maximum likelihood estimation this equation can represent the so-called "normal equations", setting $N(x)$ identically equal to the zero vector; by analogy with stochastic optimization, this situation represents the case where no "hard" (i.e. *a priori* deterministic) constraints are placed on the maximum likelihood estimator. In the general case, solutions to (1.1) could be called generalized M-estimates.

Introducing a set-valued map into the normal equations is natural from the point of view of optimization, because it permits the specification of constraints that one knows must be true (e.g. non-negativity in variance estimation), but at the same time it complicates the analysis of the asymptotic behaviour of the estimates. For more discussion and motivating examples, see Dupačová and Wets [**8**].

In this paper we develop assumptions under which there are a point x^* and a random variable $\underset{\sim}{u}$ such that $\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*)$ converges in distribution to $\underset{\sim}{u}$; furthermore, we also indicate how to compute $\underset{\sim}{u}$ from the information in (1.1). If $\underset{\sim}{u}$ turns out to be normal, then $\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*)$ is *asymptotically normal*. The presence of the set-valued mapping N, however, leads to asymptotic behaviour that is generally non-normal but that can be analyzed using the special techniques of this paper.

New developments in nonsmooth analysis, in particular the differentiation of set-valued maps, makes possible a very general study of the solutions to (1.1) by analyzing the local sensitivity of the mapping

$$\text{(1.2)} \qquad J(z) = \{x \in \mathbb{R}^n \mid 0 \in z(x) + N(x)\}$$

about $Eg(\cdot) = Eg(\cdot, \xi_1)$ (expectation with respect to the random variable ξ_1), where the perturbations are taken over a function space Z — in the case considered here, a

Banach space. The functions $E^\nu g(\cdot) = \frac{1}{\nu}\sum_{i=1}^{\nu} g(\cdot, \xi_i)$ may be viewed as random variables with values in Z, and a Banach space central limit theorem may be applied to reach the conclusion that $\sqrt{\nu}(E^\nu g - Eg)$ converges in distribution to a Gaussian random variable $\underset{\sim}{c}$. By analogy with the classical delta method, we then define an appropriate "derivative" of J at Eg and conclude that

$$\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*) \xrightarrow[\mathcal{D}]{} J'_{Eg}(\underset{\sim}{c}). \tag{1.3}$$

The basic pattern in this argument is the "generalized delta method" described in Section 2.

In this paper we derive (1.3) for $Z = \mathcal{C}(\mathbb{R}^n : \mathbb{R}^n)$, the space of bounded continuous functions from $\mathbb{R}^n$ into $\mathbb{R}^n$, $Eg : \mathbb{R}^n \to \mathbb{R}^n$ strongly monotone, and $N : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ maximal monotone. This setting does not cover all situations in stochastic programming or generalized M-estimation, but it seems at present to be the most general in which $\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*)$ can be expected to converge in distribution.

All of the early results yielding asymptotic distributions for solutions to (1.1) were developed for maximum likelihood estimation. Few papers in this field, with the notable exception of Aitchison and Silvey [**1**], considered constrained problems (and even the exceptional case had asymptotic normality as its goal). In stochastic optimization, constraints are fundamental to modelling practical decision problems. Currently there are three approaches to deriving asymptotic distributions for solutions in stochastic optimization and generalized M-estimation. One technique is based on the fundamental paper of Huber [**9**], whose result has been applied recently to stochastic optimization in Dupačová and Wets [**8**]. Essentially this technique allows one to pass to parametric analysis by assuming asymptotic normality of $E^\nu g(\underset{\sim}{x}^\nu)$. A second technique is based on the "von Mises functionals". A basic reference is Kallianpur [**10**], and a recent paper applying this concept to non-smooth generalized equations involving Clarke subgradients is (a different) Clarke [**7**]. Finally, there is the one on which the present paper is based; it was first outlined in King [**11**], where results for linear-quadratic problems were given. There are strong connections between the techniques and also some differences. Huber apparently does not require monotonicity or even continuity of $g(\cdot, \xi)$ — but every application of his result imposes these and much more. If these assumptions are granted then the results presented here are more general than those based on Huber's theorem.

The main result is presented in Section 4 along with an example and discussion. Section 2 contains the basics of the generalized delta method, and Section 3 the local analysis of the mapping J. A brief presentation of the Banach space central limit theorem appears in an Appendix. There are many concepts and definitions needed for the smooth reading of this paper, and not all readers can be expected to be fully versed in each. Accordingly, some brief space has been allotted to a description of the major prerequisites.

2. Generalized Delta Method

The definition of the mapping J in (1.2) allows us to generate the asymptotic distribution for the solution sequence based on that of the sequence of functions $\{z^\nu(\cdot)\}$, which as we have indicated, are to be regarded as elements of a Banach space Z equipped with the Borel sets $\mathcal{B}$. A discussion of central limit theory in Banach spaces appears in the Appendix. For the purposes of ths section we shall assume that there are $z^* \in Z$ and a Z-valued random variable $\underset{\sim}{w}$ with

$$\sqrt{\nu}(\underset{\sim}{z}^\nu - z^*) \xrightarrow[\mathcal{D}]{} \underset{\sim}{w}, \tag{2.1}$$

where the symbol $\mathcal{D}$ under the arrow denotes *convergence in distribution* (weak $*$-convergence of the measures μ^ν induced on Z by the random variables $\sqrt{\nu}(\underset{\sim}{z}^\nu - z^*)$ to the measure μ induced by $\underset{\sim}{w}$, which means that

$$\int_Z (v(z)\mu^\nu(dz) \to \int_Z f(z)\mu(dz)$$

for all bounded continuous $f : Z \to \mathbb{R}$, cf. Billingsley [**5**]).

