On the gradient problem of C. E. Weil

Zoltán Buczolich*, Department of Analysis, Eötvös Loránd University, Pázmány Péter Sétány 1/c, 1117 Budapest, Hungary
email: buczo@cs.elte.hu
www.cs.elte.hu/∼buczo

September 3, 2003

1 Introduction

It is well-known that if $f : (a, b) \to \mathbb{R}$ is everywhere differentiable on the open interval (a, b) then its derivative, f' is Darboux and Baire one. It is less widely known that derivatives have one more property, the so called Denjoy–Clarkson property see [11], [10]. This property states that if (α, β) is any open interval then its inverse image by the derivative, that is, $(f')^{-1}(\alpha, \beta)$ is either empty, or of positive one dimensional Lebesgue measure. This means that if inverse images of open intervals are nonempty then they are "heavy". Of course, if the function is continuously differentiable then this inverse image is a nonempty open set and hence heavy. So the Denjoy–Clarkson property is really interesting for functions with discontinuous derivatives. In [23] it was also shown that the Denjoy–Clarkson property is posessed by every approximate derivative and kth Peano derivative. Complete characterization of derivatives is still not available see on this topic A. Bruckner's survey article [2].

The gradient problem of C. E. Weil is about the multidimensional version of the Denjoy–Clarkson property.

Assume $n \geq 2$, $G \subset \mathbb{R}^n$ is an open set and $f : G \to \mathbb{R}$ is an everywhere differentiable function. Then $\nabla f = (\partial_1 f, ..., \partial_n f)$ maps G into $\mathbb{R}^n$. Assume $\Omega \subset \mathbb{R}^n$ is open. Is it true that $(\nabla f)^{-1}(\Omega) = \{\mathbf{p} \in G : \nabla f(\mathbf{p}) \in \Omega\}$ is either empty, or of positive n-dimensional Lebesgue measure? For continously differentiable functions, like in the one-dimensional case, the answer is obviously positive.

One other, equivalent and interesting way to look at the gradient problem is the following special case. Denote by B_1 the open unit ball in the n-dimensional Euclidean space. Does there exist a differentiable function $f : \mathbb{R} \to \mathbb{R}$ such that

* *2000 Mathematics Subject Classification:* Primary 26B05; Secondary 28A75, 37E99.
Keywords: gradient, Denjoy-Clarkson property, Lebesgue measure

its gradient equals the null vector at the origin and (Lebesgue) almost everywhere the norm of the gradient is bigger than one. This means that $(\nabla f)^{-1}(B_1)$ is nonempty, but of zero Lebesgue measure. So our differentiable function due to vanishing gradient is not changing much in a neighborhood of the origin, but almost everywhere, "with probability one" it changes rapidly.

The gradient problem was one of the well-known and famous unsolved problems in Real Analysis. It was around since the paper [23] appeared in the 1960s. I have learned it in 1987 from Clifford Weil. This was my first visit to the US and I quite clearly remember as after being introduced to him, standing in front of the Alamo in San Antonio, he suggested me to try to solve this problem. I also received the warning that this was a difficult problem... So I started to work on it. Of course, I was not working on this problem continuously but when I learned something new, I returned to it, worked for a few months and when I felt that I had no breakthrough I gave it up for some time... In case I obtained a partial result which was sufficiently interesting I published it and/or gave talks about it at different conferences. This was going on till the end of the Summer of 2002 when I finally managed to solve it. In 1990 at the Fourteenth Summer Symposium on Real Analysis the gradient problem was advertised and appeared in print in [25]. For most of the time when I worked on the gradient problem I tried to prove a theorem stating an n-dimensional version of the Denjoy-Clarkson property but there was something always missing... A few years ago I even made a bet with L. Zajíček about the way the solution goes. In the end I had to revise my view of the problem and started to work seriously on a two-dimensional (counter)example. This lead to the solution, and I lost the bet. There is the usual moral that at a correct trial, both sides should receive a fair hearing. One should have a flexible thinking to be able to look at and analyse a problem from both ends.

In this talk I would like to go through the road leading to the final solution, while the details of the proof can be found in the research article [8].

The answer to the gradient problem is a two-dimensional counterexample. In [8] it is shown that there exists a nonempty open set $G \subset \mathbb{R}^2$, a differentiable function $f : G \to \mathbb{R}$ and an open set $\Omega_1 \subset \mathbb{R}^2$ for which there exists a $\mathbf{p} \in G$ such that $\nabla f(\mathbf{p}) \in \Omega_1$ but for almost every (in the sense of two-dimensional Lebesgue measure, λ_2,) $\mathbf{q} \in G$ the gradient $\nabla f(\mathbf{q})$ is not in Ω_1.

2 The results on the way to the solution of the gradient problem

2.1 The $\mathcal{H}_1$ Denjoy-Clarkson property

It is not too difficult to prove (see [3]) that if one replaces the n-dimensional Lebesgue measure by the one-dimensional Hausdorff measure, $\mathcal{H}_1$, then the Denjoy–Clarkson property holds in the multidimensional case as well:

Theorem 1. Theorem 1 of [3]. *Assume $G \subset \mathbb{R}^n$ is open, $f : G \to \mathbb{R}$ is a differentiable function. Then for any $\Omega \subset \mathbb{R}^n$ open set $(\nabla f)^{-1}(\Omega)$ is either empty, or of positive $\mathcal{H}_1$ measure.*

So inverse images of open sets by the gradient map at least should be heavy in the $\mathcal{H}_1$ sense.

