IIASA International Institute for Applied Systems Analysis
www.iiasa.ac.at

Persistence and Continuity of Local Minimizers

Robinson, S.M.

IIASA Collaborative Paper
February 1984

PERSISTENCE AND CONTINUITY OF LOCAL MINIMIZERS

Stephen M. Robinson

February 1984
CP-84-5

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

PREFACE

The Adaptation and Optimization Project of the System and Decision Sciences Program is concerned with the development of tools for use in optimization problems, particularly those involving uncertainties.

In this paper, the author, from the University of Wisconsin-Madison, considers how optimization problems behave when the functions defining them are changed (e.g. by continuous deformation). He presents a very simple and general approach to the continuity analysis of the marginal function and the set of minimizers of such a problem.

This is the second of two papers written during the author's visit to IIASA during the summer of 1983.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

ABSTRACT

A fundamental question in nonlinear optimization is that of how optimization problems behave when the functions defining them are changed (e.g., by continuous deformation). Many authors have contriubted to our knowledge in this area. This paper presents a very simple and general approach to the continuity analysis of the marginal function and the set of minimizers of such a problem. Two abstract properties are identified as being crucial to good behavior of a problem, and these are then shown to ensure persistence and stability of local optimizers of general nonlinear optimization problems.

PERSISTENCE AND CONTINUITY OF LOCAL MINIMIZERS

Stephen M. Robinson

University of Wisconsin-Madison

DEDICATED TO PROFESSOR DR. KARL NICKEL ON HIS 60th BIRTHDAY

1. Introduction.

In the last 20 years many authors have contributed to the literature on stability in optimization. In brief, this area deals with the question of what happens to an optimization problem when the elements of the problem are in some way deformed. For example, if the original problem had optimal solutions, one might ask wether the perturbed problem has solutions and, if so, whether they are in some sense close to those of the original problem if the deformations are in some sense small. Of course, in general the answers to these questions are "no" and "no," so people have tried to find conditions to impose on the optimization problem so that the answers become "yes" and, frequently, so that the solutions are somehow well behaved as functions of the perturbation parameters.

One of the first to consider these questions was Berge [2], who proved theorems about the continuity of solutions and optimal values (marginal functions) of general optimization problems. Others who contributed early work in this area

---

Sponsored by the U.S. National Science Foundation under Grant No. MCS-8200632 at the University of Wisconsin-Madison. This paper was written at the International Institute for Applied Systems Analysis, Laxenburg, Austria, during the author's visit there in May and June of 1983.

included Dantzig, Folkman and Shapiro [3], Evans and Gould [5], Fiacco [6], and Robinson and Day [9] to name only a few. Also, more recently a great many papers have presented more specific results about smoothness or Lipschitz continuity of solutions in the case of problems with additional restrictions placed on the data. A comprehensive overview of much work in parametric optimization is given in the book by Bank et al. [1].

Although the results of Berge were among the earliest investigations in this area, they have not been directly used by many of the later workers. This is primarily because Berge's theorems require hypotheses that experience has shown to be very difficult to verify in situations occurring in practice. On the other hand, many investigations have shown that variants of two basic properties, which might loosely be called "constraint qualification" and "compact level sets," are basic to the analysis of stability. These properties have appeared in many different forms, usually fairly specific.

The aim of this paper is to show how a slight adaptation of Berge's approach can be made to yield abstract forms of just these two properties, and thus to provide a general framework for the analysis of stability. This framework is developed in Section 2, where it is applied to the global optimization problem originally considered by Berge. It yields a theorem similar to Berge's "Maximum Theorem" except for a certain critical difference in one assumption.

In Section 3 we apply these global optimization results to produce information about the stability of a local minimizer, such as one frequently encounters in practice. We prove there the main result of the paper, which states, roughly speaking, that if the two basic properties previously mentioned hold at a local minimizer then the set of local minimizers is persistent and stable.