The "generalized delta method" to be described in this section is a review of the theory in King [**12**] that gives conditions under which the asymptotic distribution of $\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*)$, for $\underset{\sim}{x}^\nu \in G(\underset{\sim}{z}^\nu)$, can be deduced from the limit distribution $\underset{\sim}{w}$ and the first-order behaviour of G, where $G : Z \rightrightarrows X$ is a given set-valued mapping. It takes the point of view that $\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*)$ is a selection from the "difference quotients" $\sqrt{\nu}(G(\underset{\sim}{z}^\nu) - x^*)$, i.e.

$$\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*) \in \sqrt{\nu}(G(\underset{\sim}{z}^\nu) - x^*). \tag{2.2}$$

It shows first that under special circumstances the difference quotients converge in distribution as closed-valued measurable multifunctions, and, second, passes to a more specialized situation where conclusions about the limit distribution of $\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*)$ may be drawn from that of the difference quotients. The combined result will be summarized in Theorem 2.2.

The concept on which the theory is based is the convergence of closed sets in $\mathbb{R}^n$. Let $\{A_\nu\}$ be a sequence of closed subsets of $\mathbb{R}^n$ and define the (closed) sets

$$\liminf_\nu A_\nu = \{x = \lim x_\nu \mid x_\nu \in A_\nu \text{ for all but finitely many } \nu\} \tag{2.3}$$

$$\limsup_{\nu} A_\nu = \{x = \lim x_\nu \mid x_\nu \in A_\nu \text{ for infinitely many } \nu\}. \tag{2.4}$$

We say $\{A_\nu\}$ *set converges to* $A = \lim_\nu A_\nu$, if $A = \liminf A_\nu = \limsup A_\nu$. Set-convergence induces a compact, separable, and metrizable topology on the space $\mathcal{F}$ of closed subsets of $\mathbb{R}^n$. A *closed-valued multifunction* can be viewed either as a set-valued mapping $G : Z \rightrightarrows \mathbb{R}^n$ or as a function $\gamma_G : Z \to \mathcal{F}$. If $(Z, \mathcal{B})$ is a measurable space then a closed-valued multifunction $G : Z \rightrightarrows \mathbb{R}^n$ is *measurable* if for all $C \in \mathcal{F}$ one has $G^{-1}(C) := \{z \in Z \mid G(z) \cap C \neq \emptyset\}$ belongs to $\mathcal{B}$. (When the probability space is not explicit, we use the term *random closed set* and employ the notation $\underset{\sim}{G}$.) Equivalently, such a G is measurable if and only if γ_G is a Borel measurable function. If the measurable space $(Z, \mathcal{B})$ comes equipped with a measure μ, then we say that a sequence of closed-valued measurable multifunctions $\{G_\nu\}$ *converges in distribution* to G if and ony if $\{\gamma_{G_\nu}\}$ converges in distribution to γ_G. This definition is due to Salinetti and Wets [**14**], which paper is recommended to the reader who wishes a more detailed exposition of the topics of this paragraph.

A *measurable selection* g of a multifunction $G : Z \rightrightarrows \mathbb{R}^n$ is a measurable function $g : \operatorname{dom} G \to \mathbb{R}^n$ such that $g(z) \in G(z)$ for all $z \in \operatorname{dom} G$, where $\operatorname{dom} G$, the *domain* of G, is the set $G^{-1}(\mathbb{R}^n) = \{z \mid G(z) \neq \emptyset\}$. A closed-valued measurable multifunction always has selections; cf. [**6**], for example. Convergence in distribution of selections of a converging sequence of multifunctions has been studied in [**12**].

We pass next to a brief outline of the concepts of "local behaviour" of a multifunction needed for the results of this paper.

We say that a multifunction $G : Z \rightrightarrows \mathbb{R}^n$ is *locally upper Lipschitzian* at a point z if there are a modulus $\lambda \geq 0$ and a neighborhood U of z such that

$$G(z) \subset G(z) + \lambda \|z - z'\| B, \quad \forall z' \in U, \tag{2.5}$$

where B is the open ball in $\mathbb{R}^n$ and $\|\cdot\|$ is the norm in Z. This definition is due to Robinson [**15**]. The following geometric notion of a derivative of a set-valued mapping, modelled after the original tangency constructions of Fermat, has been recently introduced by Aubin [**4**]. The *contingent derivative* of a multivalued mapping $G : Z \rightrightarrows \mathbb{R}^n$ at a point $z \in \operatorname{dom} G$ and $x \in G(z)$ is the mapping $G^+_{z,x}$ whose graph is the *contingent cone* to the graph of G at $(z, x) \in Z \times \mathbb{R}^n$, i.e.

$$\limsup_{t \downarrow 0} t^{-1}[\operatorname{gph} G - (z, x)] = \operatorname{gph} G^+_{z,x}, \tag{2.6}$$

where we denote by $\operatorname{gph} G$ the set $\{(z, x) \in Z \times \mathbb{R}^n \mid x \in G(z)\}$. The contingent derivative always exists, because the lim sup of a net of sets always exists; and it has *closed graph* (equivalently, is *lower semi-continuous*), since the lim sup is always a closed set. This latter property implies that G^+ is closed-valued and measurable [**6**; III.3].

It is worth noting here some further properties and concepts related to the contingent derivative. If one has $\limsup = \liminf$ in (2.6), then, following Rockafellar [**21**], we say that G is ***proto-differentiable*** at (z, x) and we call the common limit the ***proto-derivative***, denoted $G'_{z,x}$. A stronger property that is related to true differentiability (for functions) is *semi-differentiability*, which requires that the limit

$$\lim_{\substack{t\downarrow 0\\ w'\to w}} (G(z+tw')-x)/t \tag{2.7}$$

exists for all w. When it does, it equals the proto-derivative $G'_{z,x}(w)$. For μ a Borel measure on Z, we say that G is *μ-a.s. semi-differentiable* if (2.7) holds for all w except possibly those in a set of μ-measure zero. There are strong connections between semi-differentiability and convergence in distribution for the sequence of "difference quotients", as we shall see in a moment.

