PERGAMON

Nonlinear Analysis 49 (2002) 1111–1125

Nonlinear Analysis

www.elsevier.com/locate/na

Support functions of the Clarke generalized Jacobian and of its plenary hull

Cyril Imbert

Laboratoire de Mathématiques pour l'Industrie et la Physique Université Paul Sabatier, UFR MIG, 118 route de Narbonne 31062 Toulouse Cedex, France

Received 30 July 1999; accepted 22 January 2001

Keywords: Vector-valued functions; Generalized jacobian; Support function; Plenary hull

1. Introduction

Let $\mathcal{O}$ be an open subset of $\mathbb{R}^n$ and consider a locally Lipschitz function $f:\mathcal{O}\to\mathbb{R}$. In order to tackle the first order behaviour of such nondifferentiable functions, Clarke [1] introduced the notion of *generalized gradients*. A vector of $\mathbb{R}^n$ is a generalized gradient of f at x_0 if it is an element of

$$\partial f(x_0)=\operatorname{co}\{\lim \nabla f(x_i)\colon x_i\to x_0,\ x_i\in D_f\}, \tag{1}$$

where D_f denotes the set of all the points where f is differentiable and where $\nabla f(x_i)$ denotes the gradient of f at x_i. The set defined by (1) is referred to as the *Clarke subdifferential*. It is nonempty, compact and convex.

This object has been intensively studied, generalized and used since Clarke introduced it in 1973. Perhaps one reason for this is that he was able to calculate its support function; he found what is now known as the *generalized directional derivative*:

$$f^\circ(x_0;u)=\limsup_{x\to x_0,\varepsilon\to 0^+}\frac{f(x+\varepsilon u)-f(x)}{\varepsilon}. \tag{2}$$

Clarke naturally generalized this object to vector-valued locally Lipschitz functions.

E-mail address: imbert@mip.ups-tlse.fr (C. Imbert).

0362-546X/02/$ - see front matter
PII: S0362-546X(01)00730-1

Definition 1. Consider a locally Lipschitz function $F:\mathcal{O}\to\mathbb{R}^m$ and fix $x_0\in\mathcal{O}$. The (Clarke) generalized jacobian of F at x_0 is the following set of $m\times n$ matrices:

$$\mathscr{J}F(x_0)=\mathrm{co}\{\lim JF(x_i)\colon x_i\to x_0,\ x_i\in D_f\},$$

where $JF(x_i)$ stands for the classical jacobian matrix of F at x_i.

It is nonempty, convex and compact. We next denote the set of all the $m\times n$ matrices by $M_{m,n}(\mathbb{R})$. It is a Euclidean space when equipped with the canonical scalar product for matrices: $\langle\!\langle A,B\rangle\!\rangle=\mathrm{tr}(A^{\mathrm{T}}B)$. The canonical scalar product of $\mathbb{R}^n$ is denoted by $\langle .,.\rangle$. This mathematical object has not been thoroughly investigated and exploited. Until now, there was no analytic expression of its support function, although this is an essential tool for studying it. Our first main result fills this gap. The support function of the generalized jacobian turns out to be a "generalized directional divergence".

Theorem 1. *Let $M\in M_{m,n}(\mathbb{R})$. Consider $P_\varepsilon(x)$, the hypercube in $\mathbb{R}^n$ of vertex x, whose edges issued from x are directed by vectors of the canonical basis of $\mathbb{R}^n$:*

$$P_\varepsilon(x):=\{x+\varepsilon t_1e_1+\cdots+\varepsilon t_ne_n\colon t_i\in[0,1]\ \textit{for all}\ i\},$$

$\partial P_\varepsilon(x)$ its boundary, $n(y)$ the outer normal vector at $y\in P_\varepsilon(x)$, and σ the surface Lebesgue measure on $\partial P_\varepsilon(x)$, that is to say on the faces of the hypercube.

If $n\geqslant 2$, then the support function of $\mathscr{J}F(x_0)$ in the direction M equals:

$$\limsup_{x\to x_0,\varepsilon\to 0^+}\frac{1}{\varepsilon^n}\int_{\partial P_\varepsilon(x)}\langle f(y),Mn(y)\rangle\,\mathrm{d}\sigma(y). \tag{3}$$

If $n=1$, it equals $(\langle M,F\rangle)^\circ(x_0;1)$.

Some results of $\mathscr{J}F$ were already known. First, the generalized jacobian is more exact than (i.e. is a subset of) the cartesian product $\partial f_1(x_0)\times\cdots\times\partial f_m(x_0)$, because the possible "interdependence" of the component functions has been taken into account. Secondly, Warga et al. [11] proved that the generalized jacobian is "blind to null sets" (i.e. its definition is not modified if one specifies in (1) that $x_i\notin N_0$, if N_0 has a null Lebesgue measure). Finally, Hiriart-Urruty [5] determined the support function of the images of $\mathscr{J}F(x_0)$: $\mathscr{J}F(x_0)u$, $u\in\mathbb{R}^n$. Unfortunately, a set is not uniquely determined by its images in general. This led Halkin and Sweetser [10] to introduce the notion of a *plenary* set: $\mathscr{A}\subset M_{m,n}(\mathbb{R})$ is plenary if it contains all the matrices A verifying: $Au\in\mathscr{A}u$ for any $u\in\mathbb{R}^n$. The *plenary hull* of $\mathscr{A}$, denoted by $\mathrm{plen}\,\mathscr{A}$, is the smallest plenary set containing $\mathscr{A}$. The generalized jacobian is not necessarily plenary, unless $m=1$ or $n=1$. Hence, $\mathrm{plen}\,\mathscr{J}F(x_0)$ is a new set, that is still convex and compact. It contains $\mathscr{J}F(x_0)$ and has the same images. This object is also still more precise than $\partial f_1(x_0)\times\cdots\times\partial f_m(x_0)$. Our second goal is to calculate its support function in all directions.

Theorem 2. *Under assumptions of Theorem* 1, *the support function of the plenary hull of* $\mathscr{J}F(x_0)$ *in the direction* $M \in M_{m,n}(\mathbb{R})$ *equals*

$$\inf\left\{\sum_{i=1}^{k}(\langle v_i, F\rangle)^{\circ}(x_0; u_i):\ \sum_{i=1}^{k} u_i \otimes v_i = M\right\}, \tag{4}$$

where $u \otimes v$ *denotes the matrix* vu^{T} *for any* $u \in \mathbb{R}^n$ *and* $v \in \mathbb{R}^m$.

