Orbits of geometric descent

A. Daniilidis* D. Drusvyatskiy† A.S. Lewis‡

Abstract

We prove that quasiconvex functions always admit descent trajectories bypassing all non-minimizing critical points.

1 Introduction

To motivate the discussion, consider the classical gradient dynamical system

$$\dot{x} = -\nabla f(x), \qquad \text{where } f \text{ is a } C^1\text{-smooth function on } \mathbb{R}^d. \tag{1.1}$$

This differential equation always admits solutions starting from any point x_0, while uniqueness is only assured when the gradient ∇f is Lipschitz continuous. In this case, maximal trajectories of the system never encounter a singularity of f — a point where the gradient ∇f vanishes — in finite time. Instead, bounded trajectories converge in the limit to the critical set of the function. True convergence to a limit point is a more delicate matter; it is only guaranteed under extra assumptions on the function f, such as convexity [3, 4] or analyticity [2, 8] for example.

Reparametrizing the orbits of (1.1) by arclengths, at least away from singularities, we may instead seek absolutely continuous curves $x\colon [0, \eta) \to \mathbb{R}^d$ satisfying

$$\dot{x} = -\frac{\nabla f(x)}{\|\nabla f(x)\|}, \qquad \text{for a.e. } t \in [0, \eta), \tag{1.2}$$

*DIM-CMM, Universidad de Chile, Blanco Encalada 2120, piso 5, Santiago, Chile; http://www.dim.uchile.cl/~arisd/. Research supported by the grant MTM2011-29064-C01 (Spain) and FONDECYT Regular No 1130176 (Chile).

†Department of Combinatorics and Optimization, University of Waterloo, Ontario N2L 3G1, Canada; Department of Mathematics, University of Washington, Seattle, WA 98195; http://people.orie.cornell.edu/dd379/.

‡School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; http://people.orie.cornell.edu/aslewis/. Research supported in part by National Science Foundation Grant DMS-1208338.

where $\|\cdot\|$ denotes the norm on $\mathbb{R}^d$ and we temporarily adopt the convention $\frac{0}{0} = 0$. In comparison with (1.1), this system is much more intrinsic to the geometry of the level sets of f. Indeed, whenever ∇f is nonzero at a point x, the level set $[f = f(x)]$ is a smooth hypersurface around x and the right hand side of (1.2) coincides (up to sign) with the unit normal $\hat{n}(x)$ to the level set $[f = f(x)]$ at x. Consequently the orbits of the system (1.2) may reach a singularity in finite time and continue from there onward while not stopping at inessential singularities — points x where the gradient $\nabla f(x)$ vanishes but the level set $[f = f(x)]$ is a hypersurface around x. To emphasize this distinction further, observe that the range of any smooth function can clearly be reparametrized to force a singularity at any prespecified point; on the other hand, such a reparametrization does not effect the level set portrait of the function.

A particularly important situation arises when the function f is quasi-convex — meaning its sublevel sets $[f \leq r]$ are convex. Such functions play a decisive role for example in the theory of utility functions in microeconomics; see the landmark paper [1]. In this case, we may even drop the smoothness assumption on f and instead seek, in analogy to (1.2), absolutely continuous curves $x\colon [0, \eta) \to \mathbb{R}^d$ satisfying the inclusion

$$\dot{x} \in -N_{[f \leq f(x)]}(x), \qquad \text{for a.e. } t \in [0, \eta), \tag{1.3}$$

where $N_{[f \leq f(x)]}(x)$ denotes the convex normal cone to the sublevel set. In this short note, we prove that this system (under very mild assumptions on f) always admits Lipschitz continuous trajectories starting from any point. Moreover maximally defined trajectories are either unbounded or converge to the global minimum of the function.[1]

We should note a similarity of the differential inclusion (1.3) to the classical Moreau's Sweeping process introduced in [11]; for a nice expository article see [7]. The standard assumption for the sweeping process to admit a solution (within an appropriate space of curves) is for the sweeping set mapping to be continuous and of bounded variation. Then one can reparametrize the problem so that the sweeping set mapping becomes Lipschitz continuous and then apply the standard "catching up algorithm"; see [7] for details. In contrast, in the setting of the current manuscript the sublevel set mapping $t \mapsto [f \leq t]$ is not guaranteed to have bounded variation (see [2, Section 4.3] for a counter-example). Instead, the fundamental observation driving our analysis is that the polygonal curves created by the "catching up algorithm" are automatically self-contracted (Definition 2.4) and hence have

[1]While completing this short note, we became aware of the preprint [9], where the authors address questions of a similar flavor.

finite length whenever they are bounded [4], [10, Theorem 3.3]. This insight allows us to switch to the length parametrization and then apply the standard machinery of the theory of differential inclusions.

