TAME MAPPINGS ARE SEMISMOOTH

J. Bolte, A. Daniilidis & A.S. Lewis

Dedicated to Stephen Robinson, who has so many of the best ideas first.

Abstract Superlinear convergence of the Newton method for nonsmooth equations requires a "semismoothness" assumption. In this work we prove that locally Lipschitz functions definable in an o-minimal structure (in particular semialgebraic or globally subanalytic functions) are semismooth. Semialgebraic, or more generally, globally subanalytic mappings present the special interest of being γ-order semismooth, where γ is a *positive* parameter. As an application of this new estimate, we prove that the error at the kth step of the Newton method behaves like $O(2^{-(1+\gamma)^k})$.

Key words Semismoothness, semi-algebraic function, o-minimal structure, nonsmooth Newton method, structured optimization problem, superlinear convergence.

1 Introduction

Extensions of the Newton method for solving nonsmooth equations $F(x) = 0$ have been widely studied over the last two decades: early examples are [15, 23] (see also [24]). As pointed out in [22], superlinear convergence depends on "semismoothness" of the function F, a notion extended from earlier work on optimization originating with [20]. A good survey appears in [13].

Consider a locally Lipschitz function $F : \mathbf{R}^n \to \mathbf{R}^m$, and denote the set of points in $\mathbf{R}^n$ where F is differentiable by D. (By Rademacher's theorem, the complement of D has measure zero.) Following [22], we call F *semismooth* at a point $x \in \mathbf{R}^n$ if its directional derivative

$$(1.1) \qquad F'(x; d) = \lim_{t \downarrow 0} \frac{F(x + td) - F(x)}{t}$$

exists for every vector $d \in \mathbf{R}^n$, and as $d \to 0$ with $x + d \in D$, we have

$$(1.2) \qquad F'(x + d; d) - F'(x; d) = o(d),$$

with as usual $\lim_{d \to 0, \, d \neq 0} ||d||^{-1} o(d) = 0$.

Quadratic rather than superlinear convergence of the Newton method requires *strong semismoothness* instead (see [12]), where the $o(d)$ term is replaced by $O(||d||^2)$, that is, it is bounded near 0 by a function $c \, ||d||^2$, where $c > 0$. More generally, a mapping satisfying (1.2) with $o(d) = O(||d||^{1+\gamma})$ with $\gamma > 0$ is called γ-*order semismooth* or simply γ-*semismooth*, see [22, 26]. The class of semismooth functions is very broad: see for example the discussion in [21].

If the function F satisfies a rather strong notion of "piecewise smoothness", then it must be semismooth. To be precise, let us consider functions $F_1, F_2, \ldots, F_k : \mathbf{R}^n \to \mathbf{R}^m$. Suppose, for every point $y \in \mathbf{R}^n$ near x, each function F_j is continuously differentiable at y, the function F is continuous at y, and $F(y) = F_j(y)$ for some index j. In this case, F must be semismooth at x (see [26] and also [17]).

The usefulness of the above class of piecewise smooth functions in checking semismoothness in concrete examples is restricted by the requirement that each function F_j must be continuously differentiable throughout a neighborhood of the point of interest, rather than simply on open regions where F_j agrees with F. In part due to this complication, we take here a fresh approach to recognizing semismoothness. To illustrate this, consider for example the function

$$f(x,y) = \begin{cases} \sqrt{y^2 - x^2}, & \text{if } y > 2\,|x| \\[2mm] \sqrt{3}\,|x|, & \text{if } y \leq 2\,|x|. \end{cases}$$

The above function is clearly continuous and everywhere smooth except on the sets $\{0\} \times (-\infty, 0]$ and $\{(t, 2\,|t|) : t \in \mathbf{R}\}$ with gradient no larger in norm than $\sqrt{3}$. It is easily seen that f is (globally) Lipschitz with constant $\sqrt{3}$ and everywhere directionally differentiable, and as we discuss below, f is easily recognized to be semialgebraic. Hence f is semismooth, by the result we present here (cf. Theorem 3.6). However, verifying that f is piecewise smooth (at $(0,0)$ for example) in the above sense is not immediate.

A rich class of concrete functions is provided by the notion of a *semi-algebraic* subset of $\mathbf{R}^n$, that is, a set defined by some Boolean combination of real polynomial equations and inequalities. The function F is *semi-algebraic* if its graph $\{(x,y) : y \in F(x)\}$ is semi-algebraic. The broad applicability of semi-algebraic sets follows largely from the Tarski-Seidenberg principle, which guarantees that the projection $(x_1, x_2, \ldots, x_n) \mapsto (x_2, x_3, \ldots, x_n)$ preserves the semi-algebraic property. Good references are [4, 6]. The qualitative properties of semialgebraic mappings are shared by a much bigger class called mappings *definable in an o-minimal structure* over $\mathbf{R}$, or simply definable mappings. A slightly more general notion is that of a *tame* mapping, being a mapping whose graph has a definable intersection with every "bounded box" (see Definition 2.2). O-minimal structures over $\mathbf{R}$ correspond in some sense to an axiomatization of some of the prominent geometrical properties of semialgebraic geometry [11, 9] and particularly of the stability under projection. Due to the variety of optimization problems that can be formulated within the framework of o-minimal structures, our main results are stated for tame functions.