The contents of the present paper are organized as follows: in the first section, we give notations, definitions and results that are used throughout; Section 2 is devoted to the proof of Theorem 1 and to a technical expression of the support function of $\mathscr{J}F(x_0)$ in terms of difference quotients; in Section 3, we first prove Theorem 2 and then derive a corollary; we conclude by determining whether the infimum in (4) is attained; in Section 4, the results we previously obtained are applied to the second order differentiation theory. We conclude this paper by giving some examples and by re-examining some known results.

2. Preliminaries

This section is devoted to notations, definition and results used in the paper. We denote the support function of a subset $\mathscr{A} \subset M_{m,n}(\mathbb{R})$ in the direction $M \in M_{m,n}(\mathbb{R})$ by $\sigma_{\mathscr{A}}(M)$. We recall that

$$\sigma_{\mathscr{A}}(M) = \sup\{\langle \zeta, M\rangle:\ \zeta \in \mathscr{A}\}.$$

If $\mathbb{R}^n$ is equipped with the Lebesgue measure μ, $L^1_{\mathrm{loc}}(\mathbb{R}^n, \mathbb{R})$ denotes the set of all locally μ-integrable numerical functions defined on $\mathbb{R}^n$. The closed ball of radius r centered at x is denoted by $B(x,r)$. For a given $h \in L^1_{\mathrm{loc}}(\mathbb{R}^n, \mathbb{R})$, a point $x \in \mathbb{R}^n$ is a so-called *Lebesgue point* of h if:

$$h(x) = \lim_{r \to 0^+} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} h(y)\,\mathrm{d}\mu(y).$$

The set of the Lebesgue points of h is denoted by L_h.

Theorem 3 (Rudin [9]). *Let* $h \in L^1_{\mathrm{loc}}(\mathbb{R}^n, \mathbb{R})$. *Then* μ*-almost every point in* $\mathbb{R}^n$ *is a Lebesgue point of* h.

A sequence $\{R_i\}_i$ of Borel sets in $\mathbb{R}^n$ is said to *shrink to* x *nicely* if there is a number $\alpha > 0$ with the following property: there is a sequence of balls $B(x, r_i)$ with $\lim r_i = 0$ such that $R_i \subset B(x, r_i)$ and $\mu(R_i) \geqslant \alpha \mu(B(x, r_i))$.

Proposition 1 (Rudin [9]). *Let* $h \in L^1_{\mathrm{loc}}(\mathbb{R}^n, \mathbb{R})$ *and* $x \in L_h$. *If a sequence* $\{R_i\}_{i \geqslant 1}$ *shrinks to* x *nicely*, *then the following holds*:

$$h(x) = \lim_{i \to \infty} \frac{1}{\mu(R_i)} \int_{R_i} h(y)\,\mathrm{d}\mu(y). \tag{5}$$

Example 1. The sequence $\{P_\varepsilon(x)\}_{\varepsilon>0}$ shrinks to x nicely.

Hiriart-Urruty stated and proved the following result [5].

Proposition 2. $\sigma_{\mathscr{J}F(x_0)}(vu^{\mathrm{T}}) = \sigma_{\mathscr{J}F(x_0)u}(v) = (\langle v, F\rangle)^\circ(x_0; u)$.

The next lemma gives a characterization of the plenary hull of a convex and compact subset of $M_{m,n}(\mathbb{R})$.

Lemma 1 (Hiriart-Urruty, Thibault [7]). *Let $\mathscr{A}$ be a convex and compact subset of $M_{m,n}(\mathbb{R})$. Then $A \in \operatorname{plen} \mathscr{A}$ is plenary if and only if, for any $u \in \mathbb{R}^n$ and any $v \in \mathbb{R}^m$:*

$$\langle\langle A, vu^{\mathrm{T}}\rangle\rangle \leqslant \sigma_{\mathscr{A}}(vu^{\mathrm{T}}).$$

The matrix vu^{T} represents the linear mapping, denoted by $u \otimes v$, that assigns to any $x \in \mathbb{R}^N$ the vector $\langle u, x\rangle v \in \mathbb{R}^m$. Throughout, we identify the linear mapping and its representative matrix. If $u \neq 0$ and $v \neq 0$, $u \otimes v$ is of rank 1. Conversely, any rank-1 matrix can be represented by $u \otimes v$ for some u, v. Moreover for any $M \in M_{m,n}(\mathbb{R})$:

$$\langle\langle M, u \otimes v\rangle\rangle = \langle Mu, v\rangle.$$

Thus, in Proposition 2, the support function of $\mathscr{J}F(x_0)$ is calculated in the directions of the rank-1 matrices. Combining it with Lemma 1, we obtain:

Lemma 2. *A matrix $\zeta \in M_{m,n}(\mathbb{R})$ is an element of $\operatorname{plen} \mathscr{J}F(x_0)$ if and only if, for any $u \in \mathbb{R}^n$ and any $v \in \mathbb{R}^m$:*

$$\langle \zeta u, v\rangle \leqslant (\langle v, F\rangle)^\circ(x_0; u).$$

3. The support function of the generalized jacobian

3.1. Proof of Theorem 1

Proof. By setting $G = M^{\mathrm{T}}F$, the problem is reduced to the case $m = n$ and $M = \mathrm{Id}$ where Id is the identity matrix of $M_n(\mathbb{R})$. The key part of the proof of Theorem 1 is the following claim:

Claim 1.

$$\sigma_{\mathscr{J}G(x_0)}(\mathrm{Id}) = \limsup_{x \to x_0, \varepsilon \to 0^+} \frac{1}{\varepsilon^n} \int_{P_\varepsilon(x)} \operatorname{div} G(y)\, \mathrm{d}\mu(y), \tag{6}$$

where $\operatorname{div} G(y)$ stands for $\langle\langle JG(y), \mathrm{Id}\rangle\rangle = \operatorname{tr}(JG(y))$ (it is the divergence of the function G).

