COLLOQUIUM MATHEMATICUM

VOL. LX/LXI DÉDIÉ À M. ANTONI ZYGMUND 1990

THE PLENARY HULL OF THE GENERALIZED JACOBIAN MATRIX AND THE INVERSE FUNCTION THEOREM IN SUBDIFFERENTIAL CALCULUS

BY

TELMA CAPUTTI (BUENOS AIRES)

Introduction. Some of the most important recent advances in optimization have come about as a result of systematic replacement of smoothness assumptions by convexity. This is exemplified by the work of Rockafellar [6]. It is natural to ask whether analogous results can be proven without either smoothness or convexity. A general theory of necessary conditions for such problems has been obtained [1]. The conditions are expressed, in part, by means of generalized gradients.

The classical inverse function theorem gives conditions under which a C^r function admits (locally) a C^r inverse. The purpose of this paper is to give conditions under which a Lipschitzian (not necessarily differentiable) function admits (locally) a Lipschitzian inverse by means of the characterization of the plenary hull of the generalized Jacobian matrix.

1. Locally Lipschitz functions. Let $f : B \to \mathbf{R}$ be locally Lipschitz on a bounded subset B of $\mathbf{R}^n$. It is known [7] that such a function has at almost all points x a derivative (gradient), which we denote by $\nabla f(x)$. It is easily verified that the function ∇f is bounded on bounded subsets of its domain of definition.

Let now $F : O \to \mathbf{R}^m$ be locally Lipschitz, O a nonempty open subset of $\mathbf{R}^n$. One is tempted to define the generalized derivative of $F = (f_1, \ldots, f_m)^t$ at $x_0 \in O$ by simply considering $[\partial f_1(x_0), \ldots, \partial f_m(x_0)]^t$ (for undefined concepts the reader is referred to [1]).

The usual $m \times n$ Jacobian matrix of partial derivatives, when it exists, is denoted by $JF(x)$. We topologize the vector space of $m \times n$ matrices with the norm

$$\|M\| = \max |m_{ij}| \quad \text{where } M = (m_{ij}) \ , \ 1 \leq i \leq m \ , \ 1 \leq j \leq n \, .$$

A mathematical tool is what F. H. Clarke called the generalized Jacobian matrix, defined in the following way:

DEFINITION 1.1. The generalized Jacobian matrix of F at $x_0 \in O$, denoted by $\tilde{J}F(x_0)$, is the *convex hull* of all matrices M of the form $M = \lim_{n\to\infty} JF(x_n)$ where x_n converges to x_0 in $\operatorname{dom} F'$.

In this definition, $\operatorname{dom} F'$ denotes the subset of full measure of O where F is differentiable.

$\tilde{J}F(x_0)$ is a nonempty compact convex subset of the vector space of $m \times n$ matrices, which reduces to $\{JF(x_0)\}$ whenever F is C^1 in some neighborhood of x_0.

DEFINITION 1.2. $\tilde{J}F(x_0)$ is said to be of *maximal rank* if every M in $\tilde{J}F(x_0)$ is of maximal rank.

2. Plenary hull of the generalized Jacobian matrix.

Let us denote by $\langle\langle\ ,\ \rangle\rangle$ the inner product on the vector space of $m \times n$ matrices defined by $\langle\langle M, U\rangle\rangle =$ Trace of $M \circ U^t$; it follows from Definition 1.1 that for all $U \in \mathbf{R}^{m\times n}$

$$\max_{M\in\tilde{J}F(x_0)} \langle\langle M, U\rangle\rangle = \limsup_{\substack{x\to x_0\\ x\in\operatorname{dom} F'}} \langle\langle JF(x), U\rangle\rangle\,. \tag{2.1}$$

Consider $U \in \mathbf{R}^{m\times n}$ of the form $u \otimes v : x \to \langle u, x\rangle v$, where $u \in \mathbf{R}^n$ and $v \in \mathbf{R}^m$. Then $\langle\langle M, u\otimes v\rangle\rangle$ reduces to $\langle Mu, v\rangle$ and (2.1) can be rephrased as

$$\max_{M\in\tilde{J}F(x_0)} \langle Mu, v\rangle = \limsup_{\substack{x\to x_0\\ x\in\operatorname{dom} F'}} \langle JF(x)u, v\rangle\,. \tag{2.2}$$