In the rest of this section, we establish notations and conventions that we shall need in what follows. For simplicity, we represent an abstract parametric optimization problem by a function $f:R^m \times R^n \to [-\infty, +\infty]$, with the understanding that for each fixed $p \in R^m$ we wish to minimize over x: i.e., to compute the value at p of the function $\phi:R^m \to [-\infty, +\infty]$ defined by

$$\phi(p) := \inf_x f(p,x), \tag{1.1}$$

and this ϕ, as a function of the perturbations p, is the <u>marginal function</u> associated with f. Of course, the infimum in (1.1) might not be attained, but in any case we can define

$$X(p) := \{ x \in R^n | \ f(p,x) = \phi(p)\}, \tag{1.2}$$

with the understanding that the multifunction X might take empty values for some (or all) p. If X(p) is not empty, of course, it is precisely the set of all optimal solutions of the minimization problem with parameter p.

With this notation established, we can treat in considerable generality a wide variety of optimization problems; constraints cause no difficulty since for given p, the <u>effective domain</u> of $f(p,\cdot)$,

$$\text{dom } f(p,\cdot) := \{x \mid f(p,x) < +\infty\},$$

can be regarded as the "feasible set." Indeed, it is only for $x \in \text{dom } f(p,\cdot)$ that the infimum operation in (1.1) becomes at all interesting, and thus the use of an extended real-valued function f permits easy representation of constraints. In order to ensure that dom $f(p,\cdot)$ is nonempty we frequently require $f(p,\cdot)$ to be proper: i.e., to take $-\infty$ nowhere and not to take $+\infty$ everywhere.

The use of such a function f has been common in convex analysis, where it was introduced by Rockafellar [10]. More recently, Rockafellar and Wets [11] have begun to investigate a variety of questions about the general properties of such functions (which they call variational systems). Most of the generality in [11] will not be required here, as we shall need only basic ideas of compactness and continuity.

Given this framework for optimization, the questions we want to ask can be stated very simply: given a fixed $p_0 \in R^m$, what properties need to be imposed on f in order to ensure that ϕ and X have good continuity properties at p_0? It will turn out that in this case "good" should mean that ϕ is continuous at p_0 and X is upper semicontinuous (in the sense appropriate to multifunctions) there. The next section develops these results.

2. Stability in global optimization: a revised Maximum Theorem.

In this section we prove a general theorem about persistence and stability of global minimizers. This theorem is closely related to the Maximum Theorm of Berge, but it differs from Berge's result in that it weakens certain of the hypotheses in a way suitable for use in the local minimization analysis of Section 3.

The hypotheses needed for our global optimization theorem can be conveniently stated in terms of another multifunction closely related to f. The level-set multifunction $\Lambda_\alpha: R^m \to R^n$ is defined, for any fixed $\alpha \in R$, by

$$\Lambda_\alpha(p) := \{ x \in R^n | f(p,x) \leq \alpha \}.$$

The requirements that we shall place on Λ_α involve two key concepts for multifunctions: local boundedness and semicontinuity. A multifunction $F: R^k \to R^j$ is said to be locally bounded at a point $x_0 \in R^k$ if there exists a neighborhood N of x_0 such that the set

$$F(N) := \cup_{x \in N} F(x)$$

is bounded. This idea is closely related to that of upper semicontinuity (usc) at x_0, which is said to hold if for any open set G containing $F(x_0)$, there is some neighborhood N of x_0 such that $G \supset F(N)$. In fact, if $F(x_0)$ is compact then F is locally bounded at x_0 if it is upper semicontinuous there, while if also F is closed (i.e., its graph $\{(x,y) | y \in F(x)\}$ is closed in $R^k \times R^j$), then the converse holds too. Finally, if $y_0 \in F(x_0)$ then F is said to be lower semicontinuous at (x_0,y_0) if for any neighborhood M of y_0 there is a neighborhood N of x_0 such that if $x \in N$ then F(x) meets M. This property of lower semicontinuity at a point has a dual form involving the inverse multifunction F^{-1} defined by $F^{-1}(y) := \{ x | y \in F(x) \}$. We can restate the definition of lower semicontinuity by saying that F is lsc at (x_0,y_0) if and only if for each neighborhood M of y_0, $F^{-1}(M)$ is a neighborhood of x_0, and when this holds we say F^{-1} is open at (y_0,x_0).