The main result of this note is to establish that locally Lipschitz tame (definable) functions are semismooth (Theorem 3.6). Tame sets stratify into locally finite unions of relatively open smooth manifolds, and consequently, tame functions enjoy strong piecewise smoothness properties. However, exploiting stratification to deduce semismoothness via piecewise smoothness seems not transparent due to the complication we noted above. In part due to this complication, and as a component in our ongoing study of semialgebraic/tame properties in optimization (see [1], [2], [3]), we propose here a much more basic approach based on the curve selection lemma.

It is worthwhile-mentioning that semialgebraic or globally subanalytic mappings enjoy a stronger property than semismoothness since they are actually γ-semismooth with $\gamma > 0$. The interest of this extra information is somewhat comparable to what can be observed for functions that satisfy the Łojasiewicz inequality [18, 16]. Indeed, for algorithms in which semismoothness plays a key role (Newton methods, Bundle methods) the positive parameter γ somehow measures the rate of convergence, as it is the case for the Łojasiewicz exponent in the study of subgradient methods [1]. We illustrate this idea by proving that the Newton method for semialgebraic, or more generally,

for globally subanalytic mappings converges superlinearly. More precisely, we show that the error at the k-step behaves like $O(2^{-(1+\gamma)^k})$.

2   Preliminaries

Notation Throughout this work we shall consider the Euclidean vector space $\mathbf{R}^n$ endowed with its canonical scalar product $\langle \cdot, \cdot \rangle$, and we shall denote its associated norm by $|| \cdot ||$.

Let $U \subset \mathbf{R}^n$ be a nonempty open subset of $\mathbf{R}^n$. A mapping F is said to be *directionally differentiable at $x \in U$ along d* if the following limit

$$F'(x; d) := \lim_{t \downarrow 0} \frac{F(x + td) - F(x)}{t}$$

exists. This defines a mapping $F'(\cdot; \cdot)$ with domain $\mathcal{D} \subset U \times \mathbf{R}^n$ which we call the directional derivative of F. The function F is said to be *Gâteaux differentiable* at x if $F'(x; \cdot)$ is defined everywhere on $\mathbf{R}^n$ and linear. The linear map $F'(x; \cdot)$ is then denoted by $F'_G(x)$. We say that F is *Fréchet differentiable* at $x \in U$ if it is Gâteaux differentiable at x with in addition $F(x + d) - F(x) - F'(x; d) = o(||d||)$. The Fréchet derivative is denoted by F'. In the sequel the domains of F'_G and F' are respectively denoted by D_G and D. Finally, $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ denotes the vector space of linear mappings from $\mathbf{R}^n$ to $\mathbf{R}^m$.

Let us recall a few definitions concerning o-minimal structures (see for instance van der Dries-Miller [11] and references therein).

Definition 2.1. [o-minimal structure] [9, Definition 1.5] An *o-minimal* structure on $(\mathbf{R}, +, .)$ is a sequence of Boolean algebras $\mathcal{O} = \{\mathcal{O}_n\}$ of "definable" subsets of $\mathbf{R}^n$, such that for each $n \in \mathbf{N}$

  (i) if A belongs to $\mathcal{O}_n$, then $A \times \mathbf{R}$ and $\mathbf{R} \times A$ belong to $\mathcal{O}_{n+1}$;

  (ii) if $\Pi : \mathbf{R}^{n+1} \to \mathbf{R}^n$ is the canonical projection onto $\mathbf{R}^n$ then for any A in $\mathcal{O}_{n+1}$, the set $\Pi(A)$ belongs to $\mathcal{O}_n$;

  (iii) $\mathcal{O}_n$ contains the family of algebraic subsets of $\mathbf{R}^n$, that is, every set of the form

$$\{x \in \mathbf{R}^n : p(x) = 0\},$$

  where $p : \mathbf{R}^n \to \mathbf{R}$ is a polynomial function ;

  (iv) the elements of $\mathcal{O}_1$ are exactly the finite unions of intervals and points.

A mapping $F : S \subset \mathbf{R}^n \to \mathbf{R}^m$ is said to be definable in $\mathcal{O}$ if its graph is definable in $\mathcal{O}$ as a subset of $\mathbf{R}^n \times \mathbf{R}^m$.

Definition 2.2. A subset A of $\mathbf{R}^n$ is called *tame* if for all $r > 0$ there exists an o-minimal structure $\mathcal{O}$ over $\mathbf{R}$ such that the intersection of A with $[-r, r]^n$ is definable in $\mathcal{O}$. Similarly a mapping $F : U \subset \mathbf{R}^n \to \mathbf{R}^m$ is called tame if its graph is a tame subset of $\mathbf{R}^n \times \mathbf{R}^m$.

Remark 2.3. Restrictions of tame functions to definable bounded sets do not necessarily belong to an o-minimal structure. Take for instance $F(p) = 1/p$ with $p \in \text{dom } F := \cup_{n \in \mathbb{N}^*} \{1/n\}$. Using items (ii) and (iv) of Definition 2.1 one sees that the restriction of F to $[-1, 1]$ cannot belong to any o-minimal structure. However if F is a tame mapping with the property that $F(B)$ is bounded for every bounded subset $B \subset \mathbf{R}^n$ (which is the case when F is for instance continuous), then for all $r > 0$ the mapping $F_{|\,[-r,r]^n}$ belongs to some o-minimal structure.