In another research paper, [15], Holický, Malý, Weil, and Zajíček verified the following theorem:

Theorem 2. Theorem 5 of [15]. *Let $G \subset \mathbb{R}^n$ be an open set and let f be a (Fréchet) differentiable function on G. Suppose $\Omega \subset \mathbb{R}^n$ is an open set such that $(\nabla f)^{-1}(\Omega) \neq \emptyset$. Then the following assertions hold.*

- *(i) $(\nabla f)^{-1}(\Omega)$ is porous at none of its points.*
- *(ii) If $H \subset \mathbb{R}^n$ is open and $H \cap (\nabla f)^{-1}(\Omega) \neq \emptyset$, then $L(H \cap (\nabla f)^{-1}(\Omega))$ is of one-dimensional Lebesgue mesure zero for no nonzero linear function $L : \mathbb{R}^n \to \mathbb{R}$. In particular, $H \cap (\nabla f)^{-1}(\Omega)$ is of positive one dimensional Hausdorff measure.*
- *(iii) If $H \subset \mathbb{R}^n$ is open and $H \cap (\nabla f)^{-1}(\Omega) \neq \emptyset$, then $H \cap (\nabla f)^{-1}(\Omega)$ is not σ-porous.*

Observe that in (ii) of the above theorem one can take any one dimensional projection as the linear function showing that any one-dimensional projection of $(\nabla f)^{-1}(\Omega)$ is of positive $\mathcal{H}_1$ measure, when $(\nabla f)^{-1}(\Omega)$ is nonempty.

2.2 The "paradoxical convexity" property

The above results point in the direction that the counterexample function of [8] should be quite complicated. In the next result we show that this function is even more complicated. For a while I believed that such functions do not exist, but they do. I think that this "paradoxical convexity" property of the counterxample functions to the gradient problem was widely ignored by even those people who worked on the gradient problem, though this was the property which finally lead to the convexity construction in [8]. So assume that $G_0 \subset \mathbb{R}^2$ and $f : G_0 \to \mathbb{R}$ is a differentiable counterexample function to the gradient problem. This means that there exists an $\mathbf{x}_0 \in G_0$, a $\mathbf{p}_0 \in \mathbb{R}^2$ and an $\eta_0 > 0$ such that $\nabla f(\mathbf{x}_0) = \mathbf{p}_0$ and $\lambda_2((\nabla f)^{-1}(B(\mathbf{p}_0, \eta_0))) = 0$, where λ_2 denotes the Lebesgue measure in the plane and $B(\mathbf{p}_0, \eta_0)$ is the open ball of radius η_0 centered at $\mathbf{p}_0$. Set $F_0 = cl((\nabla f)^{-1}(B(\mathbf{p}_0, \eta_0/2)))$. Since ∇f is a Baire one function on F_0 it has a point of continuity, $\mathbf{x}_1 \in F_0$. From the definition of F_0 it follows that $\nabla f(\mathbf{x}_1) \in cl(B(\mathbf{p}_0, \eta_0/2))$. Hence, $B(\nabla f(\mathbf{x}_1), \eta_0/2) \subset B(\mathbf{p}_0, \eta_0)$. Choose $\delta > 0$ such that for any $\mathbf{y} \in B(\mathbf{x}_1, \delta) \cap F_0$ we have $||\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{y})|| < \eta_0/4$. Then $B(\mathbf{x}_1, \delta) \cap F_0 \subset (\nabla f)^{-1}(B(\mathbf{p}_0, \eta_0))$ and hence $\lambda_2(B(\mathbf{x}_1, \delta) \cap F_0) = 0$. On the other hand, if we set $\mathcal{R} = \nabla f(B(\mathbf{x}_1, \delta)) \cap B(\mathbf{p}_0, \eta_0/2)$ then $\mathcal{R} \subset B(\nabla f(\mathbf{x}_1), \eta_0/4)$. Hence, $\mathcal{G} = B(\mathbf{p}_0, \eta_0/2) \setminus cl(\mathcal{R})$ is a nonempty open set, and the density of $(\nabla f)^{-1}(B(\mathbf{p}_0, \eta_0/2))$ in F_0 implies that $\mathcal{R}$ is nonempty either. By a suitable

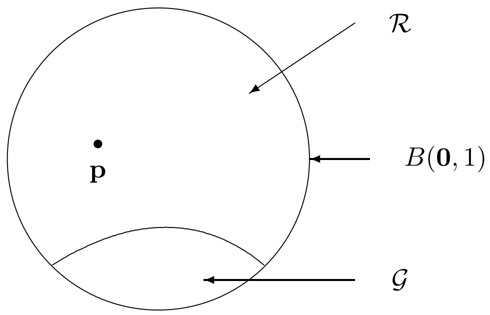


Figure 1: "Paradoxical convexity" in the range

change of variable we can assume that $\mathbf{p}_0 = \mathbf{0}$ and $\eta_0/2 = 1$. We can also take $G = B(\mathbf{x}_1, \delta)$ and work with the restriction of f onto this set. Hence, the conditions of the following theorem are satisfied.

Theorem 3. See [6]. *Assume that f is a differentiable function on $G \subset \mathbb{R}^2$. Set $\overline{\Delta}_1 = cl\{\mathbf{x} \in G : \nabla f(\mathbf{x}) \in B(\mathbf{0},1)\}$. Suppose $\overline{\Delta}_1$ is nonempty and $\lambda_2(\overline{\Delta}_1) = 0$. Put $\mathcal{R} = B(\mathbf{0},1) \cap \nabla f(G)$ and $\mathcal{G} = B(\mathbf{0},1) \setminus cl(\mathcal{R})$. Then $\mathcal{G}$ is a convex open subset of the plane and $\mathcal{G} \neq \emptyset$ implies that for any $\mathbf{p} \in int(cl(\mathcal{R}))$ we have $\mathcal{H}_1(\{\mathbf{y} : \nabla f(\mathbf{y}) = \mathbf{p}\}) > 0$.*

By a suitable linear change of variable and rescaling we can achieve that $\mathcal{G} \neq \emptyset$ and $\mathbf{0} \in \mathcal{R}$. Then $B(\mathbf{0},1)$ contains a half disk such that for any point, $\mathbf{p}$ belonging to this half disk we have $\mathcal{H}_1(\{\mathbf{x} : \nabla f(\mathbf{x}) = \mathbf{p}\}) > 0$. This shows that our differentiable function is very different from the smooth surfaces we got used to. For example if one takes the open upper half sphere $x_3 = f(x_1,x_2) = \sqrt{x_1^2 + x_2^2}$ then for any $\mathbf{p} \in \mathbb{R}^2$ we have that either $(\nabla f)^{-1}(\mathbf{p})$ is empty, or consists of a single point only. So the price we need to pay for nonempty but λ_2-measure zero ∇f preimages of open sets is that we have many points with large preimages in the sense of $\mathcal{H}_1$ measure.