2 Trajectories of convex foliations

Throughout, we denote by $\mathbb{R}^d$ the d-dimensional Euclidean space. The corresponding inner-product and norm will be denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ respectively. For any subset Q of $\mathbb{R}^d$, the symbols $\operatorname{int} Q$, ∂Q, and $\operatorname{cl} Q$ will denote the topological interior, boundary, and closure of Q, respectively. The *distance* of a point x to Q is

$$d(x, Q) := \inf_{y \in Q} d(x, y),$$

and the *metric projection* of x onto Q is

$$P_Q(x) := \{y \in Q : d(x, y) = d(x, Q)\}.$$

Given points $x, y \in \mathbb{R}^d$ we define the closed segment

$$[x, y] := \{tx + (1 - t)y : t \in [0, 1]\}.$$

A subset Q of $\mathbb{R}^d$ is *convex* if for every pair of points $x, y \in C$ the line segment $[x, y]$ lies in Q. The *convex hull* of any set $Q \subset \mathbb{R}^d$, namely the intersection of all convex sets containing Q, will be denoted by $\operatorname{conv} Q$.

The following notion, introduced in [5, Section 6.3] and further studied in [4, Section 4.1], is the focus of this short note.

Definition 2.1 (Convex foliation)**.** An ordered family of sets $\{S_t\}_{t \in [a,b]}$, indexed by an interval $[a, b] \subset \mathbb{R}$, is called a *convex foliation* provided the following properties hold.

1. The sets S_t are nonempty, closed, convex subsets of $\mathbb{R}^d$.
2. The implication
$$t_1 < t_2 \quad \Longrightarrow \quad S_{t_1} \subset \operatorname{int} S_{t_2} \qquad \text{holds.}$$
3. The equation
$$\bigcup_{t \in [a,b]} \partial S_t = S_b \setminus (\operatorname{int} S_a) \qquad \text{holds.}$$

For each point $x \in S_b \setminus (\operatorname{int} S_a)$, abusing notation slightly, we define the set S_x to be the unique set of the convex foliation satisfying $x \in \partial S_x$.

Remark 2.2. We mention in passing that any convex foliation can be represented in terms of sublevel sets of an lsc quasiconvex function $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ that is continuous on its domain and has no nonglobal extrema; conversely, sublevel sets of any such function naturally define a convex foliation.

For any convex subset Q of $\mathbb{R}^d$ and any point $\bar{x} \in Q$ the *normal cone* $N_Q(\bar{x})$ has the classical description:

$$N_Q(\bar{x}) = \Big\{ v \in \mathbb{R}^d : \ \langle v, x - \bar{x} \rangle \le 0, \quad \text{for all } x \in Q \Big\}.$$

The following is a key definition of the current work.

Definition 2.3 (Trajectories of convex foliations)**.** A curve γ is a *trajectory of a convex foliation* $\{S_t\}_{t \in [a,b]}$ if it admits an absolutely continuous parametrization $\gamma\colon I \to \mathbb{R}^d$ satisfying

$$\dot{\gamma}(\tau) \in -N_{S_{\gamma(\tau)}}(\gamma(\tau)) \qquad \text{for almost every } \tau \in I,$$

and for any $\tau_1, \tau_2 \in I$ with $\tau_1 < \tau_2$ we have $\gamma(\tau_2) \subset \operatorname{int} S_{\gamma(\tau_1)}$.

Our goal in this short note is to prove that trajectories of convex foliations always exist. The following notion turns out to be instrumental. For more details see [5].

Definition 2.4 (Self-contracted curve)**.** A curve $\gamma\colon I \to \mathbb{R}^d$ is called *self-contracted* if for any $t^* \in I$, the mapping

$$t \mapsto d(\gamma(t), \gamma(t^*)), \qquad \text{is nonincreasing on } I \cap (-\infty, t^*].$$

The following result concerning lengths of self-contracted curves will be key for us. See [10] for Lipschitz curves and [4, Theorem 3.3] for general (possibly discontinuous) self-contracted curves.

Lemma 2.5 (Lengths of self-contracted curves)**.** *Consider a self-contracted curve $\gamma\colon I \to \mathbb{R}^d$ and let $\Gamma \subset \mathbb{R}^d$ be the image of I under γ. Then we have the estimate*

$$\operatorname{length}(\gamma) \le K_d \operatorname{diam}(\Gamma),$$

where K_d is a constant that depends only on the dimension d.