Proof. The function $\operatorname{div} G$ is a locally integrable function. The set of its Lebesgue points, denoted by $L_{\operatorname{div} G}$, is therefore of full measure (Theorem 3). We already mentioned that the definition of the generalized jacobian is “blind to null sets”. Hence, we can impose in (1) that x_i lies in $L_{\mathrm{div}G}$. It follows that the support function of $\mathscr{J}G(x_0)$ equals

$$\begin{aligned}\sigma_{\mathscr{J}G(x_0)}(M) &= \limsup_{x\to x_0,\, x\in D_G\cap L_{\mathrm{div}G}} \operatorname{div} G(x)\\ &= \limsup_{x\to x_0,\, x\in D_G\cap L_{\mathrm{div}G}} \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon^n}\int_{P_\varepsilon(x)} \operatorname{div} G(y)\,\mathrm{d}\mu(y)\\ &\leqslant \limsup_{x\to x_0} \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon^n}\int_{P_\varepsilon(x)} \operatorname{div} G(y)\,\mathrm{d}\mu(y)\\ &\leqslant \limsup_{x\to x_0,\,\varepsilon\to 0^+} \frac{1}{\varepsilon^n}\int_{P_\varepsilon(x)} \operatorname{div} G(y)\,\mathrm{d}\mu(y). \qquad (7)\end{aligned}$$

(we mentioned in Example 1 that $\{P_\varepsilon(x)\}_\varepsilon$ shrinks to x nicely. We therefore applied Proposition 1). Let us prove the reverse inequality. Let L denote the right hand side of (6). There exist two sequences $x_p \to x_0$ and $\varepsilon_p \to 0^+$ such that $L = \lim_{p\to\infty} L_p$, where

$$L_p := \frac{1}{\varepsilon_p^n}\int_{P_{\varepsilon_p}(x_p)} \operatorname{div} G(y)\,\mathrm{d}\mu(y).$$

We define the integral of a matrix with integrable entries as the matrix of the integrals of entries. For instance, $JG(y)$ is matrix defined almost everywhere, due to Rademacher's theorem:

$$JG(y) = \left(\frac{\partial G_i}{\partial x_j}(y)\right)_{i\in\{1,\dots,m\},\, j\in\{1,\dots,n\}}.$$

Because G is locally Lipschitz, so are its component functions G_i; hence, their partial derivatives $\partial G_i/\partial x_j$ are locally bounded and JG is a matrix that can be integrated on the bounded domain $P_{\varepsilon_p}(x_p)$:

$$\zeta_p := \frac{1}{\varepsilon_p^n}\int_{P_{\varepsilon_p}(x_p)} JG(y)\,\mathrm{d}\mu(y) \quad \text{exists and} \quad L_p = \langle\!\langle \zeta_p, \mathrm{Id}\rangle\!\rangle.$$

We claim that

$$\zeta_p \in \operatorname{co}\{\mathscr{J}G(P_{\varepsilon_p}(x_p))\}. \qquad (8)$$

First, $\mathscr{J}G(P_{\varepsilon_p}(x_p))$ is a compact set of $M_{m,n}(\mathbb{R})$: it is closed ($\mathscr{J}G$ is closed in the sense of [2, prop 2.6.2, p. 70] and $P_{\varepsilon_p}(x_p)$ is compact) and it is a subset of the closed ball centered at the origin and of radius K, where K denotes any Lipschitz constant of G near x_0. Therefore

$$\operatorname{co}\{\mathscr{J}G(P_{\varepsilon_p}(x_p))\} = \overline{\operatorname{co}}\{\mathscr{J}G(P_{\varepsilon_p}(x_p))\}.$$

Let $N \in M_{m,n}(\mathbb{R})$:

$$\begin{aligned}\langle\!\langle \zeta_p, N\rangle\!\rangle &= \frac{1}{\varepsilon_p^n}\int_{P_{\varepsilon_p}(x_p)} \langle\!\langle JG(y), N\rangle\!\rangle \,\mathrm{d}\mu(y)\\ &\leqslant \frac{1}{\varepsilon_p^n}\int_{P_{\varepsilon_p}(x_p)} \sigma_{\mathrm{co}\{\mathscr{J}G(P_{\varepsilon_p}(x_p))\}}(N)\,\mathrm{d}\mu(y)\\ &= \sigma_{\mathrm{co}\{\mathscr{J}G(P_{\varepsilon_p}(x_p))\}}(N).\end{aligned}$$

This is true for an arbitrary N, thus (8) holds true.

By Carathéodory's theorem, there exist $\zeta_p^0,\dots,\zeta_p^{n^2} \in \mathscr{J}G(P_{\varepsilon_p}(x_p)), \lambda_p^0,\dots,\lambda_p^{n^2} \geqslant 0$ with $\lambda_p^0+\cdots+\lambda_p^{n^2}=1$, such that

$$\zeta_p = \lambda_p^0\zeta_p^0+\cdots+\lambda_p^{n^2}\zeta_p^{n^2}. \tag{9}$$

We may assume that $\lambda_p^i \to \lambda^i$ as $p\to\infty$. Since for each i, $\{\zeta_p^i\}_p$ is a bounded sequence, we may also assume that $\zeta_p^i \to \zeta^i$. Invoking Proposition 2.6.2 of [2, p. 70], $\zeta^i \in \mathscr{J}G(x_0)$, for all i. Therefore

$$\begin{aligned}L &= \lim_{p\to\infty}[\lambda_p^0\langle\!\langle\zeta_p^0,\mathrm{Id}\rangle\!\rangle+\cdots+\lambda_p^{n^2}\langle\!\langle\zeta_p^{n^2},\mathrm{Id}\rangle\!\rangle]\\ &= \lambda^0\langle\!\langle\zeta^0,\mathrm{Id}\rangle\!\rangle+\cdots+\lambda^{n^2}\langle\!\langle\zeta^{n^2},\mathrm{Id}\rangle\!\rangle\\ &\leqslant \left(\sum_{i=0}^{n^2}\lambda^i\right)\sigma_{\mathscr{J}G(x_0)}(\mathrm{Id}) = \sigma_{\mathscr{J}G(x_0)}(\mathrm{Id}).\end{aligned}$$

