



**HAL**  
open science

# Curiosities and counterexamples in smooth convex optimization

Jerome Bolte, Edouard Pauwels

► **To cite this version:**

Jerome Bolte, Edouard Pauwels. Curiosities and counterexamples in smooth convex optimization. Mathematical Programming, 2021, 10.1007/s10107-021-01707-1 . hal-02447733v2

**HAL Id: hal-02447733**

**<https://hal.science/hal-02447733v2>**

Submitted on 24 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Curiosities and counterexamples in smooth convex optimization

Jérôme Bolte\* and Edouard Pauwels†

Draft of January 24, 2020

## Abstract

Counterexamples to some old-standing optimization problems in the smooth convex coercive setting are provided. We show that block-coordinate, steepest descent with exact search or Bregman descent methods do not generally converge. Other failures of various desirable features are established: directional convergence of Cauchy's gradient curves, convergence of Newton's flow, finite length of Tikhonov path, convergence of central paths, or smooth Kurdyka-Łojasiewicz inequality. All examples are planar.

These examples are based on general smooth convex interpolation results. Given a decreasing sequence of positively curved  $C^k$  convex compact sets in the plane, we provide a level set interpolation of a  $C^k$  smooth convex function where  $k \geq 2$  is arbitrary. If the intersection is reduced to one point our interpolant has positive definite Hessian, otherwise it is positive definite out of the solution set. Furthermore, given a sequence of decreasing polygons we provide an interpolant agreeing with the vertices and whose gradients coincide with prescribed normals.

---

\*Toulouse School of Economics, Université Toulouse 1 Capitole, France.

†IRIT, ANITI, Université de Toulouse, CNRS. DEEL, IRT Saint Exupery, Toulouse, France.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Questions and method . . . . .	3
1.2	A digest of counterexamples . . . . .	4
<b>2</b>	<b>Preliminaries</b>	<b>6</b>
2.1	Continuous convex interpolations . . . . .	7
2.2	Bernstein approximation . . . . .	7
<b>3</b>	<b>Smooth convex interpolation</b>	<b>10</b>
3.1	Smooth parametrization of convex rings . . . . .	10
3.2	Smooth convex interpolation of smooth positively curved convex sequences	15
3.3	Considerations on Legendre functions . . . . .	19
<b>4</b>	<b>Smooth convex interpolation for sequences of polygons</b>	<b>23</b>
4.1	Smooth approximations of polygons . . . . .	24
4.2	Smooth convex interpolants of polygonal sequences . . . . .	27
4.3	More on Legendre functions and a pathological function with polyhedral domain . . . . .	28
<b>5</b>	<b>Counterexamples in continuous optimization</b>	<b>33</b>
5.1	Kurdyka-Łojasiewicz inequality may not hold . . . . .	33
5.2	Block coordinate descent may not converge . . . . .	34
5.3	Gradient descent with exact line search may not converge . . . . .	35
5.4	Tikhonov regularization path may have infinite length . . . . .	36
5.5	Secants of gradient curves at infinity may not converge . . . . .	37
5.6	Newton's flow may not converge . . . . .	41
5.7	Bregman descent (mirror descent) may not converge . . . . .	42
5.8	Central paths of Legendre barriers may not converge . . . . .	43
5.9	Hessian Riemannian gradient dynamics may not converge . . . . .	44
<b>6</b>	<b>Appendix</b>	<b>44</b>

# 1 Introduction

## 1.1 Questions and method

One of the goals of convex optimization is to provide a solution to a problem with stable and fast algorithms. The quality of a method is generally assessed by the convergence of sequences, rate estimates, complexity bounds, finite length of relevant quantities and other quantitative or qualitative ways.

Positive results in this direction are numerous and have been the object of intense research since decades. To name but a few: gradient methods e.g., [31, 32, 16], proximal methods e.g., [7], alternating methods e.g., [8, 39], path following methods e.g., [4, 33], Tikhonov regularization e.g. [24], semi-algebraic optimization e.g., [15, 13], decomposition methods e.g., [7, 8], augmented Lagrangian methods e.g., [11] and many others.

Despite this vast literature, some simple questions remain unanswered or just partly tackled, even for *smooth convex coercive* functions. Does the alternating minimization method, aka Gauss-Seidel method, converge? Does the steepest descent method with exact line search converge? Do mirror descent or Bregman methods converge? Does Newton's flow converge? Do central paths converge? Is the gradient flow directionally stable? Do smooth convex functions have the Kurdyka-Łojasiewicz property?

In this article we provide a negative answer to all these questions.

Our work draws inspiration from early works of de Finetti [22], Fenchel [21], on convex interpolation, but also from Torralba's PhD thesis [36] and the more recent [12], where some counterexamples on the Tikhonov path and Łojasiewicz inequalities are provided. The convex interpolation problem, see [22], is as follows: given a monotone sequence of convex sets<sup>1</sup> may we find a convex function interpolating each of these sets, i.e., having these sets as sublevel sets? Answers to these questions for *continuous* convex functions were provided by de Finetti, and improved by Fenchel [21], Kannai [25], and then used in [36, 12] for building counterexamples.