We can use results on chain rules so that the left-hand of (2.2) appears as the generalized gradient of a particular real-valued function. Given $v \in \mathbf{R}^m$, the *generalized gradient* of $F_v : x \to \langle F(x), v\rangle$ at x_0 can be exactly described as $\partial F_v(x_0) = \tilde{J}^t F(x_0)v$ ([3]). Therefore, for all $u \in \mathbf{R}^n$, we have

$$\max_{M\in\tilde{J}F(x_0)} \langle u, M^t v\rangle = F_v^\circ(x_0; u) \quad \text{(see [1, Definition 1.3])}. \tag{2.3}$$

Although $\tilde{J}F(x_0)$ is convex and compact, one generally cannot separate an M_0 from $\tilde{J}F(x_0)$ by using only linear mappings (in $\mathbf{R}^{m\times n}$) of the form $u \otimes v$, $u \in \mathbf{R}^n$, $v \in \mathbf{R}^m$. This led Sweetser [8] to introduce the following definition:

DEFINITION 2.1. A subset $A \subset \mathbf{R}^{m\times n}$ is *plenary* if it includes every M in $\mathbf{R}^{m\times n}$ satisfying $Mu \in Au$ for all $u \in \mathbf{R}^n$.

Since the intersection of plenary sets is plenary, Sweetser defined the *plenary hull of* A, denoted $\operatorname{plen} A$, as the smallest plenary set containing A. When $\min(m,n) > 1$, $\operatorname{plen} \tilde{J}F(x_0)$ is a convex compact (plenary) set of matrices containing $\tilde{J}F(x_0)$.

Since $\tilde{J}F(x_0)u = [\text{plen}\,\tilde{J}F(x_0)]u$ for all $u \in \mathbf{R}^n$, Hiriart-Urruty and Thibault [4] formulated the following theorem:

THEOREM 2.1. *Let $u \in \mathbf{R}^n$ and $v \in \mathbf{R}^m$. Then*

$$\max_{M \in \text{plen}\,\tilde{J}F(x_0)} \langle Mu, v\rangle = F^\circ(x_0; u, v). \tag{2.4}$$

In other words, $M \in \text{plen}\,\tilde{J}F(x_0)$ if and only if

$$\langle Mu, v\rangle \leq F^\circ(x_0; u, v) \quad \text{for all } (u,v) \in \mathbf{R}^n \times \mathbf{R}^m .$$

To summarize, let us say that $\text{plen}\,\tilde{J}F(x_0)$ is the convex compact (plenary) set of matrices satisfying $[\text{plen}\,\tilde{J}F(x_0)]u = \tilde{J}F(x_0)u$ for all $u \in \mathbf{R}^n$. When $F = (f_1, \ldots, f_m)^t$ we have

$$\tilde{J}F(x_0) \subset \text{plen}\,\tilde{J}F(x_0) \subset [\partial f_1(x_0), \ldots, \partial f_m(x_0)]^t .$$

The set $[\partial f_1(x_0), \ldots, \partial f_m(x_0)]^t$ is obviously convex, compact and plenary. It actually yields the same image set as $\tilde{J}F(x_0)$ does when the considered vectors u are the elements e_i of the canonical basis in $\mathbf{R}^n$. In other words,

$$\{x_i^*, [x_1^*, \ldots, x_i^*, \ldots, x_m^*]^t \in \tilde{J}F(x_0)\} = \partial f_i(x_0) \quad ([3]) .$$

3. The plenary hull of $\tilde{J}F(x_0)$ and the inverse function theorem.

THEOREM 3.1. *Let $F : O \to \mathbf{R}^n$, $O \subset \mathbf{R}^n$. If every matrix M in $\text{plen}\,\tilde{J}F(x_0)$ is of maximal rank, then there exist neighborhoods U and V of x_0 and $F(x_0)$ respectively, and a Lipschitzian function $G : V \to \mathbf{R}^n$ such that for all $(u,v) \in U \times V$, $F(u) = v$ if and only if $G(v) = u$.*

When F is C^1, $\tilde{J}F(x_0)$ reduces to $JF(x_0)$ and the function G above is necessarily C^1 as well. Thus we recover the classical theorem.

Remark 1. This theorem remains true (without modifications in the proof) if we impose the maximality of rank for all $M \in \tilde{J}_\Lambda F(x_0)$ where $\Lambda \subset \text{dom}\,F'$ has complement in O of null measure and $\tilde{J}_\Lambda F(x_0)$ is defined as in (1.1) except that the points x_n belong to Λ only.