Claim 1 is therefore proved. □

If $n=1$, we obtain

$$\begin{aligned}\sigma_{\mathscr{J}F(x_0)}(M) &= \limsup_{x\to x_0,\varepsilon\to 0^+}\frac{1}{\varepsilon}\int_x^{x+\varepsilon}(\langle M,F\rangle)'(y)\,\mathrm{d}\mu(y)\\ &= \limsup_{x\to x_0,\varepsilon\to 0^+}\frac{\langle M,F(x+\varepsilon)\rangle-\langle M,F(x)\rangle}{\varepsilon}\\ &= (\langle M,F\rangle)^\circ(x_0;1).\end{aligned}$$

If $n\geqslant 2$, apply a Green-Stokes formula to the locally Lipschitz function G. Eventually, we get:

$$\sigma_{\mathscr{J}G(x_0)} = \limsup_{x\to x_0,\varepsilon\to 0^+}\frac{1}{\varepsilon^n}\int_{\partial P_\varepsilon(x)}\langle G(y), n(y)\rangle\,\mathrm{d}\sigma(y).$$

Since $G=M^{\mathrm{T}}F$, proof of Theorem 1 is complete. □

Remark 1. In view of (7), We have also proved

$$\limsup_{x\to x_0,\varepsilon\to 0^+} \frac{1}{\varepsilon^n}\int_{P_\varepsilon(x)} \langle F(y), Mn(y)\rangle\, \mathrm{d}\mu(y) = \limsup_{x\to x_0} \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon^n}\int_{P_\varepsilon(x)} \langle F(y), Mn(y)\rangle\, \mathrm{d}\mu(y). \tag{10}$$

3.2. A technical formulation

The boundary of the hypercube, $\partial P_\varepsilon(x)$, is composed of $2n$ faces that can be parameterized by $[0,1]^{n-1}$. Denote the sum in which e_i does not appear by $\hat{t}_i = t_1 e_1 + \cdots + t_{n-1} e_n$. We then define:

$$F_i^+ := \{x + \varepsilon e_i + \varepsilon \hat{t}_i\colon (t_1,\ldots,t_{n-1}) \in [0,1]^{n-1}\}$$

and

$$F_i^- := \{x + \varepsilon \hat{t}_i\colon (t_1,\ldots,t_{n-1}) \in [0,1]^{n-1}\}.$$

Outer normal vectors of those two faces are e_i and $-e_i$, respectively. We now get a complete description of $\partial P_\varepsilon(x)$ when i describes $\{1,\ldots,n\}$. Through a change of variables in (3), we get

$$\begin{aligned}\sigma_{\mathscr{J}F(x_0)}(M) &= \limsup_{x\to x_0,\varepsilon\to 0^+} \frac{1}{\varepsilon^n}\sum_{i=1}^n \left[\int_{F_i^+} \langle F(y), Me_i\rangle\, \mathrm{d}\sigma(y) - \int_{F_i^-} \langle F(y), Me_i\rangle\, \mathrm{d}\sigma(y)\right] \\ &= \limsup_{x\to x_0,\varepsilon\to 0^+} \sum_{i=1}^n \int_{[0,1]^{n-1}} \frac{\langle F(x+\varepsilon e_i+\varepsilon\hat{t}_i) - F(x+\varepsilon\hat{t}_i), Me_i\rangle}{\varepsilon}\, \mathrm{d}t_1 \ldots \mathrm{d}t_{n-1}.\end{aligned} \tag{11}$$

4. Application: a new proof for a Clarke Jacobian chain rule

Known results about chain rules for generalized Jacobians were first established when one of the function was C^1 or real-valued. A general result about images appears in [2, p. 83]. To the best of our knowledge, the following result only appears in [3].

Theorem 4. *Let $\mathcal{O}$ be an open subset of $\mathbb{R}^n$ and consider two vector-valued functions $F : \mathcal{O} \subset \mathbb{R}^n \to \mathbb{R}^p$ and $G : \mathbb{R}^p \to \mathbb{R}^m$. We assume that F and G are locally Lipschitz. Then*

$$\mathscr{J}(G\circ F)(x_0) \subset \mathrm{co}\{\mathscr{J}G(F(x_0)) \circ \mathscr{J}F(x_0)\}. \tag{12}$$

Proof. These two sets are closed and convex. Thus, to get (12), we shall prove the following inequality, dealing with support functions:

$$\sigma_{\mathscr{J}(G\circ F)(x_0)}(M) \leqslant \max_{X\in\mathscr{J}G(F(x_0))} \sigma_{\mathscr{J}F(x_0)}(X^{\mathrm{T}}M). \tag{13}$$

Let $\eta > 0$. Since $\mathscr{J}G$ is an upper semicontinuous set-valued mapping, there exists $\delta > 0$ such that for any $y \in F(x_0) + \delta B$:

$$\begin{aligned} &\mathscr{J}G(y) \subset \mathscr{J}G(F(x_0)) + \delta B, \\ &G \quad \text{is Lipschitz continuous on } F(x_0) + \delta B. \end{aligned} \tag{14}$$

Choose x and ε small enough that

$$\begin{aligned} &F(P_\varepsilon(x)) \subset F(x_0) + \delta B, \\ &F \quad \text{is Lipschitz continuous on } P_\varepsilon(x). \end{aligned} \tag{15}$$

Denote by K a Lipschitz constant of F. Define a function $g_i : \mathbb{R}^p \to \mathbb{R}$ by $g_i(x) = \langle G(x), Me_i\rangle$ for any $x \in \mathbb{R}^p$. Using the technical expression we obtained in Subsection 3.2, rewrite $\sigma_{\mathscr{J}(G\circ F)(x_0)}(M)$ as

$$\limsup_{x\to x_0,\varepsilon\to 0^+} \sum_{i=1}^{n} \int_{[0,1]^{n-1}} \frac{g_i(F(x+\varepsilon\hat{t}_i+\varepsilon e_i)) - g_i(F(x+\varepsilon\hat{t}_i))}{\varepsilon}\, \mathrm{d}t_1 \ldots \mathrm{d}t_{n-1}. \tag{16}$$