We improve this work by providing, for  $k \geq 2$  arbitrary, a general  $C^k$  interpolation theorem for positively curved convex sets, imposing at the same time the positive definiteness of its Hessian out of the solution set. An abridged version could be as follows.

**Theorem** (Smooth convex interpolation<sup>2</sup>). *Let  $(T_i)_{i \in \mathbb{Z}}$  be a sequence of compact convex subsets of  $\mathbb{R}^2$ , with positively curved  $C^k$  boundary, such that  $T_i \subset \text{int } T_{i+1}$  for all  $i$  in  $\mathbb{Z}$ . Then there exists a  $C^k$  convex function  $f$  having the  $T_i$  as sublevel sets with positive definite Hessian outside of the set:*

$$\text{argmin } f = \bigcap_{i \in \mathbb{Z}} T_i.$$

We provide several additional tools (derivatives computations) and variants (status of the solution set, Legendre functions, Lipschitz continuity). Whether our result is generalizable

---

<sup>1</sup>In the sense of sets inclusion the sequence being indexed on  $\mathbb{N}$  or  $\mathbb{Z}$

<sup>2</sup>See Theorem 2 for the full version.

to general smooth convex sequences, i.e., with vanishing curvature, seems to be a very delicate question whose answer might well be negative.

Our central theoretical result is complemented by a discrete approximate interpolation result “of order one” which is particularly well adapted for building counterexamples. Given a nested collection of polygons, one can indeed build a smooth convex function having level sets interpolating its vertices and whose gradients agree with prescribed normals.

Our results are obtained by blending parametrization techniques, Minkovski summation, Bernstein approximations and convex analysis.

As sketched below, our results offer the possibility of building counterexamples in convex optimization by restricting oneself to the construction of countable collections of nested convex sets satisfying some desirable properties. In all cases failures of good properties are caused by some curvature oscillations.

## 1.2 A digest of counterexamples

Counterexamples provided in this article can be classified along three axes: structural counterexamples<sup>3</sup>, counterexamples for convex optimization algorithms and ordinary differential equations.

In the following, the term “nonconverging” sequence or trajectory means, a sequence or a trajectory with at least two distinct accumulation points. Unless otherwise stated, convex functions have full domain.

*The following results are proved for  $C^k$  convex functions on the plane with  $k \geq 2$ .*

### Structural counterexamples

- (i) **Kurdyka-Łojasiewicz:** There exists a  $C^k$  convex function whose Hessian is positive definite outside its solution set and which does not satisfy the Kurdyka-Łojasiewicz inequality. This is an improvement on [12].
- (ii) **Tikhonov regularization path:** There exists a  $C^k$  convex function  $f$  such that the regularization path

$$x(r) = \operatorname{argmin} \{ f(y) + r\|y\|^2 : y \in \mathbb{R}^2 \}, \quad r \in (0, 1)$$

has infinite length. This strengthens a theorem by Torralba [36].

- (iii) **Central path:** There exists a continuous Legendre function  $h: [-1, 1]^2 \mapsto \mathbb{R}$ ,  $C^k$  on the interior, and  $c$  in  $\mathbb{R}^2$  such that the central path

$$x(r) = \operatorname{argmin} \{ \langle c, y \rangle + rh(y) : y \in D \}$$

does not have a limit as  $r \rightarrow 0$ .

---

<sup>3</sup>By structural, we include homotopic deformations by mere summation

**Algorithmic counterexamples:**

- (iv) **Gauss-Seidel method (block coordinate descent):** There exists a  $C^k$  convex function with positive definite Hessian outside its solution set and an initialization  $(u_0, v_0)$  in  $\mathbb{R}^2$ , such that the alternating minimization algorithm

$$\begin{aligned} u_{i+1} &= \operatorname{argmin}_{u \in \mathbb{R}} f(u, v_i) \\ v_{i+1} &= \operatorname{argmin}_{v \in \mathbb{R}} f(u_{i+1}, v) \end{aligned}$$

produces a bounded nonconverging sequence  $((u_i, v_i))_{i \in \mathbb{N}}$ .

- (v) **Gradient descent with exact line search:** There exists a  $C^k$  convex function  $f$  with positive definite Hessian outside its solution set and an initialization  $x_0$  in  $\mathbb{R}^2$ , such that the gradient descent algorithm with exact line search

$$x_{i+1} = \operatorname{argmin}_{t \in \mathbb{R}} f(x_i + t \nabla f(x_i))$$

produces a bounded nonconvergent sequence.

- (vi) **Bregman or mirror descent method:** There exists a continuous Legendre function  $h: [-1, 1]^2 \mapsto \mathbb{R}$ ,  $C^k$  on the interior, a vector  $c$  in  $\mathbb{R}^2$  and an initialization  $x_0$  in  $(-1, 1)^2$  such that the Bregman recursion

$$x_{i+1} = \nabla h^*(\nabla h(x_i) - c)$$

produces a nonconverging sequence. The couple  $(h, \langle c, \cdot \rangle)$  satisfies the smoothness properties introduced in [6].