Example 2.4 (semialgebraic sets). The first example of o-minimal structures is given by the class $\mathcal{SA}$ of *semialgebraic* objects. A set $A \subset \mathbf{R}^n$ is called semialgebraic if it can be written as

$$A = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{ x \in \mathbf{R}^n : P_{ij}(x) = 0, \; Q_{ij}(x) < 0 \},$$

where the $P_{ij}, Q_{ij} : \mathbf{R}^n \to \mathbf{R}$ are polynomial functions on $\mathbf{R}^n$. The fact that $\mathcal{SA}$ is an o-minimal structure relies on the Tarski-Seidenberg principle (see [6]) which asserts that item (ii) is true in this class.

Let us also observe that *any* o-minimal structure on $\mathbf{R}$ contains the class $\mathcal{SA}$ of semialgebraic sets. In other words, $\mathcal{SA}$ is the *smallest*[1] o-minimal structure on $\mathbf{R}$.

Sets and functions belonging to some o-minimal structure enjoy many qualitative properties that are quite similar to those occurring in semialgebraic geometry. The reader is referred to [11, 9] for a comprehensive account on the topic. In this paper we will essentially use the following results.

Let $\mathcal{O}$ be an o-minimal structure on $(\mathbf{R}, +, .)$.

Monotonicity lemma [11, Theorem 4.1] *Let $f : I \subset \mathbf{R} \to \mathbf{R}$ be a definable function and $k \in \mathbf{N}$. Then there exists a finite partition of I into p disjoint intervals $I_1, \ldots, I_p$ such that f restricted to each nontrivial interval I_j, $j \in \{1, \ldots, p\}$ is C^k and either strictly monotone or constant.*

Curve selection lemma [11, Theorem 4.6] *Let A be a definable subset of $\mathbf{R}^n$ and let x be an element of $\overline{A}$ (the closure of A). Then for all $k \in \mathbf{N}$ there exist $\epsilon > 0$ and a C^k definable path $p : (-\epsilon, 1) \to \mathbf{R}^n$ such that $p(0) = x$ and $p((0, 1]) \subset A$.*

A major part of the interest in dealing with definable objects consists of their remarkable stability properties. These rely *in fine* on the projection stability assumption (iii) (Definition 2.1). Using for instance [9, Theorem 1.13] the reader can establish easily the following results:

Stability results Let $\mathcal{O}$ be an o-minimal structure over $\mathbf{R}$.

(a) *Given $A \subset \mathbf{R}^n$, $B \subset \mathbf{R}^m$ and a definable mapping $F : A \to B$ in $\mathcal{O}$, then for all $C \subset A$ and $E \subset B$ definable in $\mathcal{O}$, the sets $F(C)$ and $F^{-1}(E)$ are definable in $\mathcal{O}$.*

(b) *Let $F : U \to \mathbf{R}^m$ be a definable mapping in $\mathcal{O}$, where U is a nonempty open subset of $\mathbf{R}^n$. Then the mappings $F'(\cdot; \cdot) : \mathcal{D} \to \mathbf{R}^m$, $F'_G : D_G \to \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ and $F' : D \to \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ are definable in $\mathcal{O}$.*

---

[1]This is due to axiom (iii). Sometimes this axiom is omitted from the definition of an o-minimal structure, allowing smaller classes than $\mathcal{SA}$, for instance the structure $\mathcal{SL}$ of *semilinear sets*.

(c) *If $F : U \subset \mathbf{R}^n \to \mathbf{R}^m$ and $G : V \subset F(U) \to \mathbf{R}^p$ are definable mappings then $G \circ F : U \to \mathbf{R}^p$ is definable in $\mathcal{O}$.*

Example 2.5. (a) **(Globally subanalytic sets)** There exists an o-minimal structure, denoted by $\mathbf{R}_{\mathrm{an}}$, that contains all sets of the form $\{(x,t) \in [-1,1]^n \times \mathbf{R} : f(x) = t\}$ where $f : [-1,1]^n \to \mathbf{R}$ ($n \in \mathbf{N}$) is an analytic function that can be extended analytically on a neighborhood of the square $[-1,1]^n$. The sets belonging to this structure are called *globally subanalytic sets*. Early applications of this structure in optimization can be found in [10, 19].

(b) **(log-exp structure)** There exists an o-minimal structure containing $\mathbf{R}_{\mathrm{an}}$ and the graph of $\exp : \mathbf{R} \to \mathbf{R}$. This structure is denoted by $\mathbf{R}_{\mathrm{an,exp}}$.

Remark 2.6. In view of Example 2.5 (a) real-analytic mappings are obviously tame. They are not however definable in some o-minimal structure in general. Consider for instance $f(x) = \sin x$ ($x \in \mathbf{R}$) whose zero set, *i.e.* $\{x \in \mathbf{R} : \sin x = 0\}$, is discrete and infinite. By using Definition 2.1 (iv) and the stability result (a) we see that there does not exist an o-minimal structure over $\mathbf{R}$ in which f is definable.

Due to their "polynomial nature" semialgebraic and globally subanalytic functions of one variable admit the so-called Puiseux development.

Puiseux development [11] *Let $f : (0,1) \to \mathbf{R}$ be a globally subanalytic function. Then there exist some integers $p,q \in \mathbf{Z}$, with $q > 0$, a sequence $\{a_i\}_{i=p,p+1,\ldots}$ with $a_p \neq 0$ and a real number $\epsilon > 0$ such that*

$$f(x) = \sum_{i=p}^{+\infty} a_i \, x^{i/q} \quad \text{for all } x \in (0,\epsilon).$$

Checking the semialgebraicity of a set in practice is often easy. On the other hand, one needs to be careful when analytic functions come into play, as it is the case with $\mathbf{R}_{\mathrm{an}}$. A formal approach to global subanalyticity can be found in [11].