We present first a result needed for the computation of contingent derivatives. From the definition, it is clear that $G^+(w)$ contains the lim sup of the difference quotients taken along the single ray $\{tw : t > 0\}$, i.e.

$$\limsup_{t\downarrow 0}(G(z+tw)-x)/t \subset G^+_{z,x}(w). \tag{2.8}$$

To obtain equality in (2.8) one requires Lipschitzian and differentiability properties of G, as in [**21**; Section 5]. For the situations considered in this paper, where single-valuedness plays a strong role, one has the following result. We say that a closed-valued measurable multifunction $G : Z \rightrightarrows \mathbb{R}^n$ is *(a.s.) single-valued* if the set $\{z \in \operatorname{dom} G \mid G(z) \text{ is not a singleton}\}$ is empty (a set of measure zero).

Proposition 2.1. *Let $G : Z \rightrightarrows \mathbb{R}^n$ be locally upper Lipschitzian and single-valued at z, with $G(z) = \{x\}$. If the contingent derivative $G^+_{z,x}$ is (a.s.) single-valued, then G is (a.s.) semi-differentiable at (z, x) and*

$$\limsup_{t\downarrow 0}(G(z+tw)-x)/t = G'_{z,x}(w) \tag{2.9}$$

for (almost) all $w \in Z$.

Proof. All conclusions except (2.9) are in [**12**; 4.1]; and (2.9) is a simple corollary of that proof. □

We are now ready to state the main convergence result. The proof is identical to that in [**12**; 4.3] and will not be given here.

Theorem 2.2. *Let the sequence $\{\underset{\sim}{z}_\nu\}$ of the random variables in the Banach space Z satisfy a central limit property*

$$\sqrt{\nu}(\underset{\sim}{z}_\nu - z^*) \xrightarrow[\mathcal{D}]{} \underset{\sim}{w}; \tag{2.10}$$

and let the closed-valued measurable multifunction $G : Z \rightrightarrows \mathbb{R}^n$ be locally upper Lipschitzian and single-valued at z^*, with $\{x^*\} = G(z^*)$. Suppose further that:

$$z^* \in \operatorname{int}\operatorname{dom} G; \tag{2.11}$$

$$G^+_{z^*,x^*}(\underset{\sim}{w}) \text{ is a.s. single-valued.} \tag{2.12}$$

Then G is semi-differentiable at (z^*, x^*) and for all measurable selections $\underset{\sim}{x}^\nu$ of $G(\underset{\sim}{z}^\nu)$ and $\underset{\sim}{u}$ of $G'_{z^*,x^*}(\underset{\sim}{w})$ one has

$$\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*) \xrightarrow[\mathcal{D}]{} \underset{\sim}{u} \tag{2.13}$$

as random variables in $\mathbb{R}^n$. □

Remark 2.3. The assumption (2.12) implies by the Appendix of [**12**] that any selection of $\underset{\sim}{u}$ of $G^+(\underset{\sim}{w})$ is measurable. Furthermore, if any other multifunction F has $F(\underset{\sim}{w}) \supset G^+(\underset{\sim}{w})$ a.s. and $F(\underset{\sim}{w})$ is a.s. single-valued, then any selection of $F(\underset{\sim}{w})$ will also satisfy (2.13) — the reason being that (2.11) implies $\operatorname{dom} G^+ = Z$, see [**12**; 4.2], and thus $F(\underset{\sim}{w}) = G^+(\underset{\sim}{w})$ a.s.

3. Computation of the Contingent Derivative

The subject of this section is a simple computation of the contingent-derivative of the mapping J defined by (1.2), under continuity and monotonicity assumptions on z^* and N. The results in this section appeared in part in King [**11**].

A mapping $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be a *monotone* operator if for all points $x \in \operatorname{dom} T$ and $y \in T(x)$ one has

$$(x - x') \cdot (y - y') \geq 0 \quad \forall x' \in \operatorname{dom} T,\ \forall y' \in T(x'). \tag{3.1}$$

A monotone operator T is said to be *maximal monotone* if $\operatorname{gph} T$ is maximal in the partial ordering (by subsets) of all monotone operators whose graph contains $\operatorname{gph} T$. The most important examples of maximal monotone operators are the subgradients of convex functions; cf. Rockafellar [**18**]. A continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$ that is monotone is maximal monotone, and a maximal monotone operator that is everywhere single-valued is continuous. Monotone operators T with the property that for some sufficiently small $\delta > 0$ the operator $T - \delta I$ is monotone, where $Ix = x$ is the identity operator, are said to be *strongly monotone.* Strong monotonicity imparts stability to generalized operators, as the following proposition shows.

Proposition 3.1. *Let $N : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone operator, and let $z^* \in \mathcal{C}(\mathbb{R}^n : \mathbb{R}^n)$ be strongly monotone on $X = \operatorname{dom} N$. Then the solution mapping*

$$J(z) = \{x \in \mathbb{R}^n \,|\, 0 \in z(x) + N(x)\}$$

is single-valued at z^*, and for all bounded neighborhoods D of $J(z^*)$, the mapping J is nonempty, single-valued, and locally upper Lipschitzian on a neighborhood of z^* in $\mathcal{C}(D : \mathbb{R}^n)$, the Banach space of continuous functions $z : D \to \mathbb{R}^n$ equipped with the sup norm.

Proof. Since z^* is strongly monotone, there exists $\delta > 0$ such that $F(x) := -z^*(x) - \delta x$ is monotone on X. We may rewrite $J(z^*)$ as $G^{-1}(0)$, where $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined to be $G(x) = F(x) + \delta x + N(x)$. By Rockafellar [**17**], $F + N$ is maximal monotone on X, and a result of Minty [**13**] allows us to conclude that G^{-1} is single-valued and Lipschitz continuous on all of $\mathbb{R}^n$, with global Lipschitz modulus δ^{-1}. In particular, $G^{-1}(0) = J(z^*)$ is a singleton. Let D be a fixed bounded neighborhood of $J(z^*)$, and let $z \in \mathcal{C}(D : \mathbb{R}^n)$ be such that $\alpha = \sup\{|z(x) - z^*(x)| \,|\, x \in D\}$ is finite. We have

$$J(z) \subset \bigcup_{a \in \alpha B} G^{-1}(a),$$

where B is the open unit ball in $\mathbb{R}^n$. It follows that

$$J(z) \subset J(z^*) + \delta^{-1}\alpha B.$$

Hence, J is locally upper Lipschitzian as a mapping from $\mathcal{C}(D : \mathbb{R}^n)$ into $\mathbb{R}^n$. Define now the continuous function $J_z : \mathbb{R}^n \to \mathbb{R}^n$ by

$$J_z(x) = G^{-1}(z^*(x) - z(x)).$$

If we choose $z \in \mathcal{C}(D : \mathbb{R}^n)$ with $\sup\{|z^*(x) - z(x)| \,|\, x \in D\}$ less than δ, then J_z is a contraction mapping with a unique fixed point $x_z = J_z(x_z)$ in D. This x_z is obviously the only point in $J(z)$, and the proof is complete. □