We arrive at the main result of this short note.

Theorem 2.6 (Trajectories of convex foliations exist)**.** *Consider a convex foliation $\{S_t\}_{t\in[a,b]}$. Then for any point $x_0 \in S_b$ there exists a self-contracted curve $\gamma\colon [0,L] \to \mathbb{R}^d$ that is a trajectory of the convex foliation and satisfies $\gamma(0) = x_0$ and $\gamma(L) \in S_a$.*

Proof. Before we begin, we record the following result which will be used in the sequel. The proof is based on a standard convexity argument and will be omitted. We defer to [12, Definition 5.4] for the relevant definitions of continuity of set-values mappings.

Claim 2.7. *If $\{S_t\}_{t\in[a,b]}$ is a convex foliation, then the mappings $t \mapsto S_t$ and $x \mapsto N_{S_x}(x)$ are continuous in a set-valued sense.*

Consider a partition $a = \tau_n < \tau_{n-1} < \ldots < \tau_1 < \tau_0 = b$ of the interval $[a,b]$. Now inductively define the points

$$x_i = \mathrm{proj}_{S_{\tau_i}}(x_{i-1}) \qquad \text{for } i = 1,\ldots,n. \tag{2.1}$$

and consider the polygonal line

$$\Gamma_n = \bigcup_{i=0}^{n-1} [x_i, x_{i+1}].$$

Let $\gamma_n\colon [0,L_n] \to \mathbb{R}^d$ be the arclength parametrization of Γ_n. The following is true.

Claim 2.8. *The curves γ_n are self-contracted and satisfy $L_n \leq K_d \,\mathrm{dist}\,(x_0, S_a)$, where K_d is a constant depending only on the dimension d.*

Proof. Fix an index $i \in \{0,\ldots,n-1\}$. Since $S_{\tau_{i+1}}$ is convex and we have $x_i - x_{i+1} \in N_{S_{\tau_{i+1}}}(x_{i+1})$, it follows that for every fixed $x \in S_{\tau_{i+1}}$, the function

$$\theta \mapsto \|x_{i+1} + \theta(x_i - x_{i+1}) - x\|, \quad \theta \geq 0,$$

is non-decreasing. In particular, for any point $x \in S_a$ we have

$$\|x_i - x\| \geq \|x_{i+1} - x\|.$$

Since i was arbitrary, we deduce $\mathrm{dist}(x_0, S_a) \geq \|x_{i+1} - \mathrm{proj}_{S_a}(x_0)\|$ and consequently all the curves γ_n are contained in a ball of radius $\mathrm{dist}(x_0, S_a)$ around $\mathrm{proj}_{S_a}(x_0)$.

Consider now real numbers $0 \leq e < f < g \leq L$. In the case that $\gamma(e)$, $\gamma(f)$, $\gamma(g)$ all lie in a single line segment $[x_i, x_{i+1}]$, the inequality

$$\|\gamma(g) - \gamma(f)\| \leq \|\gamma(g) - \gamma(e)\|,$$

is obvious. Hence we may suppose that there are indices $0 \leq i_1 \leq i_2 \leq i_3 \leq n$, that are not all the same, and satisfying

$$\gamma(e) \in [x_{i_1}, x_{i_1+1}], \qquad \gamma(f) \in [x_{i_2}, x_{i_2+1}], \qquad \gamma(g) \in [x_{i_3}, x_{i_3+1}].$$

Observe that the inclusion

$$\gamma(g) \in S_{\tau_i}, \qquad \text{holds whenever } i_1 \leq i < i_2,$$

Consequently for such indices i, we have

$$\|x_i - \gamma(g)\| \leq \|x_{i_1} - \gamma(g)\|.$$

It follows immediately that the polygonal curve γ is self-contracted. The bound on the length of Γ_n now follows directly from Lemma 2.5. $\square$