3 Differentiability and semismoothness results

Let $\mathcal{O}$ be an o-minimal structure on $(\mathbf{R}, +, .)$.

3.1 Fréchet and Gâteaux derivative of tame mappings

The monotonicity lemma implies several elementary but very useful results concerning functions of one variable.

Lemma 3.1. *Let $\phi, \psi : [0, \epsilon) \to \mathbf{R}$ be definable functions continuous at 0 with $\phi(0) = \psi(0) = 0$.*

(i) **(local differentiation of inequalities)** *If $\psi \geq \phi \geq 0$ on $[0, \epsilon)$ there exists $\epsilon_1 \in (0, \epsilon)$ such that ψ and ϕ are differentiable on $(0, \epsilon_1)$ with $\psi'(t) \geq \phi'(t)$ for all t in $(0, \epsilon_1)$.*

(ii) (**de l'Hôpital inverse rule**) *Let $l \in \mathbf{R}$ and assume that $\psi'(t) > 0$ for all $t \in (0, \epsilon)$ sufficiently small. Then*

$$\lim_{t \to 0^+} \frac{\phi(t)}{\psi(t)} = l \quad \Longrightarrow \quad \lim_{t \to 0^+} \frac{\phi'(t)}{\psi'(t)} = l.$$

Proof. (i) By the monotonicity lemma the function $t \mapsto \psi(t) - \phi(t)$ is C^1 and monotone on $(0, \epsilon_1)$ for ϵ_1 sufficiently small. Thus, $(\psi - \phi)'(t)$ has a constant sign (or is equal to zero) for all $t > 0$ small enough. By integrating and using the assumption $\psi \geq \phi$ we obtain that for all $t > 0$ small enough, either $\psi'(t) = \phi'(t)$ (thus $\psi = \phi$), or $\psi'(t) - \phi'(t) \geq 0$.

(ii) We have $(\phi(t) - l\psi(t))/\psi(t) \to 0$ as $t \downarrow 0$, where $\phi - l\psi$ vanishes at zero. Replacing, if necessary, $\phi - l\psi$ by $\tilde{\phi}$, we see that there is no loss of generality in assuming that $l = 0$. Since ϕ has a constant sign, replacing if necessary ϕ by its opposite, we may also suppose that $\phi \geq 0$. Using the monotonicity lemma we obtain that $\phi'(t) \geq 0$ for all $t \in (0, \epsilon)$. Since by assumption $\lim_{t \to 0^+} \phi(t)/\psi(t) = l = 0$, for any $\delta > 0$ and all t small enough we have $\delta\psi(t) - \phi(t) \geq 0$. Applying (i) we obtain $\delta\psi'(s) - \phi'(s) \geq 0$ for all $s > 0$ sufficiently small. Thus

$$0 \leq \lim_{t \to 0^+} \frac{\phi'(t)}{\psi'(t)} \leq \delta,$$

where the existence of the limit is due to the monotonicity lemma. The result follows by letting δ go to zero. $\qquad\square$

Applying Lemma 3.1 (ii) with $\psi(t) = t$ one obtains the following

Corollary 3.2 (right continuity of the derivative). *Let $\phi : [0, \epsilon) \to \mathbf{R}$ be a definable function, continuous at 0 with $\phi(0) = 0$, and let us assume that the limit*

$$\phi'(0^+) = \lim_{t \to 0^+} \frac{\phi(t)}{t}$$

is finite (owing to the monotonicity lemma the limit always exists). Then

$$\lim_{t \to 0^+} \phi'(t) = \phi'(0^+).$$

Some regularity properties that are true for functions on the real-line fail to hold when considering functions of several variables. For instance by using Corollary 3.2 we see that the Fréchet differentiability at 0 of a definable path $\gamma : (-1, 1) \to \mathbf{R}^n$ implies that γ is C^1 around $t = 0$, which is no longer true in higher dimensions, see Fischer [14]. Similarly, a definable mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ which is Fréchet differentiable at 0, is not necessarily Fréchet differentiable in a neighborhood of 0: take for instance $f(x, y) = x^2 + |x| y^2$, where $(x, y) \in \mathbf{R}^2$. In an o-minimal framework, Fréchet differentiability enjoys the following interesting characterization that we now proceed to describe. From now on, let U denote a nonempty open definable neighborhood of 0 in $\mathbf{R}^n$.

Lemma 3.3 (Fréchet differentiability of definable mappings). *Let $\eta : U \to \mathbf{R}^m$ be a definable mapping. The following assertions are equivalent.*

(i) For all C^1 definable curves $p : (-1, 1) \to U$ such that $p(0) = 0$ and $||p(t)|| > 0$ for $t > 0$, we have

$$(3.3) \qquad \lim_{t \to 0^+} \frac{\eta(p(t))}{||p(t)||} = 0.$$

(ii) η is Fréchet differentiable at 0 and $\eta'(0) = 0$.

Proof. It suffices obviously to establish that (i) implies (ii). We argue by contradiction so that there exist a sequence $d_k \neq 0$ converging to 0 and $\epsilon > 0$ such that $||\eta(d_k)|| \geq \epsilon ||d_k||$ for all integers k. Let us consider the definable set

$$A := \{d \in \mathbf{R}^n \setminus \{0\} : ||\eta(d)|| \geq \epsilon \, ||d||\},$$

and let us note that the point 0 belongs to the closure of A. Applying the curve selection lemma we deduce that there exist $\gamma > 0$ and a definable C^1 path $p : (-\gamma, 1) \to \mathbf{R}^m$ such that $p((0, 1)) \subset A$ and $p(0) = 0$, which contradicts the assumption (i). $\qquad \square$

Let us note that *every definable locally Lipschitz mapping F admits directional derivatives*. This follows simply from the monotonicity lemma applied to each bounded curve of the form

$$(0, +\infty) \ni t \mapsto \frac{F(x + td) - F(x)}{t},$$

where $x \in U, d \in \mathbf{R}^n$. Locally Lipschitz, directionally differentiable functions are called *Bouligand differentiable* [13, Section 3.1], and have the following well-known property [25]. We include the brief proof for future reference.