Since $int(cl(\mathcal{R}))$ is uncountable Theorem 3 also implies that $(\nabla f)^{-1}(B(\mathbf{0},1))$ is of non-sigma finite $\mathcal{H}_1$ measure.

Therefore, it is possible that $(\nabla f)^{-1}(B(\mathbf{0},1)))$ is nonempty and of zero λ_2 measure, but it is at least of non-sigma finite $\mathcal{H}_1$ measure. So there is a gap which needs to be explored.

Question 1. Assume that $G \subset \mathbb{R}^n$ is an open set, $f : G \to \mathbb{R}$ is a differentiable function and $\Omega \subset \mathbb{R}^n$ is such an open set that $(\nabla f)^{-1}(\Omega) \neq \emptyset$. Then what can we say about the Hausdorff-dimension of $(\nabla f)^{-1}(\Omega)$.

2.3 Functions with many tangent planes

Now we return to Theorem 3. Assume that the assumptions of this theorem are satisfied and $\mathbf{p} \in int(cl(\mathcal{R}))$. Put $f^{\mathbf{p}}(\mathbf{x}) = f(\mathbf{x}) - \mathbf{p} \cdot \mathbf{x}$, where $\cdot$ denotes the usual dot/scalar product in $\mathbb{R}^2$. We use the following notation for the level sets $f^{\mathbf{p}}_c \stackrel{\text{def}}{=} \{\mathbf{x} : f^{\mathbf{p}}(\mathbf{x}) = c\}$. By scrutinizing the proof of Theorem 3 one can see that for a set of positive λ_1 measure of c's, there exists $\mathbf{x}_c$ on the level set $f^{\mathbf{p}}_c$ such that $\nabla f^{\mathbf{p}}(\mathbf{x}_c) = \mathbf{0}$, that is, $\nabla f(\mathbf{x}_c) = \mathbf{p}$. This means that we can find many points where the tangent plane to the surface given by the graph of f has gradient $\mathbf{p}$.

This turned my attention to differentiable functions with many tangent planes. Results in this direction were published in [7]. The main result of this paper is the following.

Theorem 4. See [7]. *There exists a C^1 function $f : \mathbb{R}^2 \to \mathbb{R}$, a nowhere dense closed set $E \subset [0,1] \times [0,1]$ of zero λ_2 measure, and a nonempty open set $H \subset \mathbb{R}^3$ such that for any $(a,b,c) \in H$ we can find and $(x_0, y_0) \in E$ for which the equation of the tangent plane to the surface $z = f(x,y)$ at the point $(x_0, y_0, f(x_0, y_0))$ is $z = ax + by + c$.*

So we can expect more than differentiability. There are even C^1 functions for which we have the above construction. Observe that H in the parameter space is a three dimensional open set and E is a measure zero two dimensional one so we gain not only measure, but dimension as well, we have a "Peano" map into the parameter space. In dimension one we got used to the fact that C^1 functions are sufficiently smooth. However in higher dimensions it is not true any more. In the theory of partial differential equations there are several results like the Sobolev Lemma ([20] Th. 7.25) where smoothness assumptions and consequences depend on the dimension.

Recall the Morse–Sard theorem [17], [21].

Theorem 5. Theorem 1.3 of [14]. *Let M, N be manifolds of dimension m, n and $f : M \to N$ a C^r map. If $r > \max\{0, m-n\}$ then $f(\Sigma_f)$ has measure zero in N, where Σ_f denotes the critical points of f. The set of regular values of f is residual and therefore dense.*

Hence, if $f : \mathbb{R}^2 \to \mathbb{R}$ is a C^2 function then for any $(a,b) \in \mathbb{R}^2$ the set of critical values $c_{(a,b)}$ of $f(x,y) - ax - by$ is of λ_1 measure zero. Thus, it cannot contain an interval and by Fubini's theorem the set

$$\{(a,b,c) \in \mathbb{R}^3 : \text{ the plane } z = ax + by + c \text{ is tangent to } z = f(x,y)\}$$

is of λ_3 measure zero. Therefore, it has empty interior.

C^1 functions in the plane can exhibit many other pathological properties. One is the famous example of Whitney [24] which gives a C^1 function defined on $\mathbb{R}^2$ and a continuous nondegenerate arc γ such that $\nabla f = \mathbf{0}$ on γ but f is not constant along γ. If we relax the assumption to differentiability Körner [16] constructed a non constant differentiable function defined on the plane for

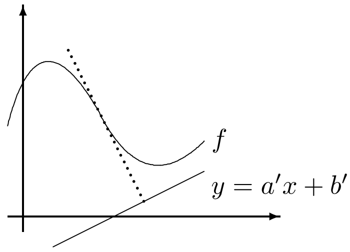


Figure 2: Projection π' onto $y = a'x + b'$

which any two points in $\mathbb{R}^2$ can be connected by a continuous curve, γ such that $\nabla f = \mathbf{0}$ everywhere on γ but at endpoints.