Apply Lebourg's mean value theorem (see [2, Thm 2.3.7, p. 41]) to g_i between $F(x+\varepsilon\hat{t}_i)$ and $F(x+\varepsilon\hat{t}_i+\varepsilon e_i)$. There then exist $y_i \in [F(x+\varepsilon\hat{t}_i); F(x+\varepsilon\hat{t}_i+\varepsilon e_i)]$ and $p_i \in \partial g_i(y_i)$ such that

$$\frac{g_i(F(x+\varepsilon\hat{t}_i+\varepsilon e_i)) - g_i(F(x+\varepsilon\hat{t}_i))}{\varepsilon} = \frac{\langle F(x+\varepsilon\hat{t}_i+\varepsilon e_i) - F(x+\varepsilon\hat{t}_i), p_i\rangle}{\varepsilon}.$$

By definition of g_i, there exists $\zeta_i \in \mathscr{J}G(y_i)$ such that $p_i = \zeta_i^{\mathrm{T}} Me_i$. Consequently, $\sigma_{\mathscr{J}(G\circ F)(x_0)}(M)$ equals

$$\limsup_{x\to x_0,\varepsilon\to 0^+} \sum_{i=1}^{n} \int_{[0,1]^{n-1}} \frac{\langle F(x+\varepsilon\hat{t}_i+\varepsilon e_i) - F(x+\varepsilon\hat{t}_i), \zeta_i^{\mathrm{T}} Me_i\rangle}{\varepsilon}\, \mathrm{d}t_1 \ldots \mathrm{d}t_{n-1}. \tag{17}$$

Observe that $y_i \in \mathrm{co}\, F(P_\varepsilon(x))$; it follows from (15) that $y_i \in F(x_0) + \delta B$. By (14), we conclude that $\zeta_i \in \mathscr{J}G(F(x_0)) + \eta B$. There then exist $X_i \in \mathscr{J}G(F(x_0))$ and $Y_i \in \eta B$ such that $\zeta_i = X_i + Y_i$. It therefore follows from (17) that

$$\begin{aligned} \sigma_{\mathscr{J}(G\circ F)(x_0)}(M) &\leqslant \max_{X\in\mathscr{J}G(x_0)} \limsup_{x,\varepsilon} \sum \int \frac{\langle F(\ldots) - F(\ldots), X^{\mathrm{T}} Me_i\rangle}{\varepsilon}\, \mathrm{d}t \\ &\quad + \max_{Y\in\mathscr{J}G(x_0)} \limsup_{x,\varepsilon} \sum \int \frac{\langle F(\ldots) - F(\ldots), Y^{\mathrm{T}} Me_i\rangle}{\varepsilon}\, \mathrm{d}t \\ &\leqslant \max_{X\in\mathscr{J}G(x_0)} \sigma_{\mathscr{J}F(x_0)}(X^{\mathrm{T}}M) + nK|M|\eta. \end{aligned}$$

Equality (13) follows by letting $\eta \to 0^+$. □

5. The support function of plen $\mathscr{J}F(x_0)$

In this section, we first prove Theorem 2. We next give a straightforward corollary. We conclude the section by determining whether the infimum in (4) is attained.

Proof of Theorem 2. Let us define a mapping $\Phi : M_{m,n}(\mathbb{R}) \to \mathbb{R}$ by the following formula:

$$\Phi(M) = \inf\left\{ \sum_{i=1}^{k} (\langle v_i, F\rangle)^\circ(x_0; u_i) \colon \sum_{i=1}^{k} u_i \otimes v_i = M \right\}.$$

We must prove that $\sigma_{\text{plen}\,\mathscr{J}F(x_0)} = \Phi$. We observe that Φ is real-valued, sublinear and positively homogeneous of degree 1. We therefore conclude that Φ is the support function of some compact and convex set Σ of $M_{m,n}(\mathbb{R})$. We must prove that $\Sigma = \text{plen}\,\mathscr{J}F(x_0)$.

Let $\zeta \in \text{plen}\,\mathscr{J}F(x_0)$ and consider any decomposition of M in sum of rank-1 matrices:

$$M = u_1 \otimes v_1 + \cdots + u_k \otimes v_k.$$

It follows from Lemma 2 that

$$\begin{aligned}\langle\langle \zeta, M\rangle\rangle &= \langle\langle \zeta, u_1 \otimes v_1\rangle\rangle + \cdots + \langle\langle \zeta, u_k \otimes v_k\rangle\rangle \\ &\leqslant (\langle v_1, F\rangle)^\circ(x_0; u_1) + \cdots + (\langle v_k, F\rangle)^\circ(x_0; u_k).\end{aligned}$$

Then $\langle\langle \zeta, M\rangle\rangle \leqslant \Phi(M)$ for any $M \in M_{m,n}(\mathbb{R})$. This implies that $\zeta \in \Sigma$.

Now let $\zeta \in \Sigma$ and consider any $u \in \mathbb{R}^n$ and any $v \in \mathbb{R}^m$:

$$\langle\langle \zeta, u \otimes v\rangle\rangle \leqslant \Phi(u \otimes v) \leqslant (\langle v, F\rangle)^\circ(x_0; u) = \sigma_{\mathscr{J}F(x_0)}(u \otimes v).$$

Lemma 2 implies that $\zeta \in \text{plen}\,\mathscr{J}F(x_0)$. The proof is complete. □

Corollary 1. *Under the assumptions of Theorem* 2, *for any vectors* $u, u_1, \ldots, u_k$ *in* $\mathbb{R}^n$ *and* $v, v_1, \ldots, v_k$ *in* $\mathbb{R}^m$ *such that*

$$u \otimes v = u_1 \otimes v_1 + \cdots + u_k \otimes v_k,$$

the following holds:

$$(\langle v, F\rangle)^\circ(x_0; u) \leqslant (\langle v_1, F\rangle)^\circ(x_0; u_1) + \cdots + (\langle v_k, F\rangle)^\circ(x_0; u_k). \tag{18}$$

Remark 2. In the next subsection, we study the case when equality holds in (18).

5.1. *Study of the infimum in* (4)

We would like to know whether the infimum in (4) is attained. We will see that the answer is "yes except for a few matrices".