Proposition 3.4 (conical approximation). *If $F : U \to \mathbf{R}^m$ is a Bouligand differentiable mapping (so in particular if F is a definable locally Lipschitz mapping), then for all $x \in U$,*

$$||F(x + d) - F(x) - F'(x; d)|| = o_x(||d||).$$

Proof. With no loss of generality we assume that $x = 0$. Let us introduce the mapping $\eta(d) = F(d) - F(0) - F'(0, d), d \in U$. Using the Lipschitz property of F we have for all d, e in a neighborhood of 0 in $\mathbf{R}^n$,

$$(3.4) \qquad ||F'(0; d) - F'(0; e)|| = || \lim_{t \to 0^+} \frac{F(td) - F(te)}{t} || \leq L ||d - e||$$

where $L > 0$ is a Lipschitz constant of F around 0. Since F is locally Lipschitz, we deduce from (3.4) that η is also locally Lipschitz continuous. On the other hand the definition of η implies that for all $d \in U$, $\eta(td)/t \to 0 = \eta'(0; d)$ as $t \downarrow 0$. Hence η is Gâteaux differentiable with $\eta'_G(0) = 0$. Since η is a locally Lipschitz function in finite dimensions, it is also Fréchet differentiable. Thus the conclusion follows. $\qquad \square$

7

Remark 3.5. The special case of semialgebraic (respectively, subanalytic) locally Lipschitz continuous mappings is of particular interest. In this case, the function $t \longmapsto o_x(t)$ of Proposition 3.4 is also semialgebraic (respectively, subanalytic) and admits a Puiseux development of the form

$$o_x(t) = \sum_{i=p}^{+\infty} a_i \, t^{i/q}, \quad \text{for all } t \in (0, \epsilon),$$

for some $\epsilon > 0$ and integers $p > q > 0$ with $a_p \neq 0$ (recall that $\lim_{t \to 0^+} t^{-1} o_x(t) = 0$). We deduce that

$$F(x + d) - F(x) - F'(x; d) = O_x(||d||^{1+\gamma}),$$

where $\gamma = \frac{p}{q} - 1 > 0$. In other words there exist a positive constant c and $\epsilon > 0$ such that

$$||F(x + d) - F(x) - F'(x; d)|| \leq c \, ||d||^{1+\gamma}$$

for all d, $||d|| \leq \epsilon$. Similar results could be derived for any Lipschitz continuous mapping definable in a "polynomially bounded" o-minimal structure over $\mathbf{R}$ [11, p. 510 and Property 4.12].

One could wonder if definable Gâteaux differentiable mappings are automatically Fréchet differentiable. The classical example of the (definable) function $f : \mathbf{R}^2 \to \mathbf{R}$ with

$$f(x, y) = \begin{cases} \frac{x \, y^3}{x^2 + y^4}, & \text{if } (x, y) \neq (0, 0) \\ \\ 0, & \text{if } (x, y) \neq (0, 0) \end{cases}$$

reveals that this is false in general.

3.2   Semismoothness results

Theorem 3.6. *Any locally Lipschitz tame (resp. definable) mapping $F : U \to \mathbf{R}^m$ ($U \subset \mathbf{R}^m$) is semismooth.*

Proof. Our aim is to show that for all points x in U the mapping

$$d \mapsto \eta(d) := F'(x + d; d) - F'(x; d)$$

is Fréchet differentiable at 0 with $\eta'(0) = 0$. With no loss of generality we assume that $x = 0$. Since the problem is of local nature with F being continuous we can also assume that F is definable (see Remark 2.3). In view of Lemma 3.3 it suffices therefore to prove that

$$\lim_{t \to 0^+} \frac{F'(p(t); p(t)) - F'(0; p(t))}{||p(t)||} = 0$$

for any C^1 definable curve $p : (-1, 1) \to \mathbf{R}^m$ such that $p(0) = 0$ and $p(t) \in U \setminus \{0\}$ for all $t \in (0, 1)$. Let $p : (-1, 1) \to U$ be a definable C^1 path such that $p(0) = 0$, $p(t) \neq 0$ if $t > 0$. Since the curve

$$(0, 1) \ni t \mapsto \frac{p(t)}{||p(t)||}$$

8

is definable, the monotonicity lemma ensures its convergence (as t goes to 0) to some vector u in the unit sphere of $\mathbf{R}^m$. Hence there exists a continuous definable curve $\theta : (0,1) \to \mathbf{R}^m$ such that

$$p(t) = ||p(t)||(u + \theta(t)),$$

with $\theta(t) \to 0$ as $t \downarrow 0$. Let us set $r(t) = ||p(t)||$ for all $t \in [0,1)$. The monotonicity lemma applied to each coordinate of $p(t)$ and the fact that $p(t) \neq 0$ yield that for $t > 0$ small we have $r'(t) \neq 0$. For $t > 0$ sufficiently small the Lipschitz property of F yields

$$\frac{||F(p(t)) - F(r(t)u)||}{r(t)} \leq L\,||\theta(t)||,$$

which tends to 0 as $t \searrow 0^+$. We thus obtain

$$F'(0; u) = \lim_{t \downarrow 0} \frac{F(r(t)u) - F(0)}{r(t)} = \lim_{t \downarrow 0} \frac{F(p(t)) - F(0)}{r(t)}.$$