Incidentally, another form of lower semicontinuity, widely used in the literature, defines F to be lsc at x_0 if it is lsc at (x_0,y) for each y in $F(x_0)$. This is, for example, the form used by Berge [2]. However, it has been found by experience that this stronger form of lower semicontinuity is difficult or impossible to verify in actual practice, whereas the form given here, which seems to have been introduced by Dolecki [4], can be verified in many common situations.

The theorem we shall establish here rests on three principal assumptions, aside from the fundamental one of lower semicontinuity of f. The first says that we are looking at a point p_0 at which the function $f(p_0,\cdot)$ is bounded below in the second variable: that is, the marginal function does not take the value $-\infty$ at p_0. The second assumption says that if γ is a real number greater than $\phi(p_0)$ than for all p close to p_0 there will exist some x with $f(p,x) \leq \gamma$. This hypothesis, although stated in terms of the level set Λ_γ , is actually equivalent to the upper semicontinuity of ϕ at p_0. As we shall see later, in practice one typically obtains this property by means of a constraint qualification. Finally, the third assumption that we make is that for some α greater than $\phi(p_0)$ the level sets $\Lambda_\alpha(p)$ are uniformly bounded on some neighborhood of p_0. This assumption is quite strong, but a little later we shall see how it can be finessed in the case of local minima, which is the case one encounters most frequently in actual practice.

THEOREM 2.1: Let f be a lower semicontinuous, extended real valued function on $R^m \times R^n$ such that, for some $p_0 \varepsilon R^m$,

(i) $\phi(p_0) > -\infty$, and

(ii) For each $\gamma > \phi(p_0)$, $p_0 \varepsilon$ int dom Λ_γ, and

(iii) For some $\alpha > \phi(p_0)$, Λ_α is locally bounded at p_0.

Then there is a neighborhood N of p_0 such that for each $p \in N$, $f(p,.)$ is proper, $\phi(p)$ is finite and $X(p)$ is a nonempty compact set. Further, ϕ is continuous at p_0 and X is upper semicontinuous there.

PROOF: Assumptions (i) and (iii) together imply that $\phi(p_0)$ is finite. Condition (ii) implies that, for the γ given, there is a neighborhood of p_0 on which $\phi(p)$ does not exceed γ. But since this is to hold for each $\gamma > \phi(p_0)$. it follows that $\phi(p_0) \geqq \limsup_{p \to p_0} \phi(p)$, which means that ϕ is upper semicontinuous at p_0.

Using (iii) together with (ii) (for $\gamma = \alpha$), we can find a compact set $K \subset R^n$ and a neighborhood U_1 of p_0 in R^m such that if $p \varepsilon U_1$ then $\phi \neq \Lambda_\alpha(p) \subset K$. Choose any real number $\gamma < \phi(p_0)$ and note that $f(p_0,x) > \gamma$ for all x and, in particular, for all $x \varepsilon K$. Since f is lower semicontinuous and K is compact, there is a neighborhood N of p_0 with $N \subset U_1$, such that $f(p,x) > \gamma$ whenever $p \varepsilon N$ and $x \varepsilon K$. But for any $p \varepsilon N$, if $x \notin K$ then $f(p,x) > \alpha > \gamma$, so in fact $f(p,\cdot)$ remains everywhere strictly greater than γ. But then $f(p,\cdot)$ is proper and $\phi(p)$ is finite. Further, if we select any $\beta < \phi(p_0)$ and any β' strictly between β and $\phi(p_0)$, then an argument like the one just made will show that for p near

p_o, $f(p,\cdot)$ remains everywhere greater than β', and therefore that $\phi(p)$ must be greater than β. It follows that ϕ is lower semicontinuous at p_o, hence continuous there (since we have already shown it to be upper semicontinuous at p_o).