In the rest of this paper we shall suppose that $Z = \mathcal{C}(D : \mathbb{R}^n)$ for some suitable bounded set D, since the assumptions of 3.1 will always be in force. The computation of the contingent derivative is our next task. We shall give two results — an estimate in the general case for nondifferentiable z^* and N, and a more precise result when differentiability assumptions hold.

Proposition 3.2. *Let J be as in 3.1 and define the multifunction $F : Z \rightrightarrows \mathrm{I\!R}^n$ by*

$$F(w) = \{u \in \mathrm{I\!R}^n \mid 0 \in (z^*)^+_{x^*}(u) + w(x^*) + N^+_{x^*,-z^*(x^*)}(u)\}. \tag{3.2}$$

Suppose that z^* is locally upper Lipschitzian at x^*. Then F is closed-valued and measurable, and

$$\operatorname{gph} J^+_{z^*,x^*} \subset \operatorname{gph} F. \tag{3.3}$$

Proof. A pair (w, u) lies in the graph of G^+ if and only if there are sequences $\{t_\nu\}$, $\{w^\nu\}$, and $\{u^\nu\}$ with $t_\nu \downarrow 0$, $w^\nu \to w$ (in Z), and $u^\nu \to u$ (in $\mathrm{I\!R}^n$), respectively, satisfying

$$(w^\nu, u^\nu) \in t_\nu^{-1}[\operatorname{gph} J - (z^*, x^*)];$$

this implies

$$0 \in z^*(x^* + t_\nu u^\nu) + t_\nu w^\nu(x^* + t_\nu u^\nu) + N(x^* + t_\nu u^\nu).$$

For each ν, there is a point $a^\nu \in \mathrm{I\!R}^n$ such that

$$a^\nu = t_\nu^{-1}[z^*(x^* + t_\nu u^\nu) - z^*(x^*)] \tag{3.4}$$

and

$$-[a^\nu + w^\nu(x^* + t_\nu u^\nu)] \in t_\nu^{-1}[N(x^* + t_\nu u^\nu) + z^*(x^*)]. \tag{3.5}$$

Since z^* is locally upper Lipschitzian at x^*, the sequence $\{a^\nu\}$ must have cluster points and all these cluster points belong to $(z^*)^+_{x^*}(u)$ by (3.4) and (2.6). The sequence w^ν converges to w in the sup norm in Z, hence in particular $w^\nu(x^* + t_\nu u^\nu) \to w(x^*)$. From this and (3.4) it follows that the given point $(w, u) \in \operatorname{gph} J^+$ satisfies $u \in F(w)$, proving (3.3). That F is closed-valued and measurable will follow from F having closed graph [**6**; III.3], and so we now prove this latter claim. Let each element of the sequence of pairs $\{(w^\nu, u^\nu)\}$ belong to $\operatorname{gph} F$ and suppose $(w^\nu, u^\nu) \to (w, u)$. We aim to show $(w, u) \in \operatorname{gph} F$. Let a^ν satisfy

$$a^\nu - w^\nu(x^*) \in (z^*)^+_{x^*}(u^\nu)$$

and

$$-a^\nu \in N_{x^*,-z(x^*)}(u^*).$$

Since z^* is locally upper Lipschitzian $(z^*)^+_{x^*}$ is locally bounded, so $\{a^\nu\}$ has a cluster point, say a, that satisfies

$$a \in (z^*)^+_{x^*}(u) + w(x^*).$$

Passing to a subsequence if necessary we have $a^\nu \to a$ and $u^\nu \to u$ with $(u^\nu, -a^\nu) \in \operatorname{gph} N^+$. But N^+ has closed graph so $-a \in N^+(u)$, and the proof is complete. □

If a given locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^n$ has the property that $f^+_{x^*}$ is single-valued everywhere then Proposition 2.1 states that f is actually semi-differentiable at x^*. It is easy to see that this is equivalent to the property

$$\lim_{\substack{t \downarrow 0 \\ u' \to u}} \frac{f(x^* + tu') - f(x^*)}{t} = f^+_{x^*}(u), \tag{3.6}$$

and when this occurs we say, following Rockafellar [**20**], that f is *directionally differentiable in the Hadamard sense* at x^* and $f^+_{x^*}(\cdot)$ equals the *directional derivative* $f'(x^*;\cdot)$. (It is well known that if f is directionally differentiable in the *ordinary sense*, i.e.

$$\lim_{t \downarrow 0} \frac{f(x^* + tu) - f(x^*)}{t} = f'(x^*; u), \tag{3.7}$$

and if $f'(x^*;\cdot)$ is continuous, then f is also directionally differentiable in the Hadamard sense. This simplifies the verification of (3.6).) This derivative has also been studied in Robinson [**16**], where it was called the *Bouligand derivative.*

The computation in Proposition 3.2 can now be made precise by making differentiability assumptions on z^* and N.

Proposition 3.3. *Suppose that z^* and N satisfy the conditions of Propositions 3.1 and 3.2. Suppose moreover that z^* is directionally differentiable in the Hadamard sense at x^* and N is proto-differentiable at $(x^*, -z^*(x^*))$. Then J is semi-differentiable at z^*, where for each $w \in Z$ one has*

$$J'_{z^*,x^*}(w) = \{u \in \mathbb{R}^n \mid 0 \in (z^*)'(x^*; u) + w(x^*) + N'_{x^*,-z^*(x^*)}(u)\},$$

and J'_{z^*,x^*} is single-valued everywhere.