In light of the claim above, the lengths of the curves γ_n are bounded by a uniform constant

$$L_* := K_d \operatorname{dist}(x_0, S_a).$$

We can thus extend the domains of the curves γ_n from $[0, L_n]$ to $[0, L_*]$ (and continue to denote by γ_n the new curves for simplicity) as follows:

$$\gamma_n(s) = \gamma_n(L), \quad \text{for every } s \in [L, L_*].$$

Now let the mesh of the partition $a = \tau_n < \tau_{n-1} < \ldots < \tau_1 < \tau_0 = b$ tend to zero as n tends to ∞. Clearly each curve γ_n is 1-Lipschitz. It follows that the sequence $\{\gamma_n\}_n$ is equi-continuous and equi-bounded, and hence by the Arzela-Ascoli theorem (see for example [6, Section 7]) it has a subsequence, which we still denote by $\{\gamma_n\}_n$, that converges uniformly to a curve $\gamma\colon [0, L_*] \to \mathbb{R}^d$. It follows that γ is a self-contracted, 1-Lipschitz continuous curve, satisfying $\gamma(0) = x_0$. In particular the inequality $\|\dot{\gamma}(s)\| \leq 1$ holds almost everywhere on $[0, L_*]$. Consider now the sequence of derivatives $\{\dot{\gamma}_n\}_n$ in the Hilbert space $L^2([0, L_*], \mathbb{R}^d)$ (equipped with the $\|\cdot\|_2$-norm). Notice that the inequalities $\|\dot{\gamma}_n\|_2 \leq \sqrt{L_*}$ hold for all n. Thus the sequence $\{\dot{\gamma}_n\}_n$ has a weakly converging subsequence, which we still denote by $\{\dot{\gamma}_n\}_n$. A standard argument easily shows that this limit coincides with $\dot{\gamma}$ almost everywhere on $[0, L_*]$.

Mazur's Lemma then implies that a subsequence of convex combinations of the form $\sum_{k=n}^{K(n)} \alpha_k^n \dot{\gamma}_k$ converges strongly to $\dot{\gamma}$ as n tends to ∞. Since convergence in $L^2[0, L_*]$ implies almost everywhere pointwise convergence, we deduce that for almost every $s \in [0, L_*]$, we have

$$\Big\| \sum_{k=n}^{K(n)} \alpha_k^n \, \dot{\gamma}_k(s) - \dot{\gamma}(s) \Big\| \to 0, \quad \text{as } n \to \infty.$$

Fix such a number $s \in [0, L_*]$. Then by Carathéodory's theorem we may assume that the quantity $K(n) - (n-1)$ is bounded by $d+1$. Relabelling we then have

$$\lim_{n\to\infty} \sum_{i=1}^{d+1} \lambda_i^n \, \dot{\gamma}_i^n(s) = \dot{\gamma}(s).$$

Passing successively to subsequences, we may assume that

$$\dot{\gamma}_i^n(s) \to v_i(s), \quad \text{for all } i \in \{1, \ldots, d+1\}, \tag{2.2}$$

and similarly,

$$(\lambda_1^n, \ldots, \lambda_{d+1}^n) \to (\lambda_1, \ldots, \lambda_{d+1}).$$

Consequently we obtain the inclusion

$$\dot{\gamma}(s) \in \operatorname{conv}\{v_1, \ldots, v_{d+1}\}. \tag{2.3}$$

By construction for each $i \in \{1, \ldots, d+1\}$ and $n \in \mathbb{N}$, there exist real numbers $\tau_{i_n}^- > \tau_{i_n}^+$ and corresponding $s_{i_n}^- < s_{i_n}^+$ satisfying $S_{\gamma_{i_n}(s_{i_n}^-)} = S_{\tau_{i_n}^-}$ and $S_{\gamma_{i_n}(s_{i_n}^+)} = S_{\tau_{i_n}^+}$ and so that

$$\gamma_{i_n}(s) \in [\gamma_{i_n}(s_{i_n}^-), \gamma_{i_n}(s_{i_n}^+)], \qquad \dot{\gamma}_{i_n}(s) \in -N_{S_{\tau_{i_n}^+}}(\gamma_{i_n}(s_{i_n}^+))$$

Now observe $\|\gamma_{i_n}(s_{i_n}^-) - \gamma_{i_n}(s_{i_n}^+)\| = d(\gamma_{i_n}(s_{i_n}^-), S_{\tau_{i_n}^+})$. According to Claim 2.7 the set-valued mapping $t \mapsto S_t$ is continuous, whence we obtain $\|\gamma_{i_n}(s_{i_n}^-) - \gamma_{i_n}(s_{i_n}^+)\| \to 0$. The outer semicontinuity of the mapping $x \mapsto N_{S_x}(x)$ (Claim 2.7), along with (2.2) immediately yields

$$-\dot{\gamma}(s) \in N_{S_{\gamma(s)}}(\gamma(s)), \qquad \text{for a.e. } s \in [0, L_*]. \tag{2.4}$$

Let L be the total length of the self-contracted curve γ. We now reparametrize γ by arc-length and continue to denote the resulting curve by γ (since no confusion will arise). This curve is now defined on $[0, L]$ and satisfies equation (2.4) with $\|\dot{\gamma}(s)\| = 1$, a.e.