It is also worth to say a few words about the one-dimensional version of Theorem 4. So assume that $f : \mathbb{R} \to \mathbb{R}$ is a differentiable function. If $y = ax + b$ is the tangent of f at $(x_0, f(x_0))$ then we set $S_1(x_0) = (a, b)$. Is it possible that $S_1(\mathbb{R})$ has nonempty interior?

In dimension one C^1 functions are sufficiently smooth for the Morse–Sard theorem. Thus, the answer is no. For any $a \in \mathbb{R}$ the critical values of $f(x) - ax$ are of measure zero. Therefore, for a fixed a we have only λ_1-measure zero b's in the range of S_1.

One might expect that for differentiable fuctions maybe the answer is positive. But it is still not. This time another one of my favorite theorems in Real Analysis can be used. Part of the Denjoy-Young-Saks theorem ([22], Chap. IX, (3.7) Theorem, p. 267) implies the following. Assume $y = a'x + b'$ denotes a given line in $\mathbb{R}^2$ and π' denotes the orthogonal projection onto this line. Assume $f : \mathbb{R} \to \mathbb{R}$ is an arbitrary function and E' denotes the π' image of those points of the graph of f where the tangent to f is perpendicular to $y = a'x + b'$, see Figure 2. Then $\lambda_1(E') = 0$. This again implies that $\lambda_2(S_1(\mathbb{R})) = 0$ even for differentiable functions.

2.4 The level set structure

Trying to answer the gradient problem in dimension two I also got interested in the level set structure of functions defined on the plane. (Anyone who likes hiking encounters similar type level sets on hiking and other geographic maps.) So assume $f : \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function. At critical points the level sets can look very strange. So it seems to be natural to assume that the gradient is nonvanishing.

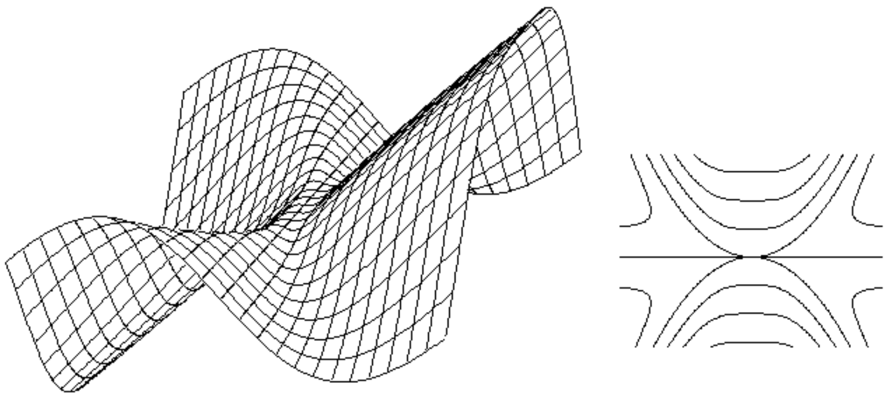


Figure 3: $f(x, y)$ with bifurcation

Assuming this, is it true that the level sets, $\{(x, y) : f(x, y) = c\}$ consist of differentiable arcs? The answer is no, see [4]. If one takes

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

then the following theorem shows that the level set can contain bifurcation points. On the left side of Figure 3 this function is plotted. For better viewing on the 3D image the y axis is pointing from left to right. On the right side of this picture the level set structure of this function is plotted in the plane, the axes are pointing in the usual directions. Properties of this function are given in the following theorem.

Theorem 6. Theorem 1 of [4]. *There exists a differentiable function* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *such that* $\nabla f(x, y) \neq \mathbf{0}$ *for all* $(x, y) \in \mathbb{R}^2$ *and* $S_0 = \{(x, y) : f(x, y) = 0\} = \{(x, y) : |y| = x^2, \text{ or } y = 0\}$, *and hence one cannot find a neighborhood* G *of* $(0, 0)$ *such that* $U \cap S_0$ *is homeomorphic to an open interval.*

However, when I was thinking about the gradient problem Theorem 6 was not a serious obstruction since by using the Baire one property of the gradient and adding a suitable linear function, like in the argument preceding Theorem 3, one can work with differentiable functions such that the gradient is bounded away from zero. This extra assumption is sufficient to get rid of bifurcation points and the following theorem holds.

Theorem 7. Theorem 2 of [4]. *Suppose that* $G \subset \mathbb{R}^2$ *is open,* $f : G \to \mathbb{R}$ *is differentiable and* $||\nabla f(x, y)|| > r > 0$ *for all* $(x, y) \in G$. *If* $c = f(x_0, y_0)$

for an $(x_0, y_0) \in G$ *then there exists a neighborhood* G_0 *of* (x_0, y_0) *such that* $S_c = G_0 \cap \{(x, y) : f(x, y) = c\}$ *is homeomorphic to an open interval and* S_c *has a tangent at each of its points.*

Assuming sufficient smoothness one might obtain a similar result by using the implicit function theorem. However, even for a generalized version of the two dimensional implicit function theorem (see [9]), an assumption is used that the partial derivative of $f(x, y)$ with respect to y is non-vanishing plus some other assumptions are also necessary about the local boundedness of the (generalized) partial derivatives. It is intersting to observe that for the inverse function theorem differentiability is sufficient. In [19] it is proved that if the mapping $f : G \to \mathbb{R}^n$ is differentiable on an open set $G \subset \mathbb{R}^n$ and $\det f'(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in G$ then f is a local diffeomorphism. This theorem is standard if f is sufficiently smooth, but when only differentianility is assumed then there is nontrivial topology, namely, Degree Theory in the background.

