The slope robustly determines convex functions

Aris Daniilidis & Dmitriy Drusvyatskiy

Abstract. We show that the deviation between the slopes of two convex functions controls the deviation between the functions themselves. This result reveals that the slope—a one dimensional construct—robustly determines convex functions, up to a constant of integration.

Key words. Convex function, subgradient, slope, stability.

AMS Subject Classification *Primary* 26B25, 49K40 ; *Secondary* 37C10, 49J52.

1 Introduction

The recent paper [2, Theorem 3.8] established the following intriguing result. Two $\mathcal{C}^2$-smooth, convex and bounded from below functions f, g defined on a Hilbert space $\mathcal{H}$ are equal up to an additive constant if and only if their gradient norms coincide:

$$\|\nabla f\| = \|\nabla g\| \qquad \Longleftrightarrow \qquad f = g + \text{cst}. \tag{1.1}$$

This result is ostensibly surprising since it readily yields that the function $x \mapsto \|\nabla f(x)\|$, which takes values in the real line, determines the entire gradient map $x \mapsto \nabla f(x)$, which takes values in $\mathcal{H}$. In the follow up work [11], the assumption on smoothness of f was further weakened to continuity with the gradient norm $\|\nabla f(x)\|$ replaced by the slope $s_f(x) := \operatorname{dist}(0, \partial f(x))$. Here $\partial f(x)$ denotes the subdifferential of the convex function f at x.[1]

In this work, we ask whether the slope (or the gradient norm in the smooth case) *robustly* determines the function itself. That is, if the slopes for two functions are close, then how close are the function values? Roughly speaking, we will show that for any two continuous convex functions f and g defined on a Hilbert space, the following estimate is true:

$$\|g - f\|_{\mathcal{U}} \;\lesssim\; \|s_g - s_f\|_{\mathcal{U}} + \sqrt{\|s_g - s_f\|_{\mathcal{U}}} + \|g - f\|_{C_f \cup C_g}.$$

Here $\mathcal{U}$ is any bounded set where f is bounded, $\|\cdot\|_{\mathcal{U}}$ denotes the sup-norm over $\mathcal{U}$, and C_f and C_g are the sets of minimizers of f and g, respectively. In particular, the deviation $\|g - f\|_{\mathcal{U}}$ exhibits a dependence on $\|s_g - s_f\|_{\mathcal{U}}$ that is at worst Hölder with exponent $1/2$. In the finite-dimensional setting $\mathcal{H} = \mathbb{R}^n$, we show that this undesirable square root dependence may be dropped:

$$\|g - f\|_{\mathcal{U}} \lesssim \|s_g - s_f\|_{\mathcal{U}} + \|g - f\|_{C_f \cup C_g}.$$

The downside is that the hidden constant in this bound depends on the length of subgradient curves initialized in $\mathcal{U}$ and at worst grows super exponentially in the dimension n.

[1] We note that further generalizations of the determination result [11] have recently been achieved: for convex continuous bounded from below functions in Banach spaces (see [12]) and for Lipschitz coercive functions in metric spaces ([5]). For the time being, we do not pursue our sensitivity analysis in this generality.

2 Notation and preliminaries

Let $\mathcal{H}$ denote a Hilbert space and let $f : \mathcal{H} \to \mathbb{R}$ be a convex continuous function. We denote the set of minimizers of f by

$$\mathcal{C}_f := \arg\min f,$$

and suppose that $\mathcal{C}_f$ is nonempty (therefore the infimum value $f_* := \inf f$ is attained). The key object we will focus on is the slope $s_f(x) = \operatorname{dist}(0, \partial f(x))$, where $\partial f(x)$ denotes the subdifferential:

$$\partial f(x) = \{v \in \mathcal{H} : \, f(y) - f(x) \geq \langle v, y - x\rangle, \quad \forall x, y \in \mathcal{H}\}. \tag{2.1}$$

Equivalently, $s_f(x)$ measures the fastest instantaneous rate of decrease of f from x.