Setting $q(t) = F(p(t))$ we deduce

$$F'(0; u) = \lim_{t \downarrow 0} \frac{q(t) - q(0)}{r(t)},$$

and applying Lemma 3.1(ii) for the definable function $r(t)$ and for each coordinate of the definable function $t \mapsto q(t) - q(0)$ we infer that

$$(3.5) \qquad\qquad F'(0; u) = \lim_{t \downarrow 0} \frac{q'(t)}{r'(t)}.$$

Using the chain rule and the monotonicity lemma we have for $t > 0$ small

$$q'(t^+) = q'(t) = F'(p(t); p'(t)),$$

which combined with (3.5) yields

$$(3.6) \qquad\qquad \lim_{t \to 0^+} F'(p(t)\,;\, \frac{p'(t)}{r'(t)}) = F'(0;\, u).$$

Thus the curve

$$(3.7) \qquad\qquad t \longmapsto \varphi(t) := ||\, F'(p(t); \frac{p'(t)}{r'(t)}) - F'(p(t); u)||$$

is well defined for $t > 0$ sufficiently small. Using (3.4) (see the proof of Proposition 3.4) we have

$$\varphi(t) = ||\, F'(p(t); u + \frac{[r(t)\theta(t)]'}{r'(t)}) - F'(p(t), u)\,|| \leq L\,||\, \frac{[r(t)\theta(t)]'}{r'(t)}\,||.$$

9

Note that for each $i = 1, \ldots, m$ we have $r(t)\theta_i(t)/r(t) = \theta_i(t) \to 0$ as $t \downarrow 0$. In view of Lemma 3.1(ii) we deduce that $r'(t)^{-1}[r(t)\theta_i(t)]' \to 0$ as $t \downarrow 0$, $i = 1, \ldots, m$, thus

$$\lim_{t \to 0^+} \varphi(t) = 0. \tag{3.8}$$

Combining (3.6) with (3.7) and (3.8) we deduce

$$\lim_{t \to 0^+} F'(p(t); u) = F'(0; u). \tag{3.9}$$

On the other hand, since F is Lipschitz around 0 we obtain for $t > 0$ sufficiently small (see (3.4))

$$||F'(p(t); p(t)) - F'(p(t); r(t)u)|| \le L\, r(t)\, ||\theta(t)||, \tag{3.10}$$

and

$$||F'(0; p(t)) - F'(0; r(t)u)|| \le L\, r(t)\, ||\theta(t)||. \tag{3.11}$$

Combining (3.10), (3.11) and using the triangle inequality we obtain

$$||F'(p(t); p(t)) - F'(0, p(t))|| \le ||F'(p(t); p(t)) - F'(p(t); r(t)u)||$$
$$+ ||F'(p(t); r(t)u) - F'(0; r(t)u)|| + ||F'(0; r(t)u) - F'(0; p(t))||$$
$$\le r(t)\, (2L||\theta(t)|| + ||F'(p(t); u) - F'(0; u)||)\,.$$

This completes the proof. $\qquad\square$

It is worth pointing out that semismooth functions need not be tame. For example, the function $f(x) = x^3 \sin(1/x)$ (with $f(0) = 0$) is continuously differentiable, so certainly semismooth, but cannot be definable, since its zero set is not locally finite, contradicting property (iv) of Definition 2.1.

Remark 3.7. Assume that F is (globally) subanalytic and Lipschitz continuous. In an analogous way as for the conic approximation result, the Puiseux lemma provides additional information. For x fixed, we have indeed

$$F'(x + d; d) - F'(x; d) = O_x(||d||^{1+\gamma}) \tag{3.12}$$

where γ is a positive rational number. Using the terminology of [22, 26] one can assert that semialgebraic or subanalytic Lipschitz continuous mappings are γ-order semismooth or γ-semismooth with $\gamma > 0$.

4 An illustration: convergence rate of the Newton method

As pointed out in the introduction, the Newton method can be run successfully for solving nonlinear equations involving semismooth data. In general, under rather mild assumption the convergence

of the method is superlinear [13]. Let us recall (see [7, page 14], for example) that a given algorithm $\{x_k\}_{k \geq 1}$ is said to converge *linearly* (respectively, *superlinearly*) to x^* if the quotient

$$q_k := \frac{||x_{k+1} - x^*||}{||x_k - x^*||}$$

satisfies $\limsup\limits_{k \to \infty} q_k < 1$ (respectively, $\lim\limits_{k \to \infty} q_k = 0$).

As an illustration of our main results, we prove under mild assumptions that the Newton method applied to a subanalytic locally Lipschitz mapping generates a sequence x^k that converges superlinearly to x^* and satisfies

$$\limsup\limits_{k \to \infty} \frac{||x_{k+1} - x^*||}{||x_k - x^*||^{1+\gamma}} < +\infty,$$

where $\gamma > 0$ (see Theorem 4.3).