If $p \in U_1$ then we know $\Lambda_\alpha(p)$ is nonempty; however, this set is also closed by lower semicontinuity of $f(p,\cdot)$, and it is contained in the compact set K and hence is itself compact. Hence $X(p)$ is nonempty; it is compact because it is a level set of $f(p,\cdot)$.

At this point we have only to show that X is upper semicontinuous at p_o. To do so, let G be any open set containing $X(p_o)$. For $p \in N$ and $x \in R^n$ define $g(p,x)$ to be $f(p,x) - \phi(p)$. This definition makes sense because ϕ is finite on N. The function g is lower semicontinuous on $\{ p_o \} \times R^n$ since we have already shown that ϕ is continuous at p_o. Further, on the compact set $K\setminus G$ we have $g(p_o,\cdot)$ strictly positive. Therefore, we can find a neighborhood U_2 of p_o, with $U_2 \subset N$, such that if $p \in U_2$ and $x \in K\setminus G$ then $g(p,x) > 0$. But then x cannot be in $X(p)$, and since we already know $X(p) \subset K$ we must have $X(p) \subset G$. Thus X is upper semicontinuous at p_o, and this completes the proof.

The conclusions of Theorem 2.1 apply to global minimization. Although they lead to strong conclusions about the behavior of the set of global minimizers, they depend on strong assumptions about the problem, some of which are unlikely to be easily verifiable in practice. Further, in practice one is often more concerned with <u>local</u> minimization, and with the persistence and good behavior of local minimizers. Therefore, in Section 3 we shall adapt the global conclusions of Theorem 2.1 to the case of local minimization, and in the process we shall see that the hypotheses become a good deal more palatable.

3. Stability in local minimization: main result.

This section develops the main result of the paper: a set of criteria for persistence and stability of local minimizers. These criteria are based on the hypotheses of Theorem 2.1, suitably extended to cover the case of local minimization. To formulate the idea of local minimization in the generality that we need here, we introduce the following definition of a strict local minimizing set:

DEFINITION 3.1: Let g be an extended real valued function on R^n. A nonempty subset M of R^n is a strict local minimizing set for g with respect to an open set $G \supset M$, if the set of minimizers of g on cl G is M.

Note that in this definition the function g must take the same value at each point of M, and that value must be strictly less than the value assumed by g at any point of the boundary of G. If M happens to be a singleton, it is usually called a strict local minimizer of g. Of course, the set of global minimizers of g is always a strict local minimizing set (take $G = R^n$).

We shall see how to adapt Theorem 2.1 to describe the behavior of strict local minimizing sets, rather than that of global minimizing sets. Roughly speaking, we shall do this by redefining the function being minimized so that it is $+\infty$ outside cl G. However, in order to apply Theorem 2.1 in this case we need to alter its hypotheses somewhat, in particular making them more local in nature. As a first step we show how to do this with hypothesis (ii). The following lemma uses a multifunction D, defined for the function f appearing in the previous formulation by $D(p) = \{ x | f(p,x) < +\infty \}$. In the usual nonlinear programming model, the set D(p) is the set of points feasible for the minimization problem with parameter p. Note that in the hypothesis of the lemma, we make the assumption that f is upper semicontinuous relative to graph D; this is done because, in general, no function like f could be expected to be upper semicontinuous relative to the entire space, since it is permitted to take values of $+\infty$. However, in the usual nonlinear programming situation, f is a relatively tractable function on graph D and is $+\infty$ elsewhere, so that our assumption of upper semicontinuity relative to graph D captures this idea of tractability.