The direction of research coming from Theorems 6 and 7 was continued by M. Elekes, who gave a talk on his results at this Symposium. The results in [12] show that if a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ has nonvanishing gradient then locally the level set $\{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) = c\}$ is homeomorphic either to an open interval or (at bifurcation points) to the union of finitely many line segments passing through a point. The bifurcation points form a discrete set.

2.5 Partials and approximate continuity of derivatives

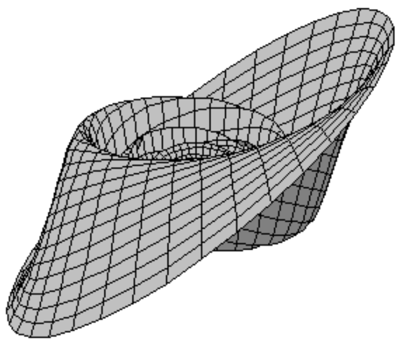


Figure 4: $f(r, \phi)$

Definition 1. The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is approximate continuous at $\mathbf{x}$ if there exists an $E \subset \mathbb{R}^n$ such that $\mathbf{x}$ is a Lebesgue density point of E and

$\lim_{\mathbf{y}\to\mathbf{x};\mathbf{y}\in E} f(\mathbf{y}) = f(\mathbf{x})$. The set of approximate continuity points of f will be denoted by A_f.

In [18] G. Petruska verified that if $f : \mathbb{R} \to \mathbb{R}$ is the derivative of a function F, then f takes each of its values on its approximate continuity points, that is, $f(\mathbb{R}) = f(A_f)$. This implies the one-dimensional Denjoy–Clarkson property. Hence I was interested in higher dimensional generalizations of this result. In [5] I proved the following results. If one considers partial derivatives then the following theorem holds.

Theorem 8. Theorem 1 of [5]. *Assume $F : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $i \in \{1, ..., m\}$. Then $f = \partial_i F$ takes each of its values on A_f.*

This implies that partials of differentiable functions have the Denjoy–Clarkson property. The assumption that F is differentiable is needed in the above theorem since the existence of partial derivatives is not sufficient as the following theorem shows.

Theorem 9. Theorem 2 of [5]. *There exists a continuous function $F : \mathbb{R}^2 \to \mathbb{R}$ such that $f = \partial F/\partial x_1$ exists everywhere and f does not take each of its values on A_f.*

For the gradient of differentiable functions we have a similar theorem.

Theorem 10. Theorem 3 of [5]. *There exists a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ such that ∇f does not take each of its values on $A_{\nabla f}$.*

The function of the above theorem can be defined in polar coordinates as

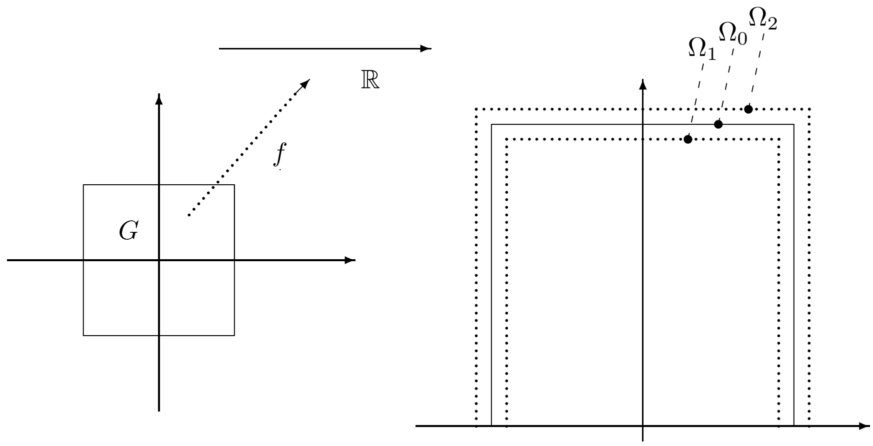
$$f(r, \phi) = \begin{cases} r^2 \sin\left(\phi + \frac{1}{r^2}\right), & \text{if} \quad r \neq (0, 0); \\ 0, & \text{if} \quad r = 0. \end{cases}$$

This function can be seen on Figure 4. The gradient of this function vanishes only at $\mathbf{0}$ and this value is not in $A_{\nabla f}$.

Question 2. Assume $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function. We call a point $\mathbf{y} \in \mathbb{R}^n$ a regular value of ∇f if there is an $\mathbf{x} \in A_{\nabla f}$ such that $\nabla F(\mathbf{x}) = \mathbf{y}$. Denote the set of regular values by $REG(\nabla F)$. The question in [5] about the density of $REG(\nabla F)$ in the range of ∇F receives a negative answer by the negative answer to the gradient problem. On the other hand our question about the characterization of $REG(\nabla F)$ remains open.

3 Outline of the construction of the counterexample to the gradient problem

We put $G = (-1, 1) \times (-1, 1)$, $\Omega_0 = [-\frac{1}{2}, \frac{1}{2}] \times [0, 2]$, $\Omega_1 = (-0.49, 0.49) \times (0, 1.99)$, and $\Omega_2 = [-0.51, 0.51] \times [0, 2.01]$. On the right side of Figure 5 the boundary of Ω_0 is marked by solid line, while the boundaries of Ω_1 and Ω_2 are marked by dotted lines.



Domain and range of f Range of ∇f, the gradient space

Figure 5: G and the sets Ω in the gradient space

Our answer to the gradient question is the following theorem. The details of the proof can be found in the research article [8], here we discuss only the main ideas of the construction.

Theorem 11. Theorem 1 of [8]. *There exists a differentiable function* $f : G \to \mathbb{R}$ *such that*
$\nabla f(0,0) = (0,1)$ *and* $\nabla f(\mathbf{p}) \notin \Omega_1$ *for almost every* $\mathbf{p} \in G$.