Definition 4.1. *Let $F : \mathbf{R}^n \to \mathbf{R}^m$ be a locally Lipschitz continuous function.*
(i) The limiting Jacobian of F at $x \in \mathbf{R}^n$ is defined as

$$\partial F(x) = \{A \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m) : \exists u^k \in D, F'(u^k) \to A, k \to +\infty\}.$$

(ii) The Clarke Jacobian of F at $x \in \mathbf{R}^n$ is defined as (see [8, p. 70])

$$\partial^\circ F(x) = \overline{co}\, \partial F(x),$$

where for all $S \subset \mathbf{R}^m$, $\overline{co}\, S$ stands for the closed convex envelope of S.
(As usual, D denotes the points of differentiability of F.)

Remark 4.2. Due to the Lipschitz property of F, the Clarke Jacobian of F is a nonempty compact convex set [13, Proposition 7.1.4], so in particular we have $\partial^\circ F(x) = co\, \partial F(x)$, for all $x \in U$.

In the remainder, we say that x is a *regular* point of F if each $A \in \partial^\circ F(x)$ has a maximal rank, that is, equal to $\min\{n, m\}$. Note that the upper semicontinuity of the multivalued mapping $\partial^\circ F$ implies that the set of regular points of F is an open subset of $\mathbf{R}^n$.

When $n = m$, an exact Newton algorithm for solving the nonsmooth nonlinear equation $F(x) = 0$ can be devised as follows:

Nonsmooth Newton algorithm
    Step 1 Choose a regular point $x^0 \in \mathbf{R}^n$.
    Step 2 If $F(x^k) = 0$ then stop.
    Step 3 Take $\Delta(x^k)$ in $\partial^\circ F(x^k)$, compute x^{k+1} via

$$F(x^k) + \Delta(x^k)(x^{k+1} - x^k) = 0,$$

$k \leftarrow k + 1$ and go to Step 2.

An important issue of the above algorithm is obviously the computation of x^{k+1} in Step 3. This is in general a delicate matter tightly linked to the convergence of the algorithm (see [13] and references therein). We will not tackle this problem here.

The following result is a stronger version of [13, Theorem 7.5.3] in the case of subanalytic mappings.

Theorem 4.3. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be a locally Lipschitz (globally) subanalytic function and a regular point $x^* \in \mathbf{R}^n$ such that $F(x^*) = 0$. Then there exists $\delta > 0$ such that for all $x^0 \in B(x^*, \delta)$ the nonsmooth Newton algorithm is well defined and generates a sequence $\{x^k\}_{k \in \mathbf{N}}$ which converges to x^*. Moreover there exists a rational number $\gamma > 0$ such that*

$$(4.13) \qquad ||x^{k+1} - x^*|| = O(\exp[-(1+\gamma)^k]),$$

which implies in particular that

$$||x^{k+1} - x^*|| \leq \frac{c}{2^{(1+\gamma)^k}}$$

for some positive constant c.

The proof relies on the following lemma:

Lemma 4.4. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be a locally Lipschitz subanalytic function and $x \in \mathbf{R}^n$. Then there exists a positive rational number γ such that*

$$(4.14) \qquad ||F(y) - F(x) - \Delta(y)(y-x)|| = O_x(||y-x||^{1+\gamma}),$$

where $\Delta(y)$ is any element of $\partial^\circ F(y)$.

Proof. Using Proposition 3.4 and Remark 3.5 we have

$$(4.15) \qquad F(y) - F(x) - F'(x; y-x) = O_x(||y-x||^{1+\gamma_1})$$

where γ_1 is a positive rational number. To obtain (4.14), it suffices therefore to establish that

$$||F'(x; d) - \Delta(x+d)d|| = O_x(||d||^{1+\gamma_2}), \text{ with } \gamma_2 > 0 \text{ in } \mathbb{Q}.$$

The constant $\gamma_2 = \gamma$ posited in Remark 3.7 does the job: indeed, let us fix $d \in \mathbf{R}^n$. By definition of the Clarke Jacobian for a Lipschitz function and the Carathéodory theorem we obtain a finite sequence $\lambda_1, \ldots, \lambda_{n^2+1} \geq 0$ with $\sum_1^{n^2+1} \lambda_i = 1$ and $n^2 + 1$ sequences $\{d^{i,k}\}_{k \in \mathbf{N}}$ with $d^{i,k} \to d$ as $k \to +\infty$ such that $x + d^{i,k} \in D$ and

$$(4.16) \qquad \Delta(x+d) = \sum_{i=1}^{n^2+1} \lambda_i \lim_{k \to +\infty} F'(x + d^{i,k}).$$

12

In view of Remark 3.7, we get

$$|| \sum_{i=1}^{n^2+1} \lambda_i F'(x; d^{i,k}) - \sum_{i=1}^{n^2+1} \lambda_i F'(x + d^{i,k}, d)||$$

$$\leq \sum_{i=1}^{n^2+1} \lambda_i ||F'(x; d^{i,k}) - F'(x + d^{i,k}; d^{i,k}) + F'(x + d^{i,k}; d - d^{i,k})||$$

$$\leq \sum_{i=1}^{n^2+1} \lambda_i \left(||F'(x; d^{i,k}) - F'(x + d^{i,k}; d^{i,k})|| + ||F'(x + d^{i,k}; d - d^{i,k})|| \right)$$

$$\leq O_x(||d||^{1+\gamma_2}) + L \sum_{i=1}^{n^2+1} ||d - d^{i,k}||$$

where $L > 0$ is the Lipschitz constant of f around $x + d$. Taking the limit $k \to \infty$ we obtain the asserted result. □

The proof of Theorem 4.3 is now standard. The above result shows, in the terminology of [13, Definition 7.5.13], that the multifunction $\partial^\circ F$ is a "$(1 + \gamma)$-order linear Newton approximation" of the function F at the point x. Superlinear convergence now follows in an analogous fashion to [13, Theorem 7.5.15]: see the discussion in [13, p. 696].