Our goal is to show that the deviation between the slopes of two convex functions controls the deviation between the functions themselves. Our arguments will make heavy use of subgradient dynamical systems, a topic we review now following [1, 3]. Namely, [1, Theorem 17.2.2] shows that for every initial point $x \in \mathcal{H}$, there exists a unique, maximally defined, injective, absolutely continuous curve $\gamma : [0, T_{\max}) \to \mathcal{H}$, such that

$$\begin{cases} \dot{\gamma}(t) \underset{\text{a.e}}{\in} -\partial f(\gamma(t)) \\ \gamma(0) = x \end{cases} \tag{GS}$$

Subgradient curves γ satisfy a number of useful properties, summarized below.

(P1) Equality
$$\|\dot{\gamma}(t)\| = s_f(\gamma(t)) \quad \text{holds for a.e. } t \in [0, T_{\max}). \tag{2.2}$$
and the slope function $t \mapsto s_f(\gamma(t))$ is nonincreasing on $[0, T_{\max})$.

(P2) The function $r(t) = f(\gamma(t))$ is convex and strictly decreasing on $[0, T_{\max})$, and
$$\lim_{t \to T_{\max}} f(\gamma(t)) = f_*.$$

(P3) The distance function $t \mapsto d(\gamma(t), \mathcal{C}_f)$ is strictly decreasing on $[0, T_{\max})$. Moreover, for every $x_* \in \mathcal{C}_f$, the function $t \mapsto \|\gamma(t) - x_*\|$ is strictly decreasing on $[0, T_{\max})$.

Property (P1) follows from [1, Theorem 17.2.2 (iii)-(iv)], (P2) is given in [1, Proposition 17.2.7 (i)], while (P3) follows easily after differentiation, using (GS) and (2.1).

Next, we will require two estimates on the length of subgradient curves. The first (Lemma 2.1) is an easy consequence of (P1) and (P2) above (we provide a proof for convenience), while the second (Proposition 2.2) was essentially proved in [10] for a particular class of Lipschitz curves (therein called Γ-curves, ultimately known as *self-contracted* curves, definition coined in [7]) and became explicit for subgradient curves in [6, 8].

Lemma 2.1 (Length estimation I)**.** *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex continuous function with nonempty set of minimizers and let $\gamma : [0, T_{\max}) \to \mathcal{H}$ be the solution of (GS). Then for every $T \in (0, T_{\max})$, setting $\gamma_T := \gamma(T)$ we have:*

$$\int_0^T |\dot{\gamma}(t)|\, dt \, \leq [s_f(\gamma_T)]^{-1} \left(f(x) - f_*\right).$$

Proof. Set $r(t) := f(\gamma(t))$ and denote by h the inverse function of the mapping $t \mapsto r(t)$ on the interval $[0, T_{\max})$. Then for the reparametrization $\tilde{\gamma}(\rho) = \gamma(h(\rho))$ we have $f(\tilde{\gamma}(\rho)) = \rho$. Differentiating gives

$$\frac{d}{d\rho}[\tilde{\gamma}(\rho)] = \frac{\partial f(\tilde{\gamma}(\rho))^{\circ}}{s_f(\tilde{\gamma}(\rho))^2}, \quad \text{for } a.e.\ \rho \in (f_*, f(x)],$$

where $\partial f(\tilde{\gamma}(\rho))^{\circ}$ is the element of $\partial f(\tilde{\gamma}(\rho))$ of minimal norm, thus $\|\partial f(\tilde{\gamma}(\rho))^{\circ}\| = s_f(\tilde{\gamma}(\rho))$. Taking into account that the function $\rho \mapsto s_f(\tilde{\gamma}(\rho))$ is increasing, we deduce:

$$\int_0^T \|\dot{\gamma}(t)\|\, dt \;=\; \int_{f(\gamma_T)}^{f(x)} \frac{1}{s_f(\tilde{\gamma}(\rho))}\, d\rho \;\leq\; \frac{f(x) - f(\gamma_T)}{s_f(\gamma_T)}$$

and the result follows. $\square$

Proposition 2.2 (Length estimation II)**.** *Assume $\mathcal{H} = \mathbb{R}^n$. There exists a constant K_n depending only on dimension such that for every $x \in \mathbb{R}^n$ the solution $\gamma(\cdot)$ of the subgradient system* (GS) *has length bounded by $K_n \cdot d(x, \mathcal{C}_f)$.*

The above result provides a universal bound K_n for the ratio between the length of a subgradient curve and its diameter, the drawback being that that the dependence of K_n on the dimension is of the order of $n^{n/2+1}$ (see [10, 9]).