LEMMA 3.2: Let f be an extended real valued function on $R^m \times R^n$, and for $p \in R^m$ let $D(p) = \{ x | f(p,x) < +\infty \}$. Let $p_0 \in R^m$, and suppose that for some $\alpha_0 > \phi(p_0)$ there exists x_0 with $f(p_0,x_0) < \alpha_0$ and such that

(i) D is lower semicontinuous at (p_0,x_0), and

(ii) f is upper semicontinuous at (p_0,x_0) relative to graph D.

Then $p_0 \in \operatorname{int} \operatorname{dom} \Lambda_{\alpha_0}$.

PROOF: Since f is usc at (p_0,x_0) relative to graph D, and since $f(p_0,x_0) < \alpha_0$, there exist neighborhoods U_1 of p_0 in R^m and V of x_0 in R^n, such that if $p \in U_1$, $x \in V$, and $(p,x) \in$ graph D then

$f(p,x) < \alpha_0$. Further, since D is lsc at (p_0,x_0) there exists a neighborhood U_2 of p_0, with $U_2 \subset U_1$, such that if $p \in U_2$ then $D(p) \cap V \neq \phi$. For any $p \in U_2$ there is then an $x \in D(p) \cap V$, and for this x we have $f(p,x) < \alpha_0$ and therefore $x \in \Lambda_{\alpha_0}(p)$. But then $p \in \operatorname{int} \operatorname{dom} \Lambda_{\alpha_0}$ as claimed, and this completes the proof.

The next theorem formulates the main result about stability for local minimization. In the theorem we use the notation ψ_A for the indicator function of a set A: $\psi_A(x)$ is zero if $x \in A$ and is $+\infty$ if $x \notin A$.

THEOREM 3.3: *Let* f *be a lower semicontinuous, extended real valued function on* $R^m \times R^n$. *Let* $p_0 \in R^m$, *assume* $f(p_0,\cdot)$ *is proper, and let* M *be a bounded, strict local minimizing set for* $f(p_0,\cdot)$ *with respect to the bounded open set* $G \subset R^n$. *Suppose that for some* $x_0 \in M$, f *is upper semicontinuous at* (p_0,x_0) *relative to* graph D, *and* D is *lower semicontinuous at* (p_0,x_0). *Define* $g(p,x) := f(p,x) + \psi_{\text{cl } G}(x)$, *and for each* $p \in R^m$ *let* $\eta(p) = \inf_x g(p,x)$ *and* $Y(p) = \{x \mid g(p,x) = \eta(p)\}$.

Then $Y(p_0) = M$, *and there exists a neighborhood* U of p_0 *such that if* $p \in U$ *then*

a. $g(p,\cdot)$ *is proper*, $\eta(p)$ *is finite, and* $Y(p)$ *is nonempty and compact*.

b. $Y(p)$ *is a strict local minimizing set for* $f(p,\cdot)$ *with respect to* G.

Further, η *is continuous at* p_0 *and* Y *is upper semicontinuous there*.

PROOF: $Y(p_0) = M$ by the definition of strict local minimizing set, which also ensures that $\eta(p_0) = g(p_0,x_0) = f(p_0,x_0)$. Since $f(p_0,\cdot)$ is proper, $\eta(p_0) > -\infty$ (since the infimum is attained), while if $\eta(p_0)$ were $+\infty$ then f would be constant ($+\infty$) on cl G, contradicting the hypothesis that M is a strict local minimizing set. Thus $\eta(p_0)$ is actually finite. We know that g is lsc on $R^m \times R^n$, since f was lsc there and cl G is a closed set. Next, for each $p \varepsilon R^m$ define $E(p):=$ dom $g(p,\cdot)$; for each p, $E(p) = D(p) \cap (\text{cl } G)$, so that graph $E \subset$ graph D, and thus g is usc at (p_0,x_0) relative to graph E since we assumed it was usc there relative to graph D. Also, since $x_0 \varepsilon G =$ int cl G, lower semicontinuity of D at (p_0,x_0) implies lower semicontinuity of E there. If $\gamma > \eta(p_0)$ then $g(p_0,x_0) > \gamma$; applying Lemma 3.2 to g we conclude that $p_0 \varepsilon$ int dom K_γ, where K_γ is the level-set multifunction associated with g:

$$K_\gamma(p):=\{x \mid g(p,x) \leq \gamma\} = \Lambda_\gamma(p) \cap (\text{cl } G).$$