**Continuous time ODE counterexamples:**

- (vii) **Continuous time Newton method:** There exists a  $C^k$  convex function with positive definite Hessian outside its solution set, and an initialization  $x_0$  in  $\mathbb{R}^2$  such that the continuous time Newton's system

$$\begin{aligned} \dot{x}(t) &= - [\nabla^2 f(x(t))]^{-1} \nabla f(x(t)), \quad t \geq 0, \\ x(0) &= x_0 \end{aligned}$$

has a solution approaching  $\operatorname{argmin} f$  which does not converge.

- (viii) **Directional convergence for gradient curves:** There exists a  $C^k$  convex function with 0 as a unique minimizer and a positive definite Hessian on  $\mathbb{R}^2 \setminus \{0\}$ , such that for any non stationary solution to the gradient system

$$\dot{x}(t) = -\nabla f(x(t))$$

the direction  $x(t)/\|x(t)\|$  does not converge.

- (ix) **Hessian Riemannian gradient dynamics:** There exists a continuous Legendre function  $h: [-1, 1]^2 \mapsto \mathbb{R}$ ,  $C^k$  on the interior, a linear function  $f$  and a nonconvergent solution to the following system

$$\dot{x}(t) = -\nabla_H f(x(t)),$$

where  $H = \nabla^2 h$  is the Hessian of  $h$  and  $\nabla_H f = H^{-1} \nabla f$  is the gradient of  $f$  in the Riemannian metric induced by  $H$  on  $(-1, 1)^2$ .

**Pathological sequences and curves** Our counterexamples lead to sequences or paths in  $\mathbb{R}^2$  which are related to a function  $f$  by a certain property (see examples above) and have a certain type of pathology. For illustration purposes, we provide sketches of the pathological behaviors we met in Figure 1.

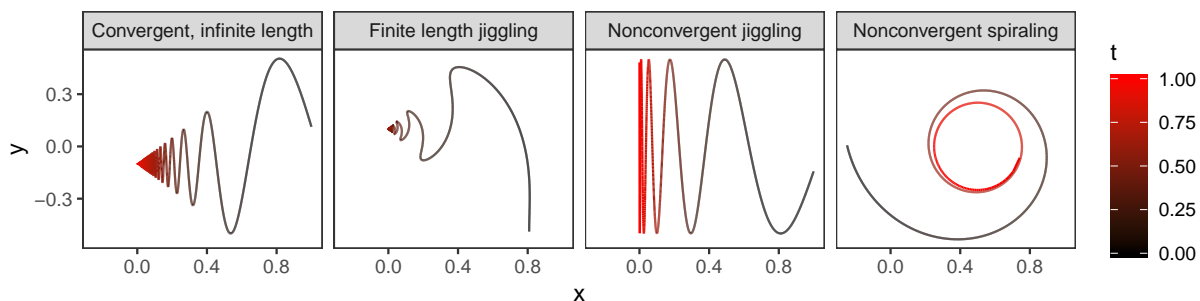


Figure 1: Rough sketches of pathological behavior for curves in the plane; for sequences similar figures would be obtained. Red colors indicate proximity with the solution set. Convergence with infinite length corresponds to counterexample (ii), finite length jiggling corresponds to counterexample (viii) (recall that gradient curves have to be self-contractant as well, see [19]), nonconvergent jiggling corresponds to counterexamples (iii), (vi), (vii), (ix) and nonconvergent spiraling corresponds to counterexamples (iv), (v) and somehow (i).

## 2 Preliminaries

Let us set  $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ . For  $p$  in  $\mathbb{N}^*$ , the Euclidean scalar product in  $\mathbb{R}^p$  is denoted by  $\langle \cdot, \cdot \rangle$ , otherwise stated the norm is denoted by  $\| \cdot \|$ . Given subsets  $S, T$  in  $\mathbb{R}^p$ , and  $x$  in  $\mathbb{R}^p$ , we define

$$\text{dist}(x, S) := \inf \{ \|x - y\| : y \in S \},$$

and the Hausdorff distance between  $S$  and  $T$ ,

$$\text{dist}(S, T) = \max \left( \sup_{x \in S} \text{dist}(x, T), \sup_{x \in T} \text{dist}(x, S) \right).$$

Throughout this note, the assertion “ $g$  is  $C^k$  on  $D$ ” where  $D$  is not an open set is to be understood as “ $g$  is  $C^k$  on an open neighborhood of  $D$ ”. Given a map  $G: X \mapsto A \times B$  for some space  $X$ ,  $[G]_1: X \mapsto A$  denotes the first component of  $G$ .

## 2.1 Continuous convex interpolations

We consider a sequence of compact convex subsets of  $\mathbb{R}^p$ ,  $(T_i)_{i \in \mathbb{Z}}$  such that  $T_{i+1} \subset \text{int } T_i$ . Finding a continuous convex interpolation of  $(T_i)_{i \in \mathbb{Z}}$  is finding a convex continuous function which makes the  $T_i$  a sequence of level sets. We call this process *continuous convex interpolation*. This questioning was present in Fenchel [21] and dates back to de Finetti [22], let us also mention the work of Crouzeix [18] revolving around this issue.