3 Main results

For any function $\omega \colon \mathcal{H} \to \mathbb{R}$ and a set $\mathcal{U} \subset \mathcal{H}$, we will use the notation

$$\|\omega|_{\mathcal{U}} := \sup_{x \in \mathcal{U}} \left(\max\{\omega(x), 0\}\right) \qquad \text{and} \qquad \|\omega\|_{\mathcal{U}} = \sup_{x \in \mathcal{U}} |\omega(x)|.$$

Note that $\|\omega|_{\mathcal{U}}$ provides a one-sided bound[2], while $\|\omega\|_{\mathcal{U}}$ is the standard two-sided sup-norm.

The following is the main theorem of the paper.

Theorem 3.1. *Let $f, g \colon \mathcal{H} \to \mathbb{R}$ be convex continuous functions. Assume $\mathcal{C}_f = \arg\min f \neq \emptyset$ and set $f_* = \min f$. For each $r > 0$ define the tube around C_f by*

$$\mathcal{U}_r := \{x \in \mathcal{H} :\, d(x, \mathcal{C}_f) \leq r\}. \tag{3.1}$$

Then for every $x \in \mathcal{U}_r$, the estimate holds:

$$g(x) - f(x) \;\leq\; \|s_g - s_f|_{\mathcal{U}_r} \,+\, \|g - f|_{\mathcal{C}_f} \,+\, 2\sqrt{d(x, \mathcal{C}_f) \cdot \|s_g - s_f|_{\mathcal{U}_r} \cdot (f(x) - f_*)}. \tag{3.2}$$

Moreover, in the finite-dimensional setting $\mathcal{H} = \mathbb{R}^n$, there exists a constant $K_n > 0$ depending only on the dimension n such that

$$g(x) - f(x) \;\leq\; K_n\, \|s_g - s_f|_{\mathcal{U}_r}\, d(x, \mathcal{C}_f) \,+\, \|g - f|_{\mathcal{C}_f}. \tag{3.3}$$

[2] Notice that $\|\cdot|_{\mathcal{U}}$ is the canonical asymmetrization of the seminorm $\|\cdot\|_{\mathcal{U}}$ of uniform convergence, see [4].

Proof. Let $x \in \mathcal{H} \setminus \mathcal{C}_f$ be arbitrary and fix $\delta > 0$. Our goal is to show the estimate

$$g(x) - f(x) \le (\|s_g - s_f|_{\mathcal{U}_r} + \delta)\, d(x, \mathcal{C}_f) + \frac{\|s_g - s_f|_{\mathcal{U}_r}}{\delta}\, (f(x) - f_*) + \|g - f|_{\mathcal{C}_f}, \tag{3.4}$$

from which (3.2) follows by setting $\delta = \sqrt{\frac{\|s_g - s_f|_{\mathcal{U}_r} \cdot (f(x) - f_*)}{d(x, \mathcal{C}_f)}}$.

We consider two cases:

(i). Suppose that $s_f(x) \le \delta$ and let $\hat{x} := \mathrm{proj}_{\mathcal{C}_f}(x)$ be the projection of $\hat{x}$ to the closed convex set $\mathcal{C}_f$ (therefore $f(\hat{x}) = f_* \le f(x)$). Then we compute

$$g(x) - g(\hat{x}) \le s_g(x)\, \|x - \hat{x}\| \le (\|s_g - s_f|_{\mathcal{U}_r} + \delta)\, d(x, \mathcal{C}_f),$$

where the first inequality follows from convexity of g. We therefore conclude

$$\begin{aligned} g(x) - f(x) &= (g(x) - g(\hat{x})) + (g(\hat{x}) - f(\hat{x})) + (f(\hat{x}) - f(x)) \\ &\le (\|s_g - s_f|_{\mathcal{U}_r} + \delta)\, d(x, \mathcal{C}_f) + \|g - f|_{\mathcal{C}_f}, \end{aligned}$$

thus verifying (3.4).

