Non-sufficiency of smoothness in the gradient conjecture

Florian Grün[*]

October 22, 2023

Abstract

It is well known that for analytic cost functions, gradient flow trajectories have finite length and converge to a single critical point. The gradient conjecture of R. Thom states that, again for analytic cost functions, whenever the gradient flow trajectory converges, the limit of its unit secants exists. One might think that already the convergence of the gradient flow trajectory to a critical point is enough to ensure that the unit secants have a limit, but this does not hold - the gradient conjecture is to a certain extend sharp.

We provide a counterexample in case of the missing analyticity assumption, that is a smooth (non-analytic) cost function f, where the limit of unit secants does not exist. In addition, f satisfies even a strong geometric length-distance convergence property.

AMS subject classification: 37N40, 37C10, 51N20

1   Introduction

Let $f : \mathbb{R}^n \to \mathbb{R}$ and consider the Euclidean gradient flow given by

$$x'(t) = -\nabla f(x(t)), \quad t \geq 0, \quad x(0) = x_0. \tag{1.1}$$

In 1984, Lojasiewicz proved in a landmark paper [5] that under the assumption f being analytic, the gradient trajectory either has finite arc-length (i.e. it converges to a limit point x_*) or diverges, for $t \to \infty$. Recent work has been done to achieve similar results for discrete gradient decent systems and related optimization schemes, see [1]. It is well known that merely smoothness of the cost function f is not sufficient to guarantee the same result. An example are the so called "Mexican hat" functions, see [1], [6, page 13] and [2], that construct a function in polar coordinates, where the gradient trajectory has the whole unit circle as its limit set.

One is not only interested whether the trajectory converges, but also in its geometry. In 1989, René Thom proposed in [7] the gradient conjecture. He conjectured that a convergent trajectory "does not spiral around its limit point".

Theorem 1.1. *Let $x(t)$ be a solution of (1.1) with f analytic and suppose that $x(t) \to x_*$. Then $x(t)$ has a tangent at x_*, that is the limit of secants $\lim_{t \to \infty} \frac{x(t) - x^*}{|x(t) - x^*|}$ exists.*

Ten years later the conjecture was proven to be true in [3], making use of techniques from subanalytic geometry originating in [5] and [4].

We now ask the question whether smoothness could be enough to have existence of the limit of secants for a converging gradient trajectory x. As it turns out, the answer is no and we construct an explicit gradient trajectory of a smooth function f, where the limit of unit secants does not exists.

Upon reparametrizing by arc-length s, without loss of generality we assume that x_* is the origin and $f(x_*) = 0$, i.e. the differential equation in (1.1) becomes $x'(s) = \frac{\nabla f(x(s))}{|\nabla f(x(s))|}$ and we consider $s \to s_0 < \infty$. In the process of the proof in [3] a related weaker statement (than the convergence of unit secants) is also shown.

Corollary 1.2. *("length-distance convergence") [3, Corollary 6.5] Let $x(s)$ be a trajectory of $\frac{\nabla f}{|\nabla f|}$ and $x(s) \to 0$ as $s \to s_0$. Denote by $\sigma(s)$ its arc-length to the origin, then*

$$\frac{\sigma(s)}{|x(s)|} \to 1 \qquad as \ s \to s_0.$$

---

[*]École Polytechnique Fédérale de Lausanne (EPFL), Insitute of Mathematics, `florian.grun@epfl.ch`

The counterexample given here also satisfies the length-distance convergence, i.e. its arc-length approaches its distance to the origin. Morally speaking, the geometry of the trajectory resembles more and more the straight line segment connecting $x(s)$ and the origin.

2 Counterexample

The function is inspired by the Mexican hat function [1] and also employs the fact that e^{-x} decays faster than any polynomial.

The main idea is to construct a spiral curve $\gamma : [2, \infty) \to \mathbb{R}^2$, written in polar coordinates as $\gamma(t) = (r(t), \theta(t))$, satisfying the desired properties and then construct a function $f(r, \theta)$ such that a gradient flow of f starting at e.g. $x_0 = \gamma(2) = (1/2, \log(\log(2)))$ follows the trajectory of γ.

2.1 The spiraling trajectory curve

Define the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ as

$$\gamma(t) = (r(t), \theta(t)) := \left(\frac{1}{t}, \log(\log(t)) \right).$$

We obtain the following results:

- $\gamma(t) \to 0$ as $t \to \infty$, since $r(t) = \frac{1}{t} \to 0$.