Proposition 3. *Let* $M \in M_{m,n}(\mathbb{R})$ *and consider* $\zeta \in \text{plen}\,\mathscr{J}F(x_0)$ *such that* $\sigma_{\text{plen}\,\mathscr{J}F(X_0)}(M) = \langle\langle \zeta, M\rangle\rangle$. *Then*

$$M \in \overline{\text{cone}}\{u \otimes v \colon \langle \zeta u, v\rangle = (\langle v, F\rangle)^\circ(x_0; u)\}.$$

Moreover, the infimum defining $\sigma_{\text{plen}\,\mathscr{J}F(x_0)}(M)$ *is attained if and only if*

$$M \in \text{cone}\{u \otimes v\colon\ \langle \zeta u, v\rangle = (\langle v, F\rangle)^0(x_0; u)\}.$$

Proof. We first derive a necessary and sufficient condition that ensures that the infimum defining $\sigma_{\text{plen}\,\mathscr{J}F(x_0)}(M)$ is attained.

Lemma 3. *Consider a decomposition of* M: $M = u_1 \otimes v_1 + \cdots + u_k \otimes v_k$.

$$\sigma_{\text{plen}\,\mathscr{J}F(x_0)}(M) = \sum_{i=1}^{k} (\langle v_i, F\rangle)^0(x_0; u_i)$$

if and only if there exists $\zeta \in \text{plen}\,\mathscr{J}F(x_0)$ *such that*

$$\forall i \in \{1, \ldots, k\}, \quad (\langle v_i, F\rangle)^0(x_0; u_i) = \langle \zeta u_i, v_i\rangle.$$

Proof. The "only if" part is straightforward. In order to prove the "if" part, let $\zeta \in \text{plen}\,\mathscr{J}F(x_0)$ such that $\sigma_{\text{plen}\,\mathscr{J}F(x_0)}(M) = \langle\!\langle \zeta, M\rangle\!\rangle$.

$$\forall i, \quad \langle \zeta u_i, v_i\rangle \leqslant (\langle v_i, F\rangle)^0(x_0; u_i),$$

$$\sum_{i=1}^{k} \langle \zeta u_i, v_i\rangle = \sum_{i=1}^{k} (\langle v_i, F\rangle)^0(x_0; u_i).$$

We conclude that $\langle \zeta u_i, v_i\rangle = (\langle v_i, F\rangle)^0(x_0; u_i)$ for all i. □

The first part of Proposition 3 remains to be demonstrated. We recall a more general result about the normal cone to a convex set defined by inequality constraints.

Lemma 4. *Consider a family* $\{s_\lambda\}_{\lambda \in \Lambda}$ *of vectors of* $\mathbb{R}^n$ *and a family* $\{\rho_\lambda\}_{\lambda \in \Lambda}$ *of real numbers. Define a convex set* C *by* $\bigcap_{\lambda \in \Lambda} \{\zeta\colon\ \langle \zeta, s_\lambda\rangle \leqslant \rho_\lambda\}$. *Then for any* $\zeta_0 \in C$:

$$N(C, \zeta_0) = \overline{\text{cone}}\{s_\lambda,\ \lambda \in \Lambda_0\},$$

where Λ_0 *denotes the set of all* λ *such that* $\langle \zeta_0, s_\lambda\rangle = \rho_\lambda$.

Proof. Let $E_\lambda = \{\zeta\colon\ \langle \zeta, s_\lambda\rangle \leqslant \rho_\lambda\}$. Then:

$$\begin{aligned} T(C, \zeta_0) &= \overline{\text{cone}}\{C - \zeta_0\} = \overline{\text{cone}} \bigcap_{\lambda \in \Lambda} \{E_\lambda - \zeta_0\} \\ &= \bigcap_{\lambda \in \Lambda_0} \{E_\lambda - \zeta_0\} = \{\zeta\colon\ \langle \zeta, s_\lambda\rangle \leqslant 0,\ \forall \lambda \in \Lambda_0\} \\ &= \{s_\lambda,\ \lambda \in \Lambda_0\}^0. \qquad \square \end{aligned}$$

We apply this result to the family R_1 of rank-1 matrices and to the corresponding family of real numbers $\{(\langle v, F\rangle)^0(x_0; u)\}_{u \otimes v \in R_1}$. Lemma 2 implies that $C = \text{plen}\,\mathscr{J}F(x_0)$. Hence, Lemma 4 implies the first part of Proposition 3. Its proof is therefore complete. □

Remark 3. The question of whether

$$\text{cone}\{u \otimes v\colon \langle \zeta u, v\rangle = (\langle v, F\rangle)^\circ(x_0; u)\}$$

is closed remains unanswered.

6. Application to second-order differentiation theory

Second-order differentiation theory provides tools that help in the understanding of optimality; in particular it permits the formulation of sufficient conditions for local optimality. Generalized Hessians, that is to say Hessians for nondifferentiable functions, are the cornerstone of this theory. Various Hessians have been introduced for $C^{1,1}$ functions i.e. differentiable functions whose gradients are locally Lipschitz continuous. They are very often closed and convex and we have already pointed out that the support function of a closed convex set is an important tool for studying it. The purpose of this subsection is to give analytical expressions of the support functions of three such sets.

Hiriart-Urruty, Strodiot and Hien Nguyen [6] introduced a Hessian in the sense of Clarke for $C^{1,1}$ functions defined in a finite dimensional setting. For $f : \mathcal{O} \subset \mathbb{R}^n \to \mathbb{R}$, $C^{1,1}$, they defined

$$\partial_H^2 f(x_0) := \mathcal{J}(Jf)(x_0).$$

Theorem 1 enables us to give the support function of this nonempty compact convex set.