Proof. The proto-derivative of a maximal monotone operator is the graph limit of the sequence of maximal monotone "difference quotient" operators, thus is itself maximal monotone; cf. Attouch [**3**]. Thus N' is maximal monotone, and $(z^*)'(x^*;\cdot)$ is evidently continuous and strongly monotone. Applying Proposition 3.1 to the multifunction F in (3.2) we find that F is everywhere single-valued and from Proposition 3.2 we know that $J^+(w) = F(w)$ for all w (since $\operatorname{dom} J^+ = Z$ by [**12**; 4.2]). Now apply Proposition 2.1 to J and conclude that J is actually semi-differentiable. □

4. Asymptotics

The main theorem is presented in this section along with illustrative examples. The target is the analysis of the sequence of solutions $\{\underset{\sim}{x}^\nu\}$ to the problem (1.1). We shall treat the $\underset{\sim}{x}^\nu$ as selections of the solution mapping J evaluated at $E^\nu g(\cdot) = \frac{1}{\nu}\sum_{i=1}^{\nu} g(\cdot,\xi_i)$, and the result will be based on the asymptotic properties of $E^\nu g$ and local properties of J about $Eg(\cdot) = \int g(\cdot,\xi)P(d\xi)$.

There are two sets of assumptions. One set of assumptions delivers the asymptotic normality of $\sqrt{\nu}(E^\nu g - Eg)$; the other set assures enough local "regularity" of the mapping J needed to apply Theorem 2.1. The assumptions interact to some extent. In particular we may suppose that everything of interest is happening in a bounded subset D of $\mathbb{R}^n$ — we shall return to this point below.

Probabilistic Assumptions

P.1 For all $x \in D$, the function $g(x,\cdot) : (\Xi, \mathcal{A}) \to \mathbb{R}^n$ is measurable.

P.2 There is some $a : \Xi \to \mathbb{R}$ with $\int_\Xi |a(\xi)|^2 P(d\xi) < \infty$ and

$$|g(x_1,\xi) - g(x_2,\xi)| \le a(\xi)|x_1 - x_2| \quad \forall x_1, x_2 \in D.$$

P.3 There is some $x \in D$ with $\int_\Xi |g(x,\xi)|^2 P(d\xi) < \infty$.

In the Appendix we show that these assumptions imply that the functions $E^\nu g$ are $\mathcal{C}(D : \mathbb{R}^n)$-valued random variables, that $Eg \in \mathcal{C}(D : \mathbb{R}^n)$, and that

$$\sqrt{\nu}(E^\nu g - Eg) \xrightarrow[\mathcal{D}]{} \underset{\sim}{w}, \tag{4.2}$$

where $\underset{\sim}{w}$ is a centered Gaussian $\mathcal{C}(D : \mathbb{R}^n)$-valued random variable with covariance equal to that of $g(\cdot, \xi_1)$.

Analytical Assumptions

A.1 The function $Eg : \mathrm{I\!R}^n \to \mathrm{I\!R}^n$ is strongly monotone on $\operatorname{dom} N$.

A.2 The operator $N : \mathrm{I\!R}^n \rightrightarrows \mathrm{I\!R}^n$ is maximal monotone.

The assumptions A.1 and P.2 imply that Eg is continuous and strongly monotone on $\operatorname{dom} N$; it therefore follows from A.2 and Proposition 3.1 that there is a unique point x^* satisfying $0 \in Eg(x^*) + N(x^*)$, and that we may without loss of generality view the perturbations $E^\nu g$ about Eg as taking place in the Banach space $\mathcal{C}(D : \mathrm{I\!R}^n)$, where D is any bounded neighborhood of x^*.

The main result now follows. This result represents the first application in the literature of the generalized delta method and the generalized differentiability techniques for set-valued maps to the problem of determining the asymptotic distribution of solutions to generalized equations of the form (1.1). The result was foreshadowed in King [11] but has since been much improved.

Theorem 4.1. *Suppose the assumptions P.1–3 and A.1–2 hold. Suppose further that the random closed set $\underset{\sim}{F}$ is a.s. single-valued, where*

$$\underset{\sim}{F} = \{u \in \mathrm{I\!R}^n \,|\, 0 \in \underset{\sim}{c} + (Eg)^+_{x^*}(u) + N^+_{x^*,-Eg(x^*)}(u)\} \tag{4.3}$$

and where $\underset{\sim}{c}$ is a normally distributed $\mathrm{I\!R}^n$-valued random variable with covariance matrix $\Sigma = \int [g(x^*,\xi) - Eg(x^*)][g(x^*,\xi) - Eg(x^*)]^T P(d,\xi)$. Then any sequence $\{\underset{\sim}{x}^\nu\}$ of measurable selections from the solution sets to (1.1) satisfies

$$\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*) \to \underset{\sim}{u}, \tag{4.4}$$

where $\underset{\sim}{u}$ is any selection from $\underset{\sim}{F}$.

Proof. Assumptions P.1–3 imply (4.2) as already noted. A simple application of the Cramer-Wald argument shows that $\underset{\sim}{w}(x^*)$ is distributed as a normal $N(0,\Sigma)$ random variable. Assumptions A.1–2 and P.2 allow us to apply Theorem 2.2 to the mapping J defined by (1.2), via Propositions 3.1 and 3.2, since $\underset{\sim}{F}$ a.s. single-valued implies $J^+(\underset{\sim}{w})$ a.s. single-valued. But by Remark 2.3, any selection of $\underset{\sim}{F}$ will satisfy (2.13) and the proof is complete. □

Corollary 4.2. *If, in addition to P.1–3 and A.1–2, the function Eg is directionally differentiable in the Hadamard sense at x^* and the mapping N is proto-differentiable at $(x^*, -Eg(x^*))$, then the conclusion (4.4) holds with $\underset{\sim}{u}$ equal to the unique selection from $\underset{\sim}{G}$, where*

$$\underset{\sim}{G} = \{u \in \mathrm{I\!R}^n \,|\, 0 \in \underset{\sim}{c} + (Eg)'(x^*;u) + N'_{x^*,-Eg(x^*)}(u)\}. \tag{4.5}$$

Proof. Follows immediately from Proposition 3.3. □

We present next an example of the application of this result to constrained estimation problems. Discussions, comparisons with other results, and extensions are presented in the series of remarks following the example.