- γ has finite length: Since $\gamma'(t) = (x(t), y(t))' = (r(t) \cos \theta(t), r(t) \sin \theta(t))'$,

$$\|\gamma'(t)\| = \left((r' \cos(\theta) - r \sin(\theta) \theta')^2 + (r' \sin(\theta) + r \cos(\theta) \theta')^2 \right)^{1/2}$$

$$= \sqrt{r'^2 \cos^2 \theta - 2r' r \cos \theta \sin \theta \theta' + r^2 \sin^2 \theta \theta'^2 + r'^2 \sin^2 \theta + 2rr' \sin \theta \cos \theta \theta' + r^2 \cos^2 \theta \theta'^2}$$

$$= \sqrt{r'(t)^2 + r(t)^2 \theta'(t)^2} = \sqrt{\frac{1}{t^4} + \frac{1}{t^4} \frac{1}{\log(t)}}.$$

Therefore,

$$\int_2^\infty \|\gamma'(t)\| dt = \int_2^\infty \sqrt{\frac{1}{t^4} + \frac{1}{t^4} \frac{1}{\log(t)}} dt \le C \int_2^\infty \frac{1}{t^2} < \infty.$$

- length-distance convergence, i.e. $\frac{\sigma(s)}{|\gamma(s)|} = \frac{\int_s^\infty \|\gamma'(t)\| dt}{|\gamma(s)|} \to 1$ as $s \to \infty$: For this, we estimate

$$\int_s^\infty \|\gamma'(t)\| dt = \int_s^\infty \sqrt{\frac{1}{t^4} + \frac{1}{t^4} \frac{1}{\log(t)}} dt$$

$$\le \int_s^\infty \frac{1}{t^2} + \int_s^\infty \frac{1}{t^2} \frac{1}{\log(t)} dt$$

$$\le \frac{1}{s} + \frac{1}{\log(s)} \int_s^\infty \frac{1}{t^2} dt$$

$$= \frac{1}{s} + \frac{1}{s} \frac{1}{\log(s)}.$$

Which in turn implies

$$\frac{\sigma(s)}{|\gamma(s)|} = \frac{\int_s^\infty \|\gamma'(t)\| dt}{|\gamma(s)|} \le \frac{\frac{1}{s} + \frac{1}{s} \frac{1}{\log(s)}}{\frac{1}{s}} = 1 + \frac{1}{\log(s)} \to 1,$$

and thus $1 \le \lim_{s \to \infty} \frac{\sigma(s)}{|\gamma(s)|} = 1$.

- However, the unit secants, which are given in polar coordinates as

$$\frac{\gamma(t) - 0}{|\gamma(t) - 0|} = \frac{1}{r(t)} (r(t), \theta(t)) = (1, \theta(t)) = (1, \log(\log(t)),$$

do not converge.

To conclude, $\gamma(t)$ converges to the origin, its arc-length approaches its distance to the origin, yet it spirals as its unit secants do not converge.

2.2 The cost function

It remains to construct a function f, having exactly γ as a gradient flow trajectory. We define the function $f : B_{1/2} = \{x \in \mathbb{R}^n : |x| < 1/2\} \to \mathbb{R}$ as

$$f(r,\theta) := e^{\frac{-1}{r}} \left(1 - a(r)\sin(\theta - \log(\log(1/r)))\right)$$
$$= e^{\frac{-1}{r}} \left(1 - \frac{\log(1/r)}{1 + r^2 \log(1/r)^2} \sin\left(\theta - \log(\log(1/r))\right)\right),$$

for the auxiliary function $a(r) = \frac{\log(1/r)}{1+r^2\log(1/r)^2}$. The exponential factor gives smoothness of the function and all its derivatives at the origin. Actually, $D^\alpha f(0,\theta) = 0$ for any multi-index α, as the factor $e^{\frac{-1}{r}}$ decays faster than any polynomial expression in $\frac{1}{r}$ and $\log(\frac{1}{r})$ increases when $r \to 0$. By e. g. Whitney extension theorem the function could be extended to $\tilde{f}$ defined on $\mathbb{R}^2$, here it is sufficient to consider the behaviour close to the origin.

We show that the gradient flow of f, starting at a point x_0 on γ, stays always on γ, i.e. the polar gradient $(\frac{\partial f}{\partial r}, \frac{\partial f}{\partial \theta}\frac{1}{r})$ at points of the form $\{(r,\theta) : \theta - \log(\log(1/r)) = 2\pi k \quad \text{for some } k \text{ in } \mathbb{Z}\}$ (which are precisely the points on γ) is parallel to $-\gamma'$. Firstly, we note

$$\gamma'(t) = (r'(t), \theta'(t)) = \left(\frac{-1}{t^2}, \frac{1}{\log(t)}\frac{1}{t}\right) = \left(-r(t)^2, \frac{r(t)}{\log(1/r(t))}\right).$$

Thus it remains to check that for some scaling function $b(r)$ (in fact $b(r) = \frac{1}{r^2}e^{\frac{-1}{r}}\left(\frac{\log^2(1/r)}{1+r^2\log^2(1/r)}\right)$) on the trajectory, that is whenever $\theta = \log(\log(1/r))$, we have

$$\frac{\partial f}{\partial r} = b(r)r^2 \qquad \text{and} \qquad \frac{\partial f}{\partial \theta}\frac{1}{r} = b(r)\frac{-r}{\log(1/r)}.$$