Such constructions have been shown to be realizable by Torralba [36], Kannai [25], using ideas based on Minkowski sum. The validity of this construction can be proved easily using the result of Crouzeix [18] which was already present under different and weaker forms in the works of de Finetti and Fenchel.

**Theorem 1** (de Finetti-Fenchel-Crouzeix). *Let  $f: \mathbb{R}^p \rightarrow \mathbb{R}$  be a quasiconvex function. The functions*

$$F_x: \lambda \mapsto \sup \{ \langle z, x \rangle : f(z) \leq \lambda \}$$

*are concave for all  $x$  in  $\mathbb{R}^p$ , if and only if  $f$  is convex.*

Our goal is to build smooth convex interpolation for sequences of smooth convex sets. To make such a construction we shall use *nonlinear Minkowski interpolation* between level sets.

We shall also rely on Bernstein approximation which we now describe.

## 2.2 Bernstein approximation

We refer to the monograph [28] by G.G. Lorentz.

The main properties of Bernstein polynomials to be used in this paper are the following:

- Bernstein approximation is linear in its functional argument  $f$  and “monotone” which allows to construct an approximation using only positive combination of a finite number of function values.
- There are precise formulas for derivatives of Bernstein approximation. They involve repeated finite differences. So approximating piecewise affine function with high enough degree leads to an approximation for which corner values of derivatives are controlled while the remaining derivatives are vanishing (up to a given order).
- Bernstein approximation is shape preserving, which means in particular that approximating a concave function preserves concavity.

The main idea to produce a smooth interpolation which preserves level sets is depicted in Figure 2 where we use Bernstein approximation to interpolate smoothly between three points and controlling the successive derivatives at the end points of the interpolation.



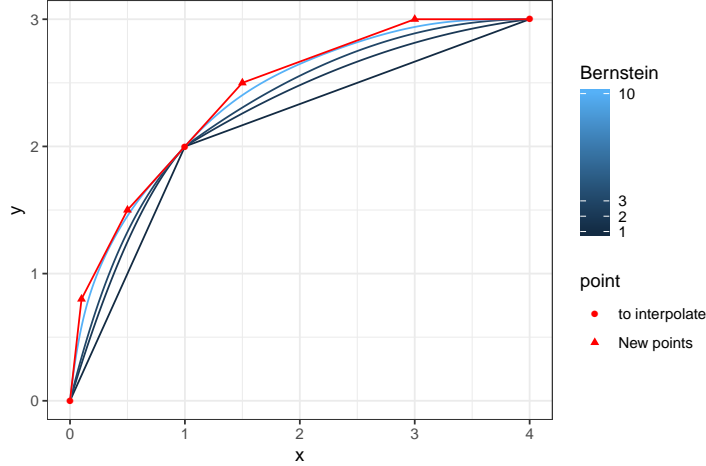


Figure 2: Illustration of Bernstein's smooth interpolation. We consider at first three points, to which we add four extra points to control derivatives at the junctions. We then build a concave, piecewise linear function and make a Bernstein approximation between the first two original points and the last two original points.

Let us now be specific. Given  $f$  defined on the interval  $[0, 1]$ , the *Bernstein polynomial* of order  $d \in \mathbb{N}^*$  associated to  $f$  is given by

$$B_d(x) = B_{d,f}(x) = \sum_{k=0}^d f\left(\frac{k}{d}\right) \binom{d}{k} x^k (1-x)^{d-k}, \text{ for } x \in [0, 1]. \quad (1)$$

**Derivatives and shape preservation:** For any  $h$  in  $(0, 1)$  and  $x$  in  $[0, 1-h]$ , we set  $\Delta_h^1 f(x) = f(x+h) - f(x)$  and recursively for all  $k$  in  $\mathbb{N}^*$ ,  $\Delta_h^k f(x) = \Delta(\Delta_h^{k-1} f(x))$ . We fix  $d \neq 0$  in  $\mathbb{N}$  and for  $h = \frac{1}{d}$  write  $\Delta_h^k = \Delta^k$ . Then for any  $m \leq d$ , we have

$$B_d(x)^{(m)} = d(d-1) \dots (d-m+1) \sum_{k=0}^{d-m} \Delta^m f\left(\frac{k}{d}\right) \binom{d-m}{k} x^k (1-x)^{d-k-m}, \quad (2)$$

for any  $x$  in  $[0, 1]$ . If  $f$  is increasing (resp. strictly increasing), then  $\Delta^1 f(x) \geq 0$  (resp.  $\Delta^1 f(x) > 0$ ) for all  $x$  and  $B_d'$  is positive (resp. strictly positive) and  $B_d$  is increasing (resp. strictly increasing). Similarly, if  $f$  is concave, then  $\Delta^2 f(x) \leq 0$  for all  $x$  so that  $B_d^{(2)} \leq 0$  and  $B_d$  is concave. From (2), we infer

$$|B_d(x)^{(m)}| \leq d(d-1) \dots (d-m+1) \sup_{k \in \{0, \dots, d-m\}} \left| \Delta^m f\left(\frac{k}{d}\right) \right| \quad (3)$$

for  $x$  in  $[0, 1]$ .