Example 4.3. Let us suppose that a minimizing solution is required for the problem

$$\begin{aligned} &\text{minimize} \quad \int_{\Xi} f(x,\xi)P(d\xi) \quad \text{over all } x \in \mathbb{R}^n \\ &\text{subject to} \quad x \in X, \end{aligned} \tag{4.6}$$

where for all $x \in \Xi$, $f(\cdot,\xi) : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuously differentiable, and where the constraint set X is a closed convex polyhedral subset of $\mathbb{R}^n$. In what follows we shall use the notations $N_C(x)$ and $T_C(x)$ for the normal and tangent cones to a given convex set C at a point $x \in C$, in the sense of convex analysis [**18**]. Suppose that the gradient mapping $\nabla f : \Xi \times \mathbb{R}^n \to \mathbb{R}^n$ satisfies the probabilistic assumptions P.1–3. It follows that all solutions to (4.6) must satisfy the first-order necessary conditions

$$0 \in \int_{\Xi} \nabla f(x,\xi)P(d\xi) + N_X(x). \tag{4.7}$$

In many applications the distribution P is not known, or it is very difficult to compute with, and a closed form representation of the objective or its gradient is unobtainable for all practical purposes. However, if a sample $\{\xi_i\}$ of independent observations with common distribution P is available then the solutions may possibly be approximated by a solution sequence $\{\underset{\sim}{x}^\nu\}$, each element of which solves the computationally more tractable problem

$$0 \in \frac{1}{\nu}\sum_{i=1}^{\nu} \nabla f(x,\xi_i) + N_X(x). \tag{4.8}$$

The approximation of solutions to (4.7) by solutions $\underset{\sim}{x}^\nu$ to (4.8) is an issue that lies within the scope of Corollary 4.2. Our foremost task is to compute the proto-derivative of N_X. With the aid of Rockafellar [**21**; 5.6], we find that for a pair $x \in X$ and $y \in N_X(x)$ we have

$$(N_X)'_{x,y}(u) = N_{X'(x,y)}(u), \tag{4.9}$$

where the mapping on the right is the normal cone to the set

$$X'(x,y) = \{u \in T_X(x) \,|\, y{\cdot}u = 0\}. \tag{4.10}$$

Set $\varphi^* = \int \nabla f(x^*,\xi)P(d\xi)$ and $\Phi^* = \int \nabla^2 f(x^*,\xi)P(d\xi)$. Corollary 4.2 states: if at the solution x^* to (4.2) one has

$$(x-x')\cdot\Phi^*(x-x') > 0, \quad \forall x, x' \in X, \tag{4.11}$$

then the solutions $\underset{\sim}{x}^\nu$ to (4.8) satisfy the asymptotic formula

$$\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*) \xrightarrow[\mathcal{D}]{} \underset{\sim}{u}, \tag{4.12}$$

where $\underset{\sim}{u}$ is the (random) solution to the *random quadratic program*

$$\begin{array}{lll} \text{minimize} & \frac{1}{2}u\cdot\Phi^* u + \underset{\sim}{c}\cdot u & \text{over all } u \text{ in } \mathbb{R}^n \\ \text{subject to} & u \in T_X(x^*) & \\ & u\cdot\varphi^* = 0 & \end{array} \tag{4.13}$$

and where the random linear perturbation $\underset{\sim}{c}$ is distributed as a normal, $N(0,\Sigma^*)$, $\mathbb{R}^n$-valued random vector with covariance $\Sigma^* = \int([\nabla f(x^*,\xi) - \varphi^*][\nabla f(x^*,\xi) - \varphi^*]^T)P(d\xi)$. As a further aid to the interpretation of the result, we offer the observations that any closed convex polyhedral subset can be expressed in the form

$$X = \{x \in \mathbb{R}^n \mid Ax \le b\}$$

for some matrix $A \in \mathbb{R}^{m\times n}$ and vector $b \in \mathbb{R}^m$, and that the tangent cone to such a set at $x^* \in X$ is given by

$$T_X(x^*) = \{u \in \mathbb{R}^n \mid A_i u \le 0, \ \forall i \text{ with } A_i x^* = b_i\},$$

where $A_i = i^{th}$ row of A. Thus (4.13), for fixed linear term c, is a convex quadratic program with linear constraints.

Remark 4.4. It is not necessary to suppose that $f(\cdot,\xi)$ is twice continuously differentiable in Example 4.3., only that the gradient mapping $E\nabla f(\cdot) := \int \nabla f(\cdot,\xi)P(d\xi)$ be Hadamard differentiable at x^*, as in (3.6), and strongly monotone near x^*. Examples with only directionally differentiable gradient mappings arise in stochastic linear-quadratic programming [**11**]. Furthermore, the maximal monotone operator can be taken to be the subgradient mapping of a convex function and the proto-derivative formulas worked out from the general results in Rockafellar [**22**]; thus, in particular, the set X could be a general closed convex set (provided some regularity conditions are satisfied at x^*).