(ii). Suppose now that $s_f(x) > \delta$ and let $\gamma \colon [0, T_{\max}) \to \mathcal{H}$ denote the unique maximal solution of the subgradient system (GS) for f. Define the function

$$a(t) := f(\gamma(t)) - g(\gamma(t)).$$

Differentiating, for *a.e.* $t \in [0, T_{\max})$, we have (*c.f.* [1, Proposition 17.2.5]):

$$\dot{a}(t) = -s_f(\gamma(t))^2 - \langle \partial g(\gamma(t))^\circ, \dot{\gamma}(t) \rangle,$$

where $\partial g(\gamma(t))^\circ$ is the element of minimal norm of $\partial g(\gamma(t))$, that is, $s_g(\gamma(t)) = \|\partial g(\gamma(t))^\circ\|$. From the Cauchy-Schwarz inequality we conclude:

$$\begin{aligned} \dot{a}(t) &\le -s_f(\gamma(t))^2 + s_g(\gamma(t)) \cdot s_f(\gamma(t)) \\ &= (s_g(\gamma(t)) - s_f(\gamma(t)))\; s_f(\gamma(t)) \\ &\le \|s_g - s_f|_{\mathcal{U}_r}\, \|\dot{\gamma}(t)\|. \end{aligned} \tag{3.5}$$

Define

$$T := \sup\, \{t \in [0, T_{\max}) :\; s_f(\gamma(t)) > \delta\}.$$

Setting $\gamma_T := \gamma(T)$ and integrating (3.5) on $[0, T]$ we obtain:

$$g(x) \le f(x) + [g(\gamma_T) - f(\gamma_T)] + \|s_g - s_f|_{\mathcal{U}_r} \int_0^T \|\dot{\gamma}(t)\|\, dt. \tag{3.6}$$

By Lemma 2.1 and the definition of T we get:

$$\int_0^T \|\dot{\gamma}(t)\|\, dt \le [s_f(\gamma_T)]^{-1}\, (f(x) - f_*) \le \delta^{-1}\, (f(x) - f_*)\,. \tag{3.7}$$

Let $\hat{\gamma} = \mathrm{proj}_{\mathcal{C}_f}(\gamma_T)$ be the projection of γ_T to the set of minimizers $\mathcal{C}_f$. Then

$$f(\hat{\gamma}) = f_* \le f(\gamma_T) \qquad \text{and} \qquad \|\gamma_T - \hat{\gamma}\| = d(\gamma_T, \mathcal{C}_f) \le d(x, \mathcal{C}_f).$$

Taking into account $s_f(\gamma_T) \le \delta$ we deduce $s_g(\gamma_T) \le \|s_g - s_f|_{\mathcal{U}_r} + \delta$ and consequently

$$g(\gamma_T) - g(\hat{\gamma}) \le s_g(\gamma_T)\,\|\gamma_T - \hat{\gamma}\| \;\le (\|s_g - s_f|_{\mathcal{U}_r} + \delta)\; d(x, \mathcal{C}_f),$$

where the first inequality follows from convexity of g. We readily obtain that:

$$g(\gamma_T) - f(\gamma_T) \;\le\; (g(\gamma_T) - g(\hat{\gamma})) + (g(\hat{\gamma}) - f_*) \;\le (\|s_g - s_f|_{\mathcal{U}_r} + \delta)\, d(x, \mathcal{C}_f) \;+\; \|g - f|_{\mathcal{C}_f}. \quad (3.8)$$

Combining (3.6), (3.7), and (3.8) yields the claimed estimate (3.4). Finally, the estimate (3.3) follows by letting $T \uparrow T_{\max}$ in (3.6) and using Proposition 2.2 to bound the length of $\gamma(\cdot)$. □

An easy consequence of the above is the following guarantee of asymptotic consistency.

Corollary 3.2 (Robust (one-sided) determination)**.** *Let $f, \{f_k\}_{k\ge 0} : \mathcal{H} \to \mathbb{R}$ be convex continuous functions and suppose that $\mathcal{C}_f$ is nonempty and bounded. Assume further that*

(i). $\limsup_{k\ge 1} \|s_{f_k} - s_f|_{\mathcal{U}} \le 0$, *for all bounded sets $\mathcal{U} \subset \mathcal{H}$; and*

(ii). $\limsup_{k\ge 1} \|f_k - f|_{C_f} \le 0$.