Remark 2. By the very definitions,

$$\partial_\Lambda F_v(x_0) = \tilde{J}_\Lambda^t F(x_0)v \quad ([3]) .$$

It is known that the generalized gradient of a real-valued function does not change when we alter the values of the function on a set of null measure [1, Proposition 1.11]. The desire to make the generalized derivative insensitive to sets of null measure led B. H. Pourciau [5] to alter Clarke's original definition by considering the Lebesgue set $\text{Leb}\,F'$ of F' instead of $\text{dom}\,F'$ in the definition of $\tilde{J}F(x_0)$, but, since F' is locally in $L^\infty(O, \mathbf{R}^m)$, almost every x in $\text{dom}\,F'$ belongs to $\text{Leb}\,F'$.

Remark 3. Let Λ and $\widetilde{J}_\Lambda F(x_0)$ be as in Remark 1. Then plen $\widetilde{J}_\Lambda F(x_0)$ = plen $\widetilde{J}F(x_0)$. So, plen $\widetilde{J}F(x_0)$ is insensitive to sets of measure zero.

Proof of Theorem 3.1.

LEMMA 1 (An exact chain rule in the finite-dimensional case [5].) *Let $F : \mathbf{R}^n \to \mathbf{R}^m$ be a locally Lipschitz function and let $g : \mathbf{R}^m \to \mathbf{R}$ be continuously differentiable. Then*

$$\partial(g \circ F) = \widetilde{J}^t F(x_0)\nabla g(F(x_0))\,.$$

LEMMA 2. *Let β be a positive number. Then for all x sufficiently near x_0,* plen $\widetilde{J}F(x) \subset$ [plen $\widetilde{J}F(x_0) + \beta M(0,1)$] *where $M(0,1)$ denotes the unit ball in the vector space of $m \times n$ matrices.*

This is a direct consequence of the definition of the plenary hull of the generalized Jacobian matrix.

LEMMA 3. *There are positive numbers r and λ with the following property: given any unit vector v in $\mathbf{R}^n$, there is a unit vector u in $\mathbf{R}^n$ such that, whenever x lies in $x_0 + rB$ and $M \in$* plen $\widetilde{J}F(x_0)$, *then $\langle Mv, u\rangle \geq \lambda$ for all M, where B denotes the open unit ball in $\mathbf{R}^n$.*

Proof. Let Σ_1 denote the unit sphere in $\mathbf{R}^n$. Then the subset (plen $\widetilde{J}F(x_0))\Sigma_1$ of $\mathbf{R}^n$ is compact and does not contain 0 since plen $\widetilde{J}F(x_0)$ is of maximal rank.

Hence for some $\lambda > 0$, (plen $\widetilde{J}F(x_0))\Sigma_1$ is distant at least 2λ from 0. For positive β sufficiently small, [plen $\widetilde{J}F(x_0) + \beta M(0,1)]\Sigma_1$ is distant at least λ from 0.

By Lemma 2, it follows that for some positive r,

$$x \in x_0 + rB \Rightarrow \text{plen}\,\widetilde{J}F(x) \subset \text{plen}\,\widetilde{J}F(x_0) + \beta M(0,1)\,.$$

We may suppose r chosen so that F satisfies the Lipschitz condition on $x_0 + r\overline{B}$.

Now let a unit vector v be given. It follows from the above that the convex set [plen $\widetilde{J}F(x_0) + \beta M(0,1)]v = [\widetilde{J}F(x_0) + \beta M(0,1)]v$, for all v in $\mathbf{R}^n$, is distant at least λ from 0. By the usual separation theorem for convex sets, there is a unit vector u such that $\langle u, Mv\rangle \geq \lambda$ for all $M \in$ plen $\widetilde{J}F(x)$.

LEMMA 4. *If x_1 and x_2 lie in $x_0 + r\overline{B}$, then*

$$\|F(x_1) - F(x_2)\| \geq \lambda\|x_1 - x_2\|\,.$$

Proof. We may suppose $x_1 \neq x_2$ and by the continuity of F that $x_1, x_2 \in x_0 + rB$.