Finally, since $K_\gamma(p) \subset$ cl G for any p, the multifunction K_γ is locally bounded at p_0.

Applying Theorem 2.1 to g, we conclude that for some neighborhood U of p_0 and for each $p \varepsilon U$, $g(p,\cdot)$ is proper, $\eta(p)$ is finite, and $Y(p)$ is a nonempty compact set. Further, η is continuous at p_0 and Y is upper semicontinuous there. But $Y(p_0)$ is contained in the open set G, so if we choose U to be small enough then $Y(p) \subset G$ for each $p \varepsilon U$, and thus for such p $Y(p)$ is actually a strict local minimizing set for $f(p,\cdot)$ with respect to G. This completes the proof of Theorem 3.3.

There are two essential assumptions in Theorem 3.3: that the local minimizing set M is bounded, and that D is lower semicontinuous at (p_0,x_0). In nonlinear programming problems encountered in practice, the first condition is typically satisfied by assuming that the second-order sufficient optimality condition holds at the point in question, although this assumption is actually stronger than is needed for Theorem 3.3.

The second condition (lower semicontinuity of D at (p_0,x_0)) is generally met by assuming that one of the standard constraint qualifications holds at x_0 for the problem with $p = p_0$. For example, the Mangasarian-Fromovitz constraint qualification [7], suitably generalized, is appropriate for this purpose. For details on the use of the second-order sufficient condition and the generalized Mangasarian-Fromovitz condition in nonlinear programming, see [8].

The result of Theorem 3.3 gives a complete and general criterion for stability in the sense of upper semicontinuity of the minimizing set. Sometimes one wants more than this: in some applications it may be necessary to establish bounds on the rate at which the set of local optimizers can vary. Such results for nonlinear programming problems are treated in [8].

REFERENCES

[1] B. Bank et al., Non-Linear Parametric Optimization (Akademie-Verlag, Berlin, DDR, 1982).

[2] C. Berge, Topological Spaces (Macmillan, New York, 1963).

[3] G.B. Dantzig, J. Folkman, and N. Shapiro, "On the Continuity of the Minimum Set of a Continuous Function," J. Math. Anal. Appl. 17 (1967) 519-548.

[4] S. Dolecki, "Semicontinuity in Constrained Optimization, Part 1: Metric Spaces," Control Cybernet. 7 (1978) 2, 5-16.

[5] J.P. Evans and F.J. Gould, "Stability in Nonlinear Programming," Operations Res. 18 (1970) 107-118.

[6] A.V. Fiacco, "Convergence Properties of Local Solutions of Sequences of Nonlinear Programming Problems in General Spaces," J. Optimization Theory Appl. 13 (1974) 1-12.

[7] O.L. Mangasarian and S. Fromovitz, "The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints," J. Math. Anal. Appl. 17 (1967) 37-47.

[8] S.M. Robinson, "Generalized Equations and Their Solutions, Part II: Applications to Nonlinear Programming," Math. Programming Studies 19 (1982) 200-221.

[9] S.M. Robinson and R.H. Day, "A Sufficient Condition for Continuity of Optimal Sets in Mathematical Programming," J. Math. Anal. Appl. 45 (1974) 506-511.

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

[11] R.T. Rockafellar and R.J.-B. Wets, Extended Real Analysis (Manuscript, in preparation).