Proposition 4.

$$\sigma_{\partial_H^2 f(x_0)}(M) = \limsup_{x \to x_0, \varepsilon \to 0+} \frac{1}{\varepsilon^n} \int_{frP_\varepsilon(x)} \langle Jf(y), Mn(y)\rangle \, \mathrm{d}\sigma(y). \tag{19}$$

In an infinite dimensional setting, Cominetti and Correa [4] defined a generalized Hessian as a set-valued function. Let X denote a Banach space and let X^* denote its topological dual. For a $C^{1,1}$ function $f : X \to \mathbb{R}$, $\partial^2 f(x_0) : X \rightrightarrows X^*$ is defined by:

$$\forall u \in X, \quad \partial^2 f(x_0)(u) = \{x \in X^*\colon \langle x, v\rangle \leqslant f^\infty(x_0; u, v)\},$$

where f^∞ is the following second-order directional derivative:

$$f^\infty(x_0; u, v) = \limsup_{\substack{x \to x_0 \\ \varepsilon \to 0^+, \delta \to 0^+}} \frac{f(x + \varepsilon u + \delta v) - f(x + \varepsilon u) - f(x + \delta v) + f(x)}{\varepsilon\delta}.$$

They proved that for $X = \mathbb{R}^n$:

$$f^\infty(x_0; u, v) = (\langle v, \nabla f\rangle)^0(x_0; u) \quad \text{and} \quad \partial^2 f(x_0)(u) = \partial_H^2 f(x_0)u. \tag{20}$$

Palés and Zeidan [8] introduced another generalized Hessian. They considered $\partial_\infty^2 f(x_0)$, the family of bounded linear operators $A : X \to X^*$ that satisfy $Au \in \partial^2 f(x_0)(u)$ for all

$u\in X$. It follows from (20) that for $X=\mathbb{R}^n$:

$$\partial^2_\infty f(x_0)=\text{plen}\,\partial^2_H f(x_0).$$

By applying Theorem 2 and using (20), we obtain:

Proposition 5. *Let $f:\mathcal{O}\subset\mathbb{R}^n\to\mathbb{R}$ be a $C^{1,1}$ function. Then for all $M\in M_n(\mathbb{R})$:*

$$\sigma_{\partial^2_\infty f(x_0)}(M)=\inf\left\{\sum_{i=1}^k f^\infty(x_0;u_i,v_i):\ M=\sum_{i=1}^k u_i\otimes v_i\right\}.$$

A third generalized Hessian were introduced by Palés and Zeidan [8]:

$$\partial^2 f(x_0):=\{B\in\mathcal{B}(X):\ B(u,v)\leqslant(\langle v,\nabla f\rangle)^0(x_0;u)\},$$

where $\mathcal{B}(X)$ denotes the set of symmetric bilinear forms on X. In the finite dimensional setting, $\mathcal{B}(X)$ can be identified with S_n, the set of symmetric $n\times n$ matrices. In view of Lemma 2, it is therefore obvious that when $X=\mathbb{R}^n$:

$$\partial^2 f(x_0)=\text{plen}\,\partial^2_H f(x_0)\cap S_n. \tag{21}$$

Theorem 2 can be applied to prove the next result.

Theorem 5.

$$\sigma_{\partial^2 f(x_0)}(M)=\inf\left\{\sum_{i=1}^k(\langle v_i,\nabla f\rangle)^\circ(x_0;u_i):\ \sum_{i=1}^k U_i\otimes V_i=\frac{M+M^{\mathrm{T}}}{2}\right\}. \tag{22}$$

Proof. The right hand side of (22) is precisely $\sigma_{\text{plen}\partial^2_H f(x_0)}(M+M^{\mathrm{T}}/2)$. It is clear that this function of M is sublinear, positively homogenous and real-valued. Hence, this is the support function of a compact convex set Σ of $M_n(\mathbb{R})$. We are going to prove that $\Sigma=\partial^2 f(x_0)$. Due to (21), it is sufficent to prove that $\Sigma=\text{plen}\,\partial^2_H f(x_0)\cap S_n$.

First, we observe that

$$\sigma_\Sigma(M)\leqslant\frac{1}{2}\sigma_{\text{plen}\partial^2_H f(x_0)}(M)+\frac{1}{2}\sigma_{\text{plen}\partial^2_H f(x_0)}(M^{\mathrm{T}}),$$

The first equality in (20) implies that $(\langle v,\nabla f\rangle)^\circ(x_0;u)=(\langle u,\nabla f\rangle)^\circ(x_0;v)$. Consequently,

$$\sigma_{\text{plen}\partial^2_H f(x_0)}(M)=\sigma_{\text{plen}\partial^2_H f(x_0)}(M^{\mathrm{T}}),$$

and $\sigma_\Sigma(M)\leqslant\sigma_{\text{plen}\partial^2_H f(x_0)}(M)$ follows. Hence Σ is a subset of plen $\partial^2_H f(x_0)$. Moreover, if A is an antisymmetric matrix, $\sigma_\Sigma(A)=0$ and $\sigma_\Sigma(-A)=0$. This implies that for any $\zeta\in\Sigma$ and any antisymmetric matrix A, $\langle\zeta,A\rangle=0$. Using the fact that the space which is orthogonal to S_n is the space of antisymmetric matrices, we can claim that Σ is a subset of S_n.

Conversely,

$$\begin{aligned}\sigma_{\partial^2 f(x_0)}(M) &= \max\{\langle\!\langle \zeta, M\rangle\!\rangle : \zeta \in \operatorname{plen} \partial_H^2 f(x_0) \cap S_n\}\\ &= \max\left\{\langle\!\langle \zeta, \frac{M+M^{\mathrm{T}}}{2}\rangle\!\rangle\right.\\ &\qquad \left. + \langle\!\langle \zeta, \frac{M-M^{\mathrm{T}}}{2}\rangle\!\rangle : \zeta \in \operatorname{plen} \partial_H^2 f(x_0)\cap S_n\right\}\\ &= \max\left\{\langle\!\langle \zeta, \frac{M+M^{\mathrm{T}}}{2}\rangle\!\rangle : \zeta \in \operatorname{plen} \partial_H^2 f(x_0)\cap S_n\right\}\\ &\leqslant \sigma_{\operatorname{plen}\partial_H^2 f(x_0)}\left(\frac{M+M^{\mathrm{T}}}{2}\right).\end{aligned}$$

We used the fact that $(M - M^{\mathrm{T}})/2$ is an antisymmetric matrix and that, consequently, it is orthogonal to symmetric ones. The proof is therefore complete. □

7. Connections with known results; examples

7.1. The special cases $m = 1$ and $n = 1$

If $n = 1$ or $m = 1$, then $\mathcal{J}F(x_0)$ is plenary: as all $m \times 1$ and $1 \times n$ matrices are of rank less than or equal to 1, this is a straightforward consequence of Lemma 2.