We compute that,

$$\frac{\partial f}{\partial r}\bigg|_\gamma = \frac{1}{r^2}e^{\frac{-1}{r}}\left(1 - a(r)\sin(\theta - \log(\log(1/r)))\right)\bigg|_\gamma$$
$$+ e^{\frac{-1}{r}}\left(-a'(r)\sin(\theta - \log(\log(1/r))) - a(r)\cos(\theta - \log(\log(1/r)))\frac{1}{r^2\log(1/r)}\right)\bigg|_\gamma$$
$$= \frac{1}{r^2}e^{\frac{-1}{r}}\left(1 - \frac{a(r)}{\log(1/r)}\right)$$
$$= \frac{1}{r^2}e^{\frac{-1}{r}}\left(1 - \frac{1}{1 + r^2\log^2(1/r)}\right)$$
$$= e^{\frac{-1}{r}}\left(\frac{\log^2(1/r)}{1 + r^2\log^2(1/r)}\right),$$
$$\frac{\partial f}{\partial \theta}\bigg|_\gamma = -e^{\frac{-1}{r}}a(r)\cos(\theta - \log(\log(1/r)))$$
$$= -e^{\frac{-1}{r}}a(r)$$
$$= -e^{\frac{-1}{r}}\frac{\log(1/r)}{1 + r^2\log^2(1/r)}.$$

Thus it follows that

$$b(r)r^2 = e^{\frac{-1}{r}}\left(\frac{\log^2(1/r)}{1 + r^2\log^2(1/r)}\right) = \frac{\partial f}{\partial r},$$
$$b(r)\frac{-r}{\log(1/r)} = \frac{-1}{r}e^{\frac{-1}{r}}\left(\frac{\log^2(1/r)}{1 + r^2\log^2(1/r)}\right) = \frac{\partial f}{\partial \theta}\frac{1}{r}.$$

Since $b(r) > 0$ for all $0 < r < 1$,

$$-\nabla_{r,\theta}f|_\gamma = -\left(\frac{\partial f}{\partial r}, \frac{\partial f}{\partial \theta}\frac{1}{r}\right)\bigg|_\gamma = b(r)\left(\frac{-1}{r^2}, \frac{r}{\log(1/r)}\right).$$

Therefore $-\nabla_{r,\theta} f|_\gamma$ is parallel to γ' for any point on γ, so the gradient flow stays on the curve (albeit with a different speed than its parametrization). Moreover, since $\frac{\partial f}{\partial r}$ and $\frac{\partial f}{\partial \theta}$ converge to zero as $r \to 0$ ($a(r) \to 0$ as the exponential term dominates), the origin is a critical point.

To conclude, the trajectory of the continuous gradient descent flow converges to a critical point, has convergent arc-length, even length-distance convergence, yet the secants do not have any limit.

It would be interesting to know relations to even stronger algebraic-geometric results like the *(analytic) finiteness conjecture for the gradient*, see [3].
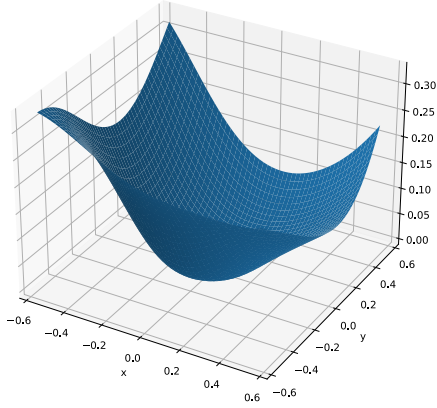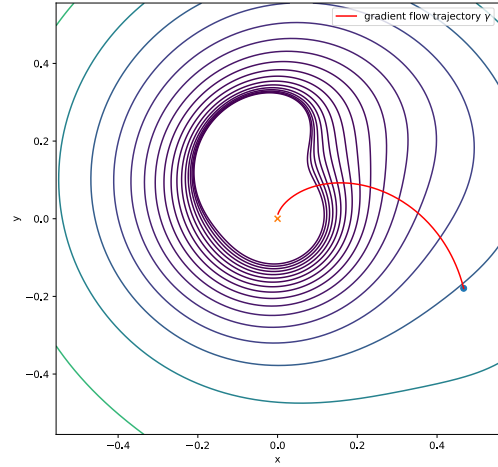

Figure 1: the cost function f

Figure 2: Contour plot of f and trajectory γ

3 Acknowledgements

References

[1] Pierre-Antoine Absil, Robert E. Mahony, and B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM J. Optim.*, 16:531–547, 2005.

[2] Haskell B. Curry. The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2:258–261, 1944.

[3] Krzysztof Kurdyka, Tadeusz Mostowski, and Adam Parusinski. Proof of the gradient conjecture of R. Thom. *Annals of Mathematics*, 152, 07 1999.

[4] Stanislaw Lojasiewicz. Ensembles semi-analytiques. 1965.

[5] Stanislaw Lojasiewicz. Trajectoires du gradient d'une fonction analytique. *Seminari di Geometria 1982-1983*, pages 115–117, 1984.

[6] Jacob Palis and Welington de Melo. *Geometric theory of dynamical systems*. Springer-Verlag, New York, 1982. Translated from the Portuguese by A. K. Manning.

[7] René Thom. Problèmes rencontrés dans mon parcours mathématique : un bilan. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 70:199–214, 1989.