To illustrate this theorem on the left side of Figure 5 the domain of f is shown and an arrow from the domain into a copy of $\mathbb{R}$ shows that f maps into $\mathbb{R}$, on the other hand on the right side of this figure one can see the range of the gradient map, the "gradient space".

Scrutinizing the proof of Theorem 11 one can also see that f is a Lipschitz function and ∇f stays in the upper half-plane.

If one relaxes the assumption in the above theorem to λ_2 almost everywhere differentiability of f it is very easy to construct some Lipschitz functions. On Figure 6 one of the favorite "constructions" of people thinking about the gradient problem can be seen, a sheet of folded paper. This corresponds to a surface which is not changing much globally, although locally with a sufficiently large number of zig-zags one can obtain almost everywhere large values of the gradient. These large gradient values are illustrated by the arrows on the figure. The surface is not differentiable on the folding edges, and elsewhere the gradient takes only two values. If we want to obtain a differentiable function we need to sand down the folding edges which yields small absolute value gradient values but we can still keep the gradient values within a line segment in the plane. We have some freedom of choosing the number of zig-zags and choosing the direction of the folding, that is, choosing the direction of the segment containing the values of the gradient. Slightly modified versions of these folded paper sheet surfaces will be converted into perturbation functions denoted by $\phi_{B_{n,k}}$. The function of Theorem 10 shown on Figure 4 corresponds to a paper folding along a "spiral" and is not far from this construction.

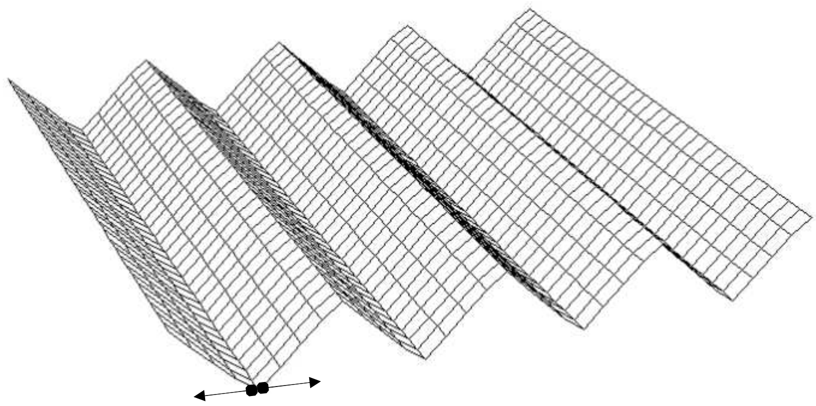


Figure 6: Folded paper

Next we outline the proof of Theorem 11. We start with a function $h_{-1}(x,y) = y$. Then $\nabla h_{-1} = (0,1)$ everywhere on G. We choose a sequence of functions h_n so that $f(x,y) \stackrel{\text{def}}{=} \sum_{n=-1}^{\infty} h_n(x,y)$ satisfies Theorem 11. Each function h_{n+1} can be regarded as a perturbation of the previous partial sum $f_n = \sum_{k=-1}^{n} h_k$.

We want to push $\nabla f(\mathbf{p})$ outside of Ω_1 for almost every $\mathbf{p} \in G$.

In the construction we will have a nested sequence of open sets G_n such that for almost every $\mathbf{p}$ outside of G_n we will have $\nabla f(\mathbf{p}) \notin \Omega_1$. We use a "stopping time argument" by not perturbing the function at almost every point, $\mathbf{p}$ any more, once $\nabla f(\mathbf{p}) \notin \Omega_0$. For these points $\sum_{n=-1}^{\infty} h_n(x,y) = \sum_{n=-1}^{n_0} h_n(x,y)$ for a suitable n_0. We show that $\lambda_2(G_n) \to 0$. The main difficulty is to show that f is differentiable at those points $\mathbf{p} \in \cap_{n=0}^{\infty} G_n$ which are subject to infinitely many perturbations.

At step n, $(n = 0, 1, ...)$ we are choosing some disjoint open squares $B_{n,k}$, called *perturbation blocks*. The sides of these squares are not necessary parallel with the coordinate axes and the choice of the proper direction of these squares is very delicate.

Here is some notation, the open square $B_{n,k}$ is centered at $\mathbf{o}_{B_{n,k}}$. Its sides are parallel to the perpendicular unit vectors $\mathbf{v}_{B_{n,k}}$ and $\mathbf{w}_{B_{n,k}}$. The vector $\mathbf{v}_{B_{n,k}}$

is the direction vector of the block, on the folded paper picture this corresponds to the direction of the folding edges. The angle between this vector and $(0,1)$ will be in $[-\pi/4, \pi/4]$. The vector $\mathbf{w}_{B_{n,k}}$ is chosen so that its first component is positive and on Figure 6 this is parallel with the gradient vectors. We will have

$$B_{n,k} \stackrel{\text{def}}{=} \{\mathbf{o}_{B_{n,k}} + \alpha \mathbf{v}_{B_{n,k}} + \beta \mathbf{w}_{B_{n,k}} \text{ for } |\alpha|, |\beta| < l_{B_{n,k}}\}.$$

For each perturbation block we choose a *perturbation function* $\phi_{B_{n,k}}$ such that this function is continuously differentiable and zero outside $B_{n,k}$, this is a sanded and tamed version of the folded paper but one should essentially think of it as the folded paper surface with smooth/sanded folding edges from Figure 6. To be more precise there will be some transient regions close to the boundary of $B_{n,k}$, but the series of the total measures of these transient regions converges and by the Borel–Cantelli lemma λ_2 almost every point is not in a transient region. For a point $\mathbf{p} \in G$ there exists at most one perturbation block, $B_n(\mathbf{p})$, containing $\mathbf{p}$. If there is no such block we set $B_n(\mathbf{p}) = \emptyset$ and $\phi_{B_n(\mathbf{p})} = 0$.