Proposition 6 (Hiriart Urruty [5], Clarke [2]). *Let $f : \mathcal{O} \subset \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function.*

- *If $n = 1$ and $M \in M_{m,1}(\mathbb{R})$, then there exists $v \in \mathbb{R}^m$ such that $M = 1 \otimes v$ and $\sigma_{\mathcal{J}F(x_0)}(1 \otimes v) = (\langle v, F\rangle)^\circ(x_0; 1)$.*
- *If $m = 1$ and $M \in M_{1,n}(\mathbb{R})$, then there exists $u \in \mathbb{R}^n$ such that $M = u \otimes 1$ and $\sigma_{\mathcal{J}F(x_0)}(u \otimes 1) = F^\circ(x_0; u)$.*

In the case $m = 1$, the connection between (3) and F° is not clear. The reason is that the problem must not be reduced to the case $m = n$ but to $n = m$ (see the beginning of the proof of Theorem 1). We therefore have two different analytic expressions of the support function of $\sigma_{\mathcal{J}F(x_0)}$.

7.2. plen $\mathcal{J}F(x_0)$ is a subset of $\partial f_1(x_0) \times \cdots \times \partial f_m(x_0)$

Considering a particular rank-1 matrices decomposition of a $m \times n$ matrix, one can easily prove that the plenary hull of the Clarke generalized jacobian is a subset of the cartesian product of the subdifferentials of the component functions. The following result is more precise than [2, Prop 2.6.2, p. 70]. Moreover, the proof is new.

Proposition 7 (Hiriart Urruty [5]). *Under assumptions of Theorem* 1, *consider* $x_0 \in \mathcal{O}$ *and* $F = (f_1, \ldots, f_m) : \mathcal{O} \to \mathbb{R}$. *Then*

$$\mathscr{J}F(x_0) \subset \text{ plen}\, \mathscr{J}F(x_0) \subset \partial f_1(x_0) \times \cdots \times \partial f_m(x_0).$$

Proof. The first inclusion is straightforward. Let M be any $m \times n$ matrix. Consider its row decomposition: $\begin{bmatrix} u_1^{\mathrm{T}} \\ \vdots \\ u_m^{\mathrm{T}} \end{bmatrix} = u_1 \otimes e_1 + \cdots + u_m \otimes e_m,\ u_i \in \mathbb{R}^n$. Theorem 2 yields

$$\begin{aligned} \sigma_{\text{plen}\,\mathscr{J}F(x_0)}(M) &\leqslant (\langle e_1, F\rangle)^\circ(x_0; u_1) + \cdots + (\langle e_m, F\rangle)^\circ(x_0; u_m) \\ &= \sigma_{\partial f_1(x_0) \times \cdots \times \partial f_m(x_0)}(M). \end{aligned}$$

7.3. A nonconvex plenary set

This example comes from [10]. Let us consider the following nonconvex set:

$$\mathscr{A} = \text{co}\left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} \bigcup \text{co}\left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

This set is plenary.

To prove it, one considers a generic matrix $\mathscr{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $Au \in \mathscr{A}u$ for all u. Choosing successively $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $u = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, one gets $c = 0$, $b = 0$, $ad = 0$ and $a, d \in [0, 1]$. This implies that $A \in \mathscr{A}$.

7.4. $\mathscr{J}F(x_0)$ can be strictly smaller than plen $JF(x_0)$

Let $\{M_i\}_{i=1}^k$ be $m \times n$ matrices. Consider

$$P = \{\zeta\colon \langle \zeta, M_i\rangle \leqslant \rho_i,\ i = 1, \ldots, k\}.$$

Such an intersection of closed half-spaces is precisely what is called a convex polyhedra. Assume that k is minimal in following sense: each intersection of less than k considered half-spaces is larger.

Proposition 8. *Under the assumptions and notations above*, P *is plenary if and only if*, M_i *is of rank lower or equal to 1*, *for* $i = 1, \ldots, k$.

It is therefore easy to construct functions whose generalized jacobians are not plenary. Considering a piecewise affine function, one can get a generalized jacobian that is polyhedral:

$$\mathscr{A} = \text{co}\left\{ \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}, \begin{bmatrix} -1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix} \right\}.$$

This set can be viewed as the intersection of $\{\langle I_2, .\rangle = 0\}$ with other half-spaces. Since $rank(I_2) = 2$, Proposition 8 implies that the general jacobian $\mathscr{A}$ is not plenary.

Acknowledgements

The author is indebted to J.-B. Hiriart-Urruty who proposed the subject of study and supervised it from beginning to end.

References

[1] F.H. Clarke, Generalized gradients and applications, Trans. Amer. Math. Soc. 205 (1975) 247–262.
[2] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[3] F.H. Clarke, Analyse non lisse et optimisation. Cours de 3ème cycle, Ecole doctorale de mathématiques, Université Paul Sabatier, 1993–1994.
[4] R. Cominetti, R. Correa, A generalized second-order derivative in nonsmooth optimization, SIAM J. Control Optim. 28 (1990) 789–809.
[5] J.-B. Hiriart-Urruty, Characterizations of the plenary hull of the generalized Jacobian matrix, Math. Program. Study 17 (1982) 1–12.
[6] J.-B. Hiriart Urruty, J.-J. Strodiot, V. Hien Nguyen, Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data, Appl. Math. Optim. 11 (1984) 43–56.
[7] J.-B. Hiriart-Urruty, L. Thibault, Existence et caractérisation de différentielles généralisées d'applications localement lipschitziennes d'unespace de Banach séparable dans un espace de Banach réflexif séparable, Notes aux C.R.A.S. 290 (23) (1980) 1091–1094.
[8] Zs. Páles, V. Zeidan, Generalized Hessian for $C^{1,1}$ functions in infinite dimensional normed spaces, Math. Program. 74 (1996) 59–78.
[9] W. Rudin, Real and Complex Analysis, McGraw Hill, 3rd edition, 1987.
[10] T.H. Sweetser, A minimal set-valued strong derivative for vector-valued Lipschitz functions, J. Optim. Theory Appl. 23 (4) (1977) 549–562.
[11] J. Warga, Fat homeomorphisms and derivate unbounded containers, J. Math. Anal. Appl. 81 (1981) 549.