Remark 4.5. If we suppose $E\nabla f(\cdot)$ is differentiable and the Hessian $H = \nabla(E\nabla f(\cdot))(x^*)$ is positive definite then Corollary 4.2 resembles standard results in maximum likelihood estimation, except in that we allow constraints to be placed on the estimators. In particular, there are interesting parallels to be drawn between our result and those of Huber [**9**] in the unconstrained situation. Our probabilistic assumptions P.1–3 correspond roughly to Huber's assumptions N1, N3(ii) and (iii), and N4, and our monotonicity assumptions correspond practically to Huber's N2 and N3(i), and they imply his condition that $\underset{\sim}{x}^\nu \to x^*$ a.s. Huber's goal is to prove that $\sqrt{\nu}(E\nabla f(x^\nu) - E\nabla f(x^*))$ has the same asymptotic distribution as $\sqrt{\nu}(E^\nu\nabla f(x^*) - E\nabla f(x^*))$, and then to derive the asymptotic distribution of $\sqrt{\nu}(x^\nu - x^*)$ via the classical delta method under the assumption that $E\nabla f(\cdot)$ is Frechét differentiable at x^* with invertible Jacobian H. We achieve the same result, namely that $\sqrt{\nu}(\underset{\sim}{x}^\nu - x^*)$ is asymptotically $N(0, (H^{-1})^T\Sigma H^{-1})$, but under our slightly different assumptions. For a further discussion of asymptotic theory in stochastic programming from Huber's perspective, see Dupačová and Wets [**8**].

Appendix

In this appendix we briefly discuss central limit theory for random variables in $C(D : \mathbb{R}^n)$, the space of continuous $\mathbb{R}^n$-valued functions on a compact subset $D \subset \mathbb{R}^n$. Further details may be found in Araujo and Giné [2], on which this presentation has been based.

For now, let Z be a separable Banach space equiped with its Borel sets $\mathcal{Z}$, and let Z^* be the dual space of continuous linear functionals on Z. If $\underset{\sim}{z}$ is a random variable taking values in Z, we say that $\underset{\sim}{z}$ is (Pettis) *integrable* if there is an element $E\,\underset{\sim}{z} \in Z$ for which $\ell(E\,\underset{\sim}{z}) = E\{\ell(\underset{\sim}{z})\}$ for all $\ell \in Z^*$, where $E\{\cdot\}$ denotes ordinary expected value. (Clearly, if $Z = C(D : \mathbb{R}^n)$ then $E\,\underset{\sim}{z}$ exists if and only if $(E\,\underset{\sim}{z})(x) = E\{\underset{\sim}{z}(x)\}$ for every $x \in D$.) The *covariance* of $\underset{\sim}{z}$, denoted cov $\underset{\sim}{z}$ is defined to be the mapping from $Z^* \times Z^*$ into $\mathbb{R}$ given by

$$(\operatorname{cov} \underset{\sim}{z})(\ell_1, \ell_2) = E\{(\ell_1(\underset{\sim}{z}) - \ell_1(E\,\underset{\sim}{z})][\ell_2(\underset{\sim}{z}) - \ell_2(E\,\underset{\sim}{z})]\}.$$

A random variable $\underset{\sim}{\mathbf{z}}$ taking values in Z will be called *Gaussian* with mean $E\,\underset{\sim}{\mathbf{z}}$ and covariance cov $\underset{\sim}{\mathbf{z}}$ provided that for all $\ell \in Z^*$ the real-valued random variable $\ell(\underset{\sim}{\mathbf{z}})$ is normal $N(\ell(E\,\underset{\sim}{\mathbf{z}}), \operatorname{cov} \ell(\underset{\sim}{\mathbf{z}}))$.

Let us now return to the specific case at hand, that of the Banach space $C(D : \mathbb{R}^n)$. The first assertion leading to (4.2) is that the functions $E^\nu g(\cdot)$ are $C(D : \mathbb{R}^n)$-valued random variables. This is a consequence of the following proposition.

Proposition A1. *Let $(S, \mathcal{S})$ be a measurable space, and let $g : D \times S \longrightarrow \mathbb{R}^n$ be continuous in the first argument, $\forall s \in S$, and measurable in the second, $\forall x \in D$. Then the mapping $s \mapsto g(\cdot, s)$ is Borel measurable as a mapping from S into $C(D : \mathbb{R}^n)$.*

Proof. It suffices to show that for every $\alpha > 0$, the set

$$\{s \mid \sup_{x \in D} |g(s, x)| \le \alpha\}$$

is a measurable subset of $\mathbb{R}^n$. This follows easily from standard results in the theory of measurable multifunctions; see, for example, Rockafellar [**19**; Theorem 2K]. □

Corollary A2. *$E^\nu g$ is a $C(D : \mathbb{R}^n)$-valued random variable for every $\nu = 1, 2, \ldots$.*

Proof. The probability space in question can be constructed in the standard way by taking a countable number of copies of $(\Xi, \mathcal{A})$, i.e. setting $S = \Xi^{(x)}$ and equipping it with the product sigma-algebra. Now write $E^\nu g(\cdot) = \frac{1}{\nu} \sum_{i=1}^{\nu} g(\cdot; \pi_i(s))$, where $\pi_i : S \longrightarrow \Xi$ is the i^{th} coordinate projection. Then each member of the sum is a $C(D : \mathbb{R}^n)$ valued random variable, since by assumption P.1 and Proposition A1 the mapping $s \mapsto g(\cdot; \pi_i(s))$ is measurable. □

The main result is a "well-known" theorem that does not seem to have been published for $\mathcal{C}(D : \mathbb{R}^n)$ with $n \geq 2$. The argument presented here was suggested by Professor R. Pyke.

Theorem A3. *Suppose that $g : D \times \Xi \to \mathbb{R}^n$ satisfies the probabilistic assumptions P.1–3. Then there exists a Gaussian random variable $\underset{\sim}{w}$ taking values in $\mathcal{C}(D : \mathbb{R}^n)$ such that*

$$\sqrt{\nu}(E^\nu g - Eg) \xrightarrow[\mathcal{D}]{} \underset{\sim}{w},$$

where for all $x \in D$, $\underset{\sim}{w}(x)$ is a normal $N(0, \Sigma(x))$ valued random variable with covariance $\Sigma(x) = \operatorname{cov}[E^1 g(x)]$.