References

[1] BOLTE, J., DANIILIDIS, A. & LEWIS, A., The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, *SIAM J. Optim.* (to appear).

[2] BOLTE, J., DANIILIDIS, A. & LEWIS, A., A nonsmooth Morse-Sard theorem for subanalytic functions, *J. Math. Anal. Appl.* **321** (2006), 729–740.

[3] BOLTE, J., DANIILIDIS, A., LEWIS, A., & SHIOTA, M., Clarke critical values of subanalytic Lipschitz continuous functions, *Ann. Polon. Math.* **87** (2005), 13–25 (memorial issue for S. Łojasiewicz).

[4] BENEDETTI, R. & RISLER, J-J., *Real Algebraic and Semi-Algebraic Sets*, Hermann, Paris, 1990.

[5] BIERSTONE, E. & MILMAN, P., Semianalytic and subanalytic sets, *IHES Publ. Math.* **67** (1988), 5–42.

[6] BOCHNAK, J., COSTE, M. & ROY, M.-F., *Real Algebraic Geometry*, Springer, Berlin, 1998.

[7] BONNANS, F., GILBERT, J.-C., LEMARÉCHAL, C. & SAGASTIZÁBAL, C., *Numerical optimization. Theoretical and applied aspects*, Mathematics & Applications **27**, Springer-Verlag, Berlin, 2002.

[8] Clarke, F., *Optimization and nonsmooth analysis,* (second edition), Classics in Applied Mathematics **5**, SIAM, Philadelphia, 1990.

[9] Coste, M., *An Introduction to o-minimal Geometry*, RAAG Notes, 81 pages, Institut de Recherche Mathématiques de Rennes, November 1999.

[10] Dedieu, J.-P., Penalty functions in subanalytic optimization, *Optimization* **26** (1992), 27–32.

[11] van den Dries, L. & Miller, C., Geometric categories and o-minimal structures, *Duke Math. J.* **84** (1996), 497–540.

[12] Facchinei, F., Fischer, A. & Kanzow, C., Inexact Newton methods for semismooth equations with applications to variational inequality problems, in: DiPillo, G. & Giannessi, F., editors, *Nonlinear Optimization and Applications*, pp. 125–139, Plenum Press, New York, 1996.

[13] Facchinei, F. & Pang, J.-S., *Finite-Dimensional Variational Inequalities and Complementarity Problems,* Volumes I and II, Springer, New York, 2003.

[14] Fischer, A., Peano-differentiable functions in o-minimal structures, PhD Dissertation, 2005.

[15] Kummer, B., Newton's method for non-differentiable functions, in: Guddat, J. et al., editors, *Advances in Mathematical Optimization*, pp. 114–125, Akademie-Verlag, Berlin, 1988.

[16] Kurdyka, K., On gradients of functions definable in o-minimal structures, *Ann. Inst. Fourier* **48** (1998), 769–783.

[17] Kuntz, L. & Scholtes, S., Structural analysis of nonsmooth mappings, inverse functions, and metric projections, *J. Math. Anal. Appl.* **188** (1994), 346–386.

[18] Łojasiewicz, S., Une propriété topologique des sous-ensembles analytiques réels, in: *Les Équations aux Dérivées Partielles*, pp. 87–89, Éditions du Centre National de la Recherche Scientifique, Paris, 1963.

[19] Luo, Z. & Pang, J.-S., Error bounds for analytic systems and their applications, *Math. Programming* **67** (1994), 1–28.

[20] Mifflin, R., Semismooth and semiconvex functions in constrained optimization, *SIAM J. Control Optim.* **15** (1997), 957–972.

[21] Pang, J.S. & Stewart, D.E., Solution dependence on initial conditions in differential variational inequalities, *Math. Programming* (to appear).

[22] Qi, L. & Sun, J., A nonsmooth version of Newton's method, *Math. Programming* **58** (1993), 353–367.

[23] Robinson, S.M., Newton's method for a class of nonsmooth functions, Industrial Engineering Working Paper, University of Wisconsin (1988).

[24] Robinson, S.M., Newton's method for a class of nonsmooth functions, *Set-Valued Analysis* **2** (1994), 291–305.

[25] Shapiro, A., On concepts of directional differentiability, *J. Optim. Theory Appl.* **66** (1990), 477–487.

[26] Sun, D. & Sun, J., Löwner's operator and spectral functions in Euclidean Jordan algebras, *Math. Oper. Res.* (to appear).

Jérôme BOLTE

Equipe Combinatoire et Optimisation (UMR 7090), Case 189
Université Pierre et Marie Curie
4 Place Jussieu, 75252 Paris Cedex 05, France.
E-mail: bolte@math.jussieu.fr ; http://www.ecp6.jussieu.fr/pageperso/bolte/

Aris DANIILIDIS

Departament de Matemàtiques, C1/320
Universitat Autònoma de Barcelona
E-08193 Bellaterra (Cerdanyola del Vallès), Spain.
E-mail: arisd@mat.uab.es ; http://mat.uab.es/~arisd

Adrian LEWIS

School of ORIE
Cornell University
Ithaca, NY 14853, USA.
E-mail: aslewis@orie.cornell.edu ; http://www.orie.cornell.edu/~aslewis