**Approximation of piecewise affine functions:** The following lemma will be extensively used throughout the proofs.

**Lemma 1** (Smoothing of piecewise lines in  $\mathbb{R}^p$ ). *Let  $q_0, q_1 \in \mathbb{R}^p$ ,  $\lambda_- < \lambda_0 < \lambda_1 < \lambda_+$  and  $0 < e_1, e_0 < 1$ . Set  $\Theta = (q_0, q_1, \lambda_-, \lambda_0, \lambda_1, \lambda_+, e_0, e_1)$  and define  $\gamma_\Theta: [0, 1] \mapsto \mathbb{R}^p$  through*

$$\gamma_\Theta(t) = \begin{cases} q_0 \left( 1 + \frac{e_0}{\lambda_0 - \lambda_-} (t(\lambda_+ - \lambda_-)) \right) & \text{if } 0 \leq t \leq \frac{\lambda_0 - \lambda_-}{\lambda_+ - \lambda_-} \\ q_0(1 + e_0) \left( \frac{\lambda_1 - \lambda_- - t(\lambda_+ - \lambda_-)}{\lambda_1 - \lambda_0} \right) + q_1(1 - e_1) \left( \frac{\lambda_- - \lambda_0 + t(\lambda_+ - \lambda_-)}{\lambda_1 - \lambda_0} \right) & \text{if } \frac{\lambda_0 - \lambda_-}{\lambda_+ - \lambda_-} \leq t \leq \frac{\lambda_1 - \lambda_-}{\lambda_+ - \lambda_-} \\ q_1 \left( 1 + \frac{(t-1)(\lambda_+ - \lambda_-)e_1}{\lambda_+ - \lambda_1} \right) & \text{if } \frac{\lambda_1 - \lambda_-}{\lambda_+ - \lambda_-} \leq t \leq 1. \end{cases}$$

The curve  $\gamma_\Theta$  in  $\mathbb{R}^{p+1}$  is the affine interpolant between the points  $q_0$ ,  $(1 + e_0)q_0$ ,  $(1 - e_1)q_1$  and  $q_1$ . For any  $m$  in  $\mathbb{N}$ , we choose  $d$  in  $\mathbb{N}^*$  such that

$$\frac{m}{d} \leq \min \left\{ \frac{\lambda_0 - \lambda_-}{\lambda_+ - \lambda_-}, 1 - \frac{\lambda_1 - \lambda_-}{\lambda_+ - \lambda_-} \right\}. \quad (4)$$

We consider a Bernstein-like reparametrization of  $\tilde{\gamma}_\Theta$  given by

$$\begin{aligned} \tilde{\gamma}_\Theta: [\lambda_-, \lambda_+] &\mapsto \mathbb{R}^p \\ \lambda &\mapsto \sum_{k=0}^d \tilde{\gamma}_\Theta \left( \frac{k}{d} \right) \binom{d}{k} \left( \frac{\lambda - \lambda_-}{\lambda_+ - \lambda_-} \right)^k \left( 1 - \frac{\lambda - \lambda_-}{\lambda_+ - \lambda_-} \right)^{d-k}. \end{aligned}$$

Then the following holds, for any  $2 \leq l \leq m$ ,  $\tilde{\gamma}_\Theta$  is  $C^m$  and

$$\begin{aligned} \tilde{\gamma}_\Theta(\lambda_-) &= q_0 & \tilde{\gamma}_\Theta(\lambda_+) &= q_1 \\ \tilde{\gamma}'_\Theta(\lambda_-) &= \frac{e_0}{\lambda_0 - \lambda_-} q_0 & \tilde{\gamma}'_\Theta(\lambda_+) &= \frac{e_1}{\lambda_+ - \lambda_1} q_1 \\ \tilde{\gamma}_\Theta^{(l)}(\lambda_-) &= 0 & \tilde{\gamma}_\Theta^{(l)}(\lambda_+) &= 0. \end{aligned}$$

Furthermore, if  $\gamma_\Theta$  has monotone coordinates (resp. strictly monotone, resp. concave, resp. convex), then so has  $\tilde{\gamma}_\Theta$ .

**Proof :** Note that the dependence of  $\tilde{\gamma}_\Theta$  in  $(q_0, q_1)$  is linear so that the dependence of  $\tilde{\gamma}_\Theta$  in  $(q_0, q_1)$  is also linear. Hence  $\tilde{\gamma}_\Theta$  is of the form  $\lambda \mapsto a(\lambda)q_0 + b(\lambda)q_1$ . We can restrict ourselves to  $p = 1$  since the general case follows from the univariate case applied coordinatewise.