Proof. Each $E^\nu g$ is a vector of continuous functions $(E^\nu g, \ldots, E^\nu g_n)$. The conditions of the theorem imply that for each $j = 1, \ldots, n$ there is a Gaussian random variable in $\mathcal{C}(D : \mathbb{R}^n)$ with zero mean and coveriance equal to $\operatorname{cov} E^1 g$, which we suggestively call $\underset{\sim}{w}_j$, such that

$$\sqrt{\nu}(E^\nu g_j - Eg_j) \xrightarrow[\mathcal{D}]{} \underset{\sim}{w}_j;$$

cf. Araujo and Giné [**2**; 7.17]. It follows that the finite-dimensional distributions of $\underset{\sim}{w}^\nu : \sqrt{\nu}(Eg^\nu - Eg)$ converge to those of $\underset{\sim}{w}$, i.e. for all finite subsets $\{x_1, \ldots, x_k\} \subset D$ one has

$$(\underset{\sim}{w}^\nu(x_1), \ldots, \underset{\sim}{w}^\nu(x_k)) \xrightarrow[\mathcal{D}]{} (\underset{\sim}{w}(x_1), \ldots, \underset{\sim}{w}(x_k)).$$

This determines the limit $\underset{\sim}{w}$ uniquely as that in the statement of the theorem. Thus by Prohorov's Theorem (Billingsley [**5**; 6.1]) it remains only to show that the sequence $\{w^\nu\}$ is *tight* in $\mathcal{C}(D : \mathbb{R}^n)$, i.e. for each $\varepsilon > 0$ there is a compact set $A \subset \mathcal{C}(D : \mathbb{R}^n)$ such that $\Pr\{w^\nu \in A\} > 1 - \varepsilon$ for all sufficiently large ν. By adapting the argument of [**5**; 8.2] for $\mathcal{C}(D : \mathbb{R}^n)$ we find that the tightness of $\{w^\nu\}$ is equivalent to the simultaneous satisfaction of the following two conditions:

(i) There exists $x \in D$ such that for each $\eta > 0$ there is $\alpha \geq 0$ with

$$\Pr\{|\underset{\sim}{w}^\nu(x)| > \alpha\} \geq \eta, \quad \forall \nu \geq 1.$$

(ii) For each positive ε and η there exist $\delta > 0$ and an integer ν_0 such that

$$\Pr\{\sup_{(x-y)<\delta} |\underset{\sim}{w}^\nu(x) - w^\nu(y)| \geq \varepsilon\} \leq \eta, \quad \forall \nu \geq \nu_0.$$

These conditions follow easily from the tightness of the coordinate sequences $\{\underset{\sim}{w}_j^\nu\}$ for $j = 1, \ldots, n$ since

$$\Pr\{|\underset{\sim}{w}^\nu(x)| > \alpha\} \leq \sum_{j=1}^{n} \Pr\left\{|\underset{\sim}{w}_j^\nu(x)| > \frac{\alpha}{\sqrt{n}}\right\},$$

and similarly for the probability in condition (ii), and hence these can be made as small as one pleases by application of conditions (i) and (ii) to the co-ordinate sequences. Thus $\{\underset{\sim}{w}^\nu\}$ is tight, and the proof is complete. □

References.

1. J. Aitchison and S.D. Silvey, "Maximum likelihood estimation of parameters subject to restraints", *Annals of Mathematical Statistics* **29**(1948), 813–828.
2. A. Araujo and E. Giné, *The Central Limit Theorem for Real and Banach Valued Random Variables*, Wiley, 1980.
3. H. Attouch, *Variational Convergence for Functions and Operators*, Pitman, 1984.
4. J-P. Aubin, "Lipschitz behaviour of solutions to convex minimization problems", *Mathematics of Operations Research* **9**(1984) 97–102.
5. P. Billingsley, *Convergence of Probability Measures*, Wiley, 1968.
6. C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Springer-Verlay Lecture Notes in Math. No. 580., 1977.
7. B.R. Clarke, "Nonsmooth analysis and Fréchet differentiability of M-functionals", *Probab. Th. Rel. Fields* **73** (1986) 197–209.
8. J. Dupačová and R.J-B Wets, "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems", *Annals of Mathematical Statistics*(1988) (to appear)
9. P.J. Huber, "The behaviour of maximum likelihood estimates under non-standard conditions", *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics*(1967), 221–233.
10. G. Kallianpur, "Von Mises functionals and maximum likelihood estimation", *Sankhya, Ser. A* **23** (1963) 149–158.
11. A.J. King, *Asymptotic Behaviour of Solutions in Stochastic Optimization: Nonsmooth Analysis and the Derivation of Non-normal Limit Distributions*, Dissertation, University of Washington, 1986.
12. A.J. King, "Generalized delta theorems for multivalued mappings and measurable selections", Working Paper WP-88- , International Institute for Applied Systems Analysis, (also to appear in *Mathematics of Operations Research*) 1988.
13. G.J. Minty, "Monotone (nonlinear) operators in Hilbert space", *Duke Mathematics Journal*(1962), 341–346.
14. G. Salinetti and R.J-B Wets, "On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima", *Mathematics of Operations Research* **11**(1986), 385–419.
15. S.M. Robinson, "Generalized equations and their solutions, part 1: basic theory", *Mathematical Programming Study* **10**(1979), 128–141.

16. S.M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity", *Mathematical Programming Study* **30**(1987) 45–66.

17. R.T. Rockafellar, "On the maximality of sums of nonlinear maximal monotone operators", *Transactions of the American Mathematical Society* **149**(1970), 75–88.

18. R.T. Rockafellar, *Convex Analysis*, Princeton U. Press, 1970.

19. R.T. Rockafellar, "Integral functionals, normal integrands and measurable selections", in *Nonlinear Operators and the Calculus of Variations*, Lecture Notes in Math. 543, Springer-Verlag, 1976, pp. 157–207.

20. R.T. Rockafellar, "Directional differentiability of the optimal value function in a nonlinear programming problem", *Mathematical Programming Study* **21**(1984), 213–216.

21. R.T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization", *Annals of the Institute of H. Poincaré: Analyse Non Lineaire*(1988) (to appear).

22. R.T. Rockafellar, "First and second order proto-differentiability in nonlinear programming", manuscript, 1987.