Then $\limsup_{k\ge 1} \|f_k - f|_{\mathcal{U}} \le 0$ *for all bounded sets $\mathcal{U} \subset \mathcal{H}$.*

Proof. Recalling from Theorem 3.1 the definition of $\mathcal{U}_r$, we observe that $\mathcal{U}_r$ is bounded. Our assumption can then be restated as follows:

$$\forall r > 0 : \; \limsup_{k\ge 1} \|s_{f_k} - s_f|_{\mathcal{U}_r} \le 0 \qquad \text{and} \qquad \limsup_{k\ge 1} \|f_k - f|_{\mathcal{C}_f} \le 0.$$

An application of Theorem 3.1 for each $r > 0$ completes the proof. □

A symmetric version of the corollary follows by an analogous argument.

Corollary 3.3 (Robust (two-sided) determination)**.** *Let $f, \{f_k\}_{k\ge 1} : \mathcal{H} \to \mathbb{R}$ be convex continuous functions such that*

$$\mathcal{C}_{f_k} \ne \emptyset,\ \forall k \ge 1 \qquad \textit{and} \qquad \mathcal{C} := \mathcal{C}_f \cup (\cup_{k\ge 1} \mathcal{C}_{f_k}) \ \textit{ is bounded.}$$

Assume further that:

(i). *s_{f_k} converge to s_f uniformly on bounded sets,*

(ii). *f_k converge to f uniformly on $\mathcal{C}$.*

Then f_k converge to f uniformly on bounded sets.

Remark 3.4 (open question). Our approach is heavily based on the existence of minimizers. We do not know if the results of this work can be extended to the class of lower semicontinuous convex functions, which are bounded for below. This is a challenging question that merits investigation.

Acknowledgements. A major part of this work has been accomplished during a research visit of the first author to the University of Washington. This author thanks the host institution for hospitality and acknowledges support from the Austrian Science Fund (FWF, P-36344-N).

References

[1] H. Attouch, G. Buttazzo, G. Michaille, *Variational Analysis in Sobolev and BV Spaces* (2nd ed), MPS-SIAM Series on Optimization (Philadelphia 2014).

[2] T. Boulmezaoud, P. Cieutat, A. Daniilidis, Gradient flows, second-order gradient systems and convexity, *SIAM J. Optim.* **28** (2018), 2049–2066.

[3] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publ. 1973.

[4] A. Daniilidis, J.M. Sepulcre, F. Venegas M, Asymmetric free spaces and canonical asymmetrizations, *Studia Math.* **261** (2021), 55–102.

[5] A. Daniilidis, D. Salas, A determination theorem in terms of the metric slope, *Proc. Amer. Math. Soc.* **150** (2022), 4325–4333.

[6] A. Daniilidis, G. David, E. Durand-Cartagena, A. Lemenant, Rectifiability of self-contracted curves in the Euclidean space and applications. *J. Geom. Anal.* **25** (2015), 1211–1239.

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

[8] M. Longinetti, P. Manselli, A. Venturi, On steepest descent curves for quasi convex families in $\mathbb{R}^n$, *Math. Nachr.* **288** (2015), 420–442.

[9] C. Gupta, S. Balakrishnan, A. Ramdas, Path length bounds for gradient descent and flow *J. Mach. Learn. Res.* **22** (2021), Paper No. 68, 63 pp.

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

[11] P. Perez-Aros, D. Salas, E. Vilches, Determination of convex functions via subgradients of minimal norm, *Math. Program.* **190** (2021), 561–583.

[12] L. Thibault, D. Zagrodny, Determining functions by slopes, *Communications in Contemporary Mathematics* (in press) `https://doi.org/10.1142/S0219199722500146`

Aris DANIILIDIS

Institute of Statistics and Mathematical Methods in Economics, VADOR E105-04
TU Wien, Wiedner Hauptstraße 8, A-1040 Wien

(on leave) DIM-CMM, CNRS IRL 2807
Beauchef 851, FCFM, Universidad de Chile

E-mail: aris.daniilidis@tuwien.ac.at
https://www.arisdaniilidis.at/

Research supported by the grants:

Austrian Science Fund (FWF P-36344N) (Austria)
CMM FB210005 BASAL funds for centers of excellence (ANID-Chile)

Dmitriy Drusvyatskiy

University of Washington
Department of Mathematics
C-138 Padelford, Seattle, WA 98195

E-mail: ddrusv@uw.edu
http://www.math.washington.edu/~ddrusv/.