Now set $h_n(\mathbf{p}) \stackrel{\text{def}}{=} \phi_{B_n(\mathbf{p})}(\mathbf{p})$ and in [8] we showed that

$$\nabla f(\mathbf{p}) = (0,1) + \sum_{n=0}^{\infty} \nabla \phi_{B_n(\mathbf{p})}(\mathbf{p}) = \sum_{n=-1}^{\infty} \nabla h_n(\mathbf{p}).$$

It is clear that

$$\nabla f_n(\mathbf{p}) = (0,1) + \sum_{k=0}^{n} \nabla \phi_{B_k(\mathbf{p})}(\mathbf{p}).$$

Next we outline how the perturbation blocks at level n are defined. We proceed by induction.

We chosse a sufficiently small $c_0 = 0.004$. We use four perturbation blocks at level 0, these are: $B_{0,1} = (0,1) \times (0,1)$, $B_{0,2} = (-1,0) \times (0,1)$, $B_{0,3} = (-1,0) \times (-1,0)$, and $B_{0,4} = (0,1) \times (-1,0)$. We have $\mathbf{v}_{B_{0,k}} = (0,1)$ and $\mathbf{w}_{B_{0,k}} = (1,0)$ for $k = 1, ..., 4$. We choose the functions $\phi_{B_{0,k}}$, $k = 1, ..., 4$.

Setting $G_0 \approx \cup_k B_{0,k}$ we define an open set and $G \setminus G_0$ is of measure zero. (Due to some technical details G_0 is a little smaller than $\cup_k B_{0,k}$, but for this heuristic argument one can think of it as being equal to it.) For $\mathbf{p} \in G_0$ set $x_0(\mathbf{p}) = 0 = \pi_x(\nabla_{rc,0} f(\mathbf{p})) \approx \pi_x(\nabla f_{-1}(\mathbf{p})) = \pi_x(\nabla h_{-1}(\mathbf{p}))$, where the projection onto the x axis is denoted by π_x, again to avoid technical difficulties one should think of x_0 as the first coordinate of $\nabla h_{-1}(\mathbf{p})$. We also define an auxiliary function $g_{0,\mathbf{p}}(x) = 1 - \frac{x^4}{4}$. These conacve down auxiliary functions can help to determine the direction vectors of the perturbation blocks at the next level.

Assume $n \geq 0$, the constant c_n and the open set G_n are given. Suppose furthermore that for all $\mathbf{p} \in G_n$ the trajectory $\{x_0(\mathbf{p}), ..., x_n(\mathbf{p})\} \subset [-1,1]$, the perturbation blocks $B_0(\mathbf{p}), ..., B_n(\mathbf{p})$ and the concave down functions $g_{n,\mathbf{p}}$ are defined. Moreover $|g'_{n,\mathbf{p}}(x)| \leq 1$ for all $x \in [-1,1]$ and ∇f_n is continuous on G_n.

Set $G_n^* = \{\mathbf{p} \in G_n : \nabla f_n(\mathbf{p}) \in \Omega_0\}$. For $\mathbf{p} \in G_n^*$ we define $\mathbf{v}_{\mathbf{p},n+1}$ as the "upward" normal vector of $g_{n,\mathbf{p}}$ at the point $(x_n(\mathbf{p}), g_{n,\mathbf{p}}(x_n(\mathbf{p})))$. We also put

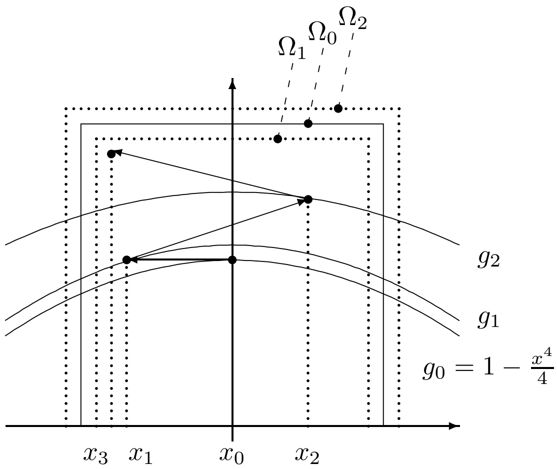


Figure 7: trajectory in the gradient space

$x^*_{n+1}(\mathbf{p}) = \pi_x(\nabla_{mrc,n+1} f(\mathbf{p})) \approx \pi_x(\nabla f_n(\mathbf{p}))$, again to simplify technical difficulties, due to some error terms, one can use this second approximate expression in this heuristic outline, instead of the exact value, $\pi_x(\nabla_{mrc,n+1} f(\mathbf{p}))$ given in [8].

We use Vitali's covering theorem to select the centers of the perturbation blocks at level $n+1$. In order to continue the perturbation of the gradient of f_n at λ_2 almost every point of G^*_n we cover almost every point of G^*_n by the perturbation blocks $B_{n+1,k}$, $k = 1, 2, ...$ These perturbation blocks are chosen so that if $\mathbf{p}$ is the center of such a block then $\mathbf{v}_{\mathbf{p},n+1}$ is the direction of this block. The values $x^*_{n+1}(\mathbf{p})$ will determine $x_{n+1}(\mathbf{p})$.

Finally, a few words about the functions g_n which play a crucial role in the determination of the direction of the perturbation blocks. The definition of these functions is very delicate. There were two sources of inspiration to define them. The first source was Theorem 3, it gave the idea of working with concave down functions. The second source was coming from one dimensional dynamical systems and learned it while I worked on the proof of [1]. I was also asked why I started and used $1 - \frac{x^4}{4}$, instead of the much more natural second order $1 - \frac{x^2}{2}$. The answer is that one needs a varying second derivative, the argument is not working if the second derivative is constant.