Now to complete the proof, assume towards a contradiction, that for some $s_1 < s_2$ and all $s \in [s_1, s_2]$ the set $S_{\gamma(s)}$ is constantly equal to some set Q. Then we have $(\delta_Q \circ \gamma)(s) = 0$, for all $s \in [s_1, s_2]$. Then by [12, Theorem 10.6] we have for almost all s and all $v(s) \in N_Q(\gamma(s))$

$$\frac{d}{dt}(\delta_Q \circ \gamma)(s) = \langle \dot{\gamma}(s), v(s) \rangle = 0.$$

In view of (2.4) this yields $\|\dot{\gamma}(s)\| = 0$ a.e. on $[s_1, s_2]$. This contradicts the fact that γ is parametrized by arclength, and concludes the proof. $\square$

Corollary 2.9 (Smooth convex foliations)**.** *Consider a convex foliation $\{S_t\}_{t\in[a,b]}$ and suppose moreover that the sets ∂S_t are C^1-smooth manifolds for each $t \in [a, b]$. Then every trajectory $\gamma\colon I \to \mathbb{R}^d$ of the convex foliation can be parametrized by arclength, at which point it becomes C^1-smooth on the interior of its domain of definition.*

Proof. Observe that for every point $x \in S_b \setminus \operatorname{int} S_a$, there exists a unitary normal vector $\hat{n}(x) \in \mathbb{R}^d$ satisfying

$$N_{S_x}(x) = \mathbb{R}_+ \hat{n}(x).$$

The assignment $x \mapsto \hat{n}(x)$ is a unitary continuous vector field on $S_b \setminus \operatorname{int} S_a$. On the other hand, when γ is parametrized by arclength, we have $\dot{\gamma}(s) = \hat{n}(\gamma(s))$ a.e. on γ's domain of definition. Since we have the representation

$$\gamma(s) = \gamma(0) + \int_0^s \dot{\gamma}(\tau)\, d\tau = \gamma(0) + \int_0^s \hat{n}(\gamma(\tau))\, d\tau,$$

we deduce that γ is a C^1-smooth curve on the interior of its domain. $\square$

Acknowledgements: The first author thanks David Marin (UAB) for useful discussions. Part of this work has been realized during a research stay of the first author at Cornell University (December 2012). This author thanks his hosts and the host institution for hospitality.

References

[1] Arrow, K.J., Debreu, G., Existence of an equilibrium for a competitive economy., *Econometrica,* **22** (1954), 265–290.

[2] Bolte, J., Daniilidis, A., Ley, O., Mazet, L., Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity, *Trans. Amer. Math. Soc.* **362** (2010), 3319–3363.

[3] Bruck, R., Asymptotic convergence of nonlinear contraction semigroups in Hilbert space, *J. Funct. Anal.* **18** (1975), 15–26.

[4] Daniilidis, A., David, G., Durand-Cartagena, E., Lemenant, A., Rectifiability of self-contracted curves in the Euclidean space and applications, *J. Geom. Anal.* (to appear in 2013) `DOI 10.1007/s12220-013-9464-z`

[5] Daniilidis, A., Ley, O., Sabourau, S., Asymptotic behaviour of self-contracted planar curves and gradient orbits of convex functions, *J. Math. Pures Appl.* **94** (2010), 183–199.

[6] Kelley, J.L., *General topology* (Springer-Verlag, New York, 1975). Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27.

[7] Kunze, M., Marques, M., & Manuel, D.P., An introduction to Moreau's sweeping process. (English summary) *Impacts in mechanical systems (Grenoble, 1999)*, 1–60, Lecture Notes in Phys., 551, Springer, Berlin, 2000.

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

[9] Longinetti, M., Manselli, P., Venturi, A., On steepest descent curves for quasiconvex families in R^n, preprint 29 p., arXiv:1303.3721 (March 2013)

[10] Manselli, P. Pucci, C., Maximum length of steepest descent curves for quasi-convex functions, *Geom. Dedicata* **38** (1991), 211–227.

[11] Moreau, J.-J., Evolution problem associated with a moving convex set in a Hilbert space. *J. Differential Equations* **26** (1977), no. 3, 347374.

[12] Rockafellar, T., Wets, R., *Variational analysis* (Springer, 1998).