Set $v = (x_2 - x_1)/\|x_2 - x_1\|$, $\alpha = \|x_2 - x_1\|$ so that $x_2 = x_1 + \alpha v$. Let π be the plane perpendicular to v and passing through x_1. The set P of points x in $x_0 + rB$ where F' fails to exist is of measure zero, and hence by Fubini's theorem, for almost every x in π the ray $x + tv$, $t \geq 0$, meets P in a set of null one-dimensional measure. Choose an x with the above property and sufficiently close to x_1 so that $x + tv$ lies in $x_0 + rB$ for every t in $[0, \alpha]$. Then the function $t \to F(x + tv)$ is Lipschitzian for t in $[0, \alpha]$ and has a.e. on this interval the derivative $JF(x + tv)v$. Thus

$$F(x + \alpha v) - F(x) = \int_0^\alpha JF(x + tv)v \, dt.$$

Let u be as in Lemma 3. We deduce that

$$\langle u, F(x + \alpha v) - F(x) \rangle = \left\langle u, \int_0^\alpha JF(x + tv)v \, dt \right\rangle \geq \int_0^\alpha \lambda \, dt = \lambda\alpha.$$

Recalling the definition of α, we arrive at

$$\|F(x + \alpha v) - F(x)\| \geq \lambda \|x_2 - x_1\|.$$

This may be done for x arbitrarily close to x_1. Since F is continuous, the lemma follows.

LEMMA 5. *$F(x_0 + rB)$ contains $F(x_0) + (r\lambda/2)B$.*

Proof. Let y be any point in $F(x_0) + (r\lambda/2)B$, and let the minimum of $\|y - F(x)\|^2$ over $x_0 + r\overline{B}$ be attained at x. We claim that x belongs to $x_0 + rB$. Indeed, otherwise

$$\begin{aligned} r\lambda/2 > \|y - F(x_0)\| &\geq \|F(x) - F(x_0)\| - \|y - F(x)\| \\ &\geq \lambda\|x - x_0\| - \|y - F(x)\| \\ &\geq \lambda r - \|y - F(x_0)\| > \lambda r - r\lambda/2 = r\lambda/2, \end{aligned}$$

which is a contradiction. Thus x yields a local minimum for the function $\|y - F(x)\|^2$, and consequently [1, Corollary 1.10],

$$0 \in \partial \|y - F(x)\|^2.$$

We now use Lemma 1 to conclude that 0 belongs to the set

$$\tilde{J}^t F(x)(y - F(x)) \quad ([3]),$$

which coincides with $[\text{plen}\, \tilde{J}^t F(x)](y - F(x))$ for all vectors in $\mathbf{R}^n$ by Theorem 2.1. But Lemma 3 implies that every matrix in plen $\tilde{J}F(x)$ is nonsingular, hence the above is possible only if $F(x) = y$.

We now set $V = F(x_0) + (r\lambda/2)B$, and we define G on V as follows: $G(v)$ is the unique x in $x_0 + rB$ such that $F(x) = v$. For U we can choose any neighborhood of x_0 satisfying $F(U) \supset V$. The theorem is now seen to follow, since Lemma 4 implies that G is Lipschitz with constant λ^{-1}.

REFERENCES

[1] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. 205 (1975), 247–262.

[2] J. B. Hiriart-Urruty, *Contributions à la programmation mathématique: cas détérministe et stochastique*, thèse, Université de Clermont-Ferrand II, 1977.

[3] —, *New concepts in nondifferentiable programming*, Analyse Non Convexe (Proc. Colloq. Pau 1977), Bull. Soc. Math. France Mém. 60 (1979), 57–85.

[4] J. B. Hiriart-Urruty et L. Thibault, *Existence et caractérisation de différentielles généralisées d'applications localement lipschitziennes d'un espace de Banach séparable dans un espace de Banach réflexif séparable*, C. R. Acad. Sci. Paris Sér. A-B 290 (1980), 1091–1094.

[5] B. H. Pourciau, *Analysis and optimization of Lipschitz continuous mappings*, J. Optim. Theory Appl. 22 (1977), 311–351.

[6] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, N.J., 1970.

[7] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton Univ. Press, Princeton, N.J., 1970.

[8] T. H. Sweetser, *A minimal set-valued strong derivative for vector-valued Lipschitz functions*, J. Optim. Theory Appl. 23 (1977), 549–562.

DEPARTAMENTO DE MATEMÁTICA
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD DE BUENOS AIRES
1428 BUENOS AIRES, ARGENTINA

Reçu par la Rédaction le 30.9.1989