If  $p = 1$ , then  $\tilde{\gamma}_\Theta = B_{f_{\Theta,d}} \circ A$ , where  $A: \lambda \mapsto \frac{\lambda - \lambda_-}{\lambda_+ - \lambda_-}$ . We have  $\tilde{\gamma}_\Theta(0) = q_0$ ,  $\tilde{\gamma}_\Theta(1) = q_1$  and  $\Delta^1 f_\Theta(0) = \frac{\lambda_+ - \lambda_-}{\lambda_0 - \lambda_-} \frac{e_0}{d} q_0$ ,  $\Delta^1 f_\Theta(1 - \frac{1}{d}) = \frac{\lambda_+ - \lambda_-}{\lambda_+ - \lambda_1} \frac{e_1}{d} q_1$  and  $\Delta^{(l)} f_\Theta(0) = \Delta^{(l)} f_\Theta(1 - \frac{1}{d}) = 0$ . The results follow from the expressions in (1) and (2) and the chain rule for  $\tilde{\gamma}_\Theta = B_{f_{\Theta,d}} \circ A$ .

The last property of  $\tilde{\gamma}_\Theta$  is due to the shape preserving property of Bernstein approximation and the fact that  $\tilde{\gamma}_\Theta = B_{f_{\Theta,d}} \circ L$ .  $\square$

**Remark 1.** (a) **[Affine image]** Using the notation of Lemma 1, if  $(\lambda_-, q_0)$ ,  $(\lambda_0, (1+e_0)q_0)$ ,  $(\lambda_1, (1-e_1)q_1)$  and  $(\lambda_+, q_1)$  are aligned, then the interpolation is actually an affine function. (b) **[Degree of the interpolants]** Observe that the degree of the Bernstein interpolant is connected to the slopes of the piecewise path  $\lambda$  by (4).

### 3 Smooth convex interpolation

Being given a subset  $S$  of  $\mathbb{R}^p$ , we denote by  $\text{int}(S)$  its interior,  $\bar{S}$  its closure and  $\text{bd}S = \bar{S} \setminus \text{int}(S)$  its boundary. Let us recall that the *support function* of  $S$  is defined through

$$\sigma_S(x) = \sup \{ \langle y, x \rangle : y \in S \} \in \mathbb{R} \cup \{+\infty\}.$$

#### 3.1 Smooth parametrization of convex rings

A *convex ring* is a set of the form  $C_1 \setminus C_2$  where  $C_1 \subset C_2$  are convex sets. Providing adequate parameterizations for such objects is key for interpolating  $C_1$  and  $C_2$  by some (regular) convex function.

The following assertion plays a fundamental role.

**Assumption 1.** *Let  $T_-, T_+ \subset \mathbb{R}^2$  be convex, compact with  $C^k$  boundary ( $k \geq 2$ ) and positive curvature. Assume that,  $T_- \subset \text{int}(T_+)$  and  $0 \in \text{int}(T_-)$ .*

The positive curvature assumption ensures that the boundaries can be parametrized by their normal, that is, for  $i = -, +$ , there exists

$$c_i: \mathbb{R}/2\pi\mathbb{Z} \mapsto \text{bd}(T_i)$$

such that the normal to  $T_i$  at  $c_i(\theta)$  is the vector  $n(\theta) = (\cos(\theta), \sin(\theta))^T$  and  $\dot{c}_i(\theta) = \rho_i(\theta)\tau(\theta)$  where  $\rho_i > 0$  and  $\tau(\theta) = (-\sin(\theta), \cos(\theta))$ . In this setting, it holds that  $c_i(\theta) = \text{argmax}_{y \in T_i} \langle n(\theta), y \rangle$ . The map  $c_i$  is the inverse of the Gauss map and is  $C^{k-1}$  (see [35] Section 2.5).

**Lemma 2** (Minkowski sum of convex sets with positive curvature). *Let  $T_-, T_+$  be as in Assumption 1 with normal parametrizations as above. For  $a, b \geq 0$  with  $a + b > 0$  set  $T = aT_- + bT_+$ .*

*Then  $T$  has positive curvature and its boundary is given by*

$$\text{bd}T = \{ a c_-(\theta) + b c_+(\theta) : \theta \in \mathbb{R}/2\pi\mathbb{Z} \},$$

*with the natural parametrization  $\mathbb{R}/2\pi\mathbb{Z} \ni \theta \rightarrow a c_-(\theta) + b c_+(\theta)$ .*

**Proof :** We may assume  $ab > 0$  otherwise the result is obvious. Let  $x$  be in  $\text{bd}T$  and denote by  $n(\theta)$  the normal vector at  $x$  for a well chosen  $\theta$ , so that  $x = \text{argmax} \{ \langle y, n(\theta) \rangle : y \in T \}$ . Observe that the definition of the Minkowski sum yields

$$\begin{aligned} \max_{y \in T} \langle y, n(\theta) \rangle &= \max_{(v,w) \in T_- \times T_+} \langle av + bw, n(\theta) \rangle \\ &= a \max_{v \in T_-} \langle v, n(\theta) \rangle + b \max_{w \in T_+} \langle w, n(\theta) \rangle \end{aligned}$$

so that

$$\begin{aligned} \langle x, n(\theta) \rangle &= a \langle c_-(\theta), n(\theta) \rangle + b \langle c_+(\theta), n(\theta) \rangle \\ &= \langle ac_-(\theta) + bc_+(\theta), n(\theta) \rangle \end{aligned}$$

which implies by extremality of  $x$  that  $x = ac_-(\theta) + bc_+(\theta)$ . Conversely, for any such  $x$ ,  $n(\theta)$  defines a supporting hyperplane to  $T$  and  $x$  must be on the boundary of  $T$ . The other results follow immediately.  $\square$

In the following fundamental proposition, we provide a smooth parametrization of the convex ring  $T_+ \setminus \text{int} T_-$ . The major difficulty is to control tightly the derivatives at the boundary so that the parametrizations can be glued afterward to build smooth interpolants.

**Proposition 1** ( $C^k$  parametrization of convex rings). *Let  $T_-, T_+$  be as in Assumption 1 with their normal parametrization as above. Fix  $k \geq 2$ ,  $\lambda_- < \lambda_0 < \lambda_1 < \lambda_+$  and  $0 < e_0, e_1 < 1$ . Choose  $d$  in  $\mathbb{N}^*$ , such that*

$$\frac{k}{d} \leq \min \left\{ \frac{\lambda_0 - \lambda_-}{\lambda_+ - \lambda_-}, 1 - \frac{\lambda_1 - \lambda_-}{\lambda_+ - \lambda_-} \right\}.$$

Consider the map

$$\begin{aligned} G: [\lambda_-, \lambda_+] \times \mathbb{R}/2\pi\mathbb{Z} &\mapsto \mathbb{R}^2 \\ (\lambda, \theta) &\mapsto \tilde{\gamma}_\Theta(\lambda) \end{aligned} \quad (5)$$

with  $\Theta = (c_-(\theta), c_+(\theta), \lambda_-, \lambda_0, \lambda_1, \lambda_+, e_0, e_1)$  and  $\tilde{\gamma}$  as given by Lemma 1. Assume further that:

( $\mathcal{M}$ ) For any  $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ ,  $\lambda \mapsto \langle G(\lambda, \theta), n(\theta) \rangle$  has strictly positive derivative on  $[\lambda_-, \lambda_+]$ .

Then the image of  $G$  is  $\mathcal{R} := T_+ \setminus \text{int}(T_-)$ ,  $G$  is  $C^k$  and satisfies, for any  $2 \leq l \leq k$  and any  $m$  in  $\mathbb{N}^*$ ,

$$\begin{aligned} \frac{\partial^m G}{\partial \theta^m}(\lambda_-, \theta) &= c_-^{(m)}(\theta) & \frac{\partial^m G}{\partial \theta^m}(\lambda_+, \theta) &= c_+^{(m)}(\theta) \\ \frac{\partial^{m+1} G}{\partial \lambda \partial \theta^m}(\lambda_-, \theta) &= c_-^{(m)}(\theta) \frac{e_0}{\lambda_0 - \lambda_-} & \frac{\partial^{m+1} G}{\partial \lambda \partial \theta^m}(\lambda_+, \theta) &= c_+^{(m)}(\theta) \frac{e_1}{\lambda_+ - \lambda_1} \\ \frac{\partial^{l+m} G}{\partial \lambda^l \partial \theta^m}(\lambda_-, \theta) &= 0 & \frac{\partial^{l+m} G}{\partial \lambda^l \partial \theta^m}(\lambda_+, \theta) &= 0. \end{aligned}$$

Besides  $G$  is a diffeomorphism from its domain on to its image. Set  $\mathcal{R} \ni x \mapsto (f(x), \theta(x))$  to be the inverse of  $G$ . Then  $f$  is  $C^k$  and in addition, for all  $x$  in  $\mathbb{R}$ ,

$$\begin{aligned} \nabla f(x) &= \frac{1}{\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \rangle} n(\theta(x)) \\ \nabla \theta(x) &= \frac{1}{\langle \frac{\partial G}{\partial \theta}(f(x), \theta(x)), \tau(\theta(x)) \rangle} \tau(\theta(x)) \\ &\quad - \frac{\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), \tau(\theta(x)) \rangle}{\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \rangle \langle \frac{\partial G}{\partial \theta}(f(x), \theta(x)), \tau(\theta(x)) \rangle} n(\theta(x)) \\ \nabla^2 f(x) &= \frac{\langle \frac{\partial G}{\partial \theta}(f(x), \theta(x)), \tau(\theta(x)) \rangle}{\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \rangle} \nabla \theta(x) \nabla \theta(x)^T \\ &\quad - \frac{\langle \frac{\partial^2 G}{\partial \lambda^2}(f(x), \theta(x)), n(\theta(x)) \rangle}{\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \rangle} \nabla f(x) \nabla f(x)^T, \end{aligned} \quad (6)$$

where all denominators are positive.

**Remark 2.** Note that  $G$  is actually well defined and smooth on an open set containing its domain. As we shall see it is also a diffeomorphism from an open set containing its domain onto its image.

**Proof :** Note that by construction, we have  $G(\lambda, \theta) = a(\lambda)c_-(\theta) + b(\lambda)c_+(\theta)$  for some polynomials  $a$  and  $b$  which are nonnegative on  $[\lambda_-, \lambda_+]$ . The formulas for the derivatives follow easily from this remark, the form of  $a$  and  $b$  and Lemma 1.

Set, for any  $\lambda$  in  $[\lambda_-, \lambda_+]$ ,  $T_\lambda = a(\lambda)T_- + b(\lambda)T_+$ . The resulting set  $T_\lambda$  is convex and has a positive curvature by Lemma 2, and for  $\lambda$  fixed  $G(\lambda, \cdot)$  is the inverse of the Gauss map of  $T_\lambda$ , which constitutes a parametrization by normals of the boundary.

Assume that  $\lambda < \lambda'$ , using the monotonicity assumption  $(\mathcal{M})$ , we have for any  $\theta, \theta'$ ,

$$\begin{aligned} \langle n(\theta'), G(\lambda, \theta) \rangle &\leq \sup_{y \in T_\lambda} \langle n(\theta'), y \rangle \\ &= \langle n(\theta'), G(\lambda, \theta') \rangle \\ &< \langle n(\theta'), G(\lambda', \theta') \rangle \end{aligned}$$

so that  $G(\lambda, \theta) \neq G(\lambda', \theta')$ . Furthermore, we have by definition of  $G(\lambda', \theta')$

$$G(\lambda, \theta) \in \bigcap_{\theta' \in \mathbb{R}/2\pi\mathbb{Z}} \{y, \langle y, n(\theta') \rangle \leq \langle n(\theta'), G(\lambda', \theta') \rangle\} = T_{\lambda'},$$

where the equality follows from the convexity of  $T_{\lambda'}$ . By convexity and compactness, this entails that  $T_\lambda = \text{conv}(\text{bd}(T_\lambda)) \subset \text{int}(T_{\lambda'})$ .

Let us show that the map  $G$  is bijective, first consider proving surjectivity. Let  $f$  be defined on  $T_+ \setminus \text{int}(T_-)$  through

$$f: x \mapsto \inf \{\lambda : \lambda \geq \lambda_-, x \in T_\lambda = a(\lambda)T_- + b(\lambda)T_+\}. \quad (7)$$

Since  $a(\lambda_+) = 0, b(\lambda_+) = 1$  this function is well defined and by compactness and continuity the infimum is achieved. It must hold that  $x$  belongs to  $\text{bd}(T_{f(x)})$ , indeed, if  $f(x) = \lambda_-$ , then  $x$  belongs to  $\text{bd}(T_-)$  and otherwise, if  $x$  is in  $\text{int}(T_{\lambda'})$  for  $\lambda' > \lambda_-$ , then  $f(x) < \lambda'$ . We deduce that  $x$  is of the form  $G(f(x), \theta)$  for a certain value of  $\theta$ , so that  $G$  is surjective.

As for injectivity, we have already seen a first case, the monotonicity assumption  $(\mathcal{M})$  ensures that  $\lambda \neq \lambda'$  implies  $G(\lambda, \theta) \neq G(\lambda', \theta')$  for any  $\theta, \theta'$ . Furthermore, we have the second case, for any  $\lambda$  in  $[\lambda_-, \lambda_+]$  and any  $\theta$ ,  $G(\lambda, \theta) = \arg \max_{y \in T_\lambda} \langle y, n(\theta) \rangle$  so that  $\theta \neq \theta'$  implies  $G(\lambda, \theta) \neq G(\lambda, \theta')$ . So in all cases,  $(\lambda, \theta) \neq (\lambda', \theta')$  implies that  $G(\lambda, \theta) \neq G(\lambda', \theta')$  and  $G$  is injective.

Let us now show that the map  $G$  is a local diffeomorphism by estimating its Jacobian map.

Since  $0 \in \text{int}(T_-)$ , we have for any  $\lambda, \theta$ ,

$$\begin{aligned} 0 &< \sup_{y \in T_-} \langle y, n(\theta) \rangle \\ &= \langle G(\lambda_-, \theta), n(\theta) \rangle \\ &\leq a(\lambda) \langle c_-(\theta), n(\theta) \rangle + b(\lambda) \langle c_+(\theta), n(\theta) \rangle, \end{aligned}$$

and both scalar products are positive so that  $a(\lambda) + b(\lambda) > 0$ . Hence, for any  $\theta$  in  $\mathbb{R}/2\pi\mathbb{Z}$ ,

$$\frac{\partial G}{\partial \theta}(\lambda, \theta) = (a(\lambda)\rho_-(\theta) + b(\lambda)\rho_+(\theta))\tau(\theta), \quad (8)$$

with  $a(\lambda)\rho_-(\theta) + b(\lambda)\rho_+(\theta) > 0$ . Furthermore by assumption  $\lambda \rightarrow \max_{y \in T_\lambda} \langle y, n(\theta) \rangle = \langle G(\lambda, \theta), n(\theta) \rangle$  has strictly positive derivative in  $\lambda$ , whence

$$\left\langle