Assume $\mathbf{p} \in G$ is fixed and it is subject to infinitely many perturbations. One can think of $(x_{n,\mathbf{p}}, y_{n,\mathbf{p}}) = (x_n, y_n)$, $n = 0, 1, ...$, as the approximate value $\nabla f_{n-1}(\mathbf{p})$, again we ignore some small error terms. In order to verify that f is differentiable at $\mathbf{p}$ one has to show that (x_n, y_n) converges. (In the language of dynamical systems we need to show that if the "trajectory in the gradient space", (x_n, y_n), is not escaping from Ω_0 then it converges.) For ease of notation we write g_n, instead of $g_{n,\mathbf{p}}$. See Figure 7.

The functions g_n, for all n are concave (down) on $\mathbb{R}$ and they are all Lipschitz 1 functions on $[-1, 1]$.

They satisfy the following lemmas:

Lemma 12. Lemma 2 of [8]. *For all $x \in \mathbb{R}$ and $n = 0, 1, ...$ we have $g_{n+1}(x) \geq g_n(x)$.*

Lemma 13. Lemma 3 of [8]. *If $x_n \in [-1, 1]$ and $x_n \to x^*$ then there exists $y^* \in \mathbb{R} \cup \{+\infty\}$ such that $y_n \to y^*$.*

Lemma 14. Lemma 4 of [8]. *If $x_n \in [-1, 1]$, $\liminf x_n = x_* < \limsup x_n = x^*$ and $c = (x_* + x^*)/2$ then $g_n(c) \to \infty$. This, by the uniform Lipschitz property of g_n, implies $y_n \to \infty$ as well.*

These lemmas imply convergence of (x_n, y_n) and differentiability of the function constructed. Lemma 12 says that the higher n the higher the graph of g_n. This and the non-uniform concavity of the g_n in Lemma 13 show that convergence of the first coordinates implies that second coordinates have a finite or infinite limit. Finally Lemma 14 can be used to show that if (x_n, y_n) is a bounded sequence then the first coordinates converge and Lemma 13 is applicable.

References

[1] K. M. Brucks and Z. Buczolich, "Trajectory of the turning point is dense for a co-σ-porous set of tent maps", *Fund. Math.* **165** (2000) no. 2, 95-123.

[2] A. Bruckner, "The problem of characterizing derivatives revisited", *Real. Anal. Exch.* **21** (1995/96), no. 1, 112-133.

[3] Z. Buczolich, "The n-dimensional gradient has the 1-dimensional Denjoy-Clarkson property", *Real Anal. Exch.*, **18** (1992/93), no. 1, 221-224.

[4] Z. Buczolich, "Level sets of functions $f(x, y)$ with nonvanishing gradient", *J. Math. Anal. Appl.*, **185** (1994), no. 1, 27-35.

[5] Z. Buczolich, "Approximate continuity points of derivatives of functions of several variables", *Acta Math. Hungar.* **67** (1995) no. 3, 229-235.

[6] Z. Buczolich, "Another note on the gradient problem of C. E. Weil", *Real Anal. Exch.*, **22** (1996/97), no. 2, 775-784.

[7] Z. Buczolich, "Functions of two variables with large tangent plane sets", *J. Math. Anal. Appl.*, **220** (1998), no. 2, 562-570.

[8] Z. Buczolich, "Solution to the gradient problem of C. E. Weil", to appear in *Revista Matematica Iberoamericana*, preprint available from `www.cs.elte.hu/∼buczo`.

[9] F. Cater, “A global implicit function theorem”, *Real. Anal. Exchange* **16**, No. 1, (1990/91), 268-273.

[10] J. A. Clarkson, “A property of derivatives”, *Bull. Amer. Math. Soc.* **53** (1947) 124-125.

[11] A. Denjoy, “Sur une proprieté des fonctions dérivées”, *Enseignement. Math.* **18** (1916) 320-328.

[12] M. Elekes, “Level sets of differentiable functions of two variables with non-vanishing gradient”, *J. Math. Anal. Appl.*, **270** (2002) 369-382.

[13] H. Federer, *Geometric measure theory*, Die Grundlehren der matematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York (1969).

[14] M. W. Hirsch, *Differential Topology*, Graduate Texts in Mathematics, **33**, Springer-Verlag, Berlin/New York, 1976.

[15] P. Holický, J. Malý, L. Zajíček, C. E. Weil, “A note on the gradient problem”, *Real Anal. Exch.*, **22** (1996/97), no. 1, 225-235.

[16] T. W. Körner, “A dense arcwise connected set of critical points– Molehills out of mountains”, *J. London Math. Soc.* (2) **38** (1988), 442-452.

[17] A. P. Morse, “The behavior of a function on its critical set”, *Ann. of Math.* (2) **40** (1939), 62-70.

[18] G. Petruska, “Derivatives take every value on the set of approximate continuity points”, *Acta Math. Hungar.* **42** (1983) no. 3-4, 355-360.

[19] M. Radulescu and S. Radulescu, “Local inversion theorems without assuming continuous differentiability, *J. Math. Anal. Appl.* **138** (1989), 581-590.

[20] W. Rudin, *Functional Analysis*, McGraw-Hill Series in Higher Mathematics, *McGraw-Hill Book Co.,* (1973).

[21] A. Sard, “The measure of the critical values of differentiable maps”, *Bull. Amer. Math. Soc.* **48** (1942), 883-890.

[22] S. Saks, *Theory of the integral*, Dover, New York, 1964.

[23] C. E. Weil, “On properties of derivatives”, *Trans. Amer. Math. Soc.* **114** (1965), 363-376.

[24] H. Whitney, “A function not constant on a connected set of critical points”, *Duke Math. J.*, **1**, (1935), 514-517.

[25] “Query 1”, *Real Anal. Exch.*, **16** (1990/91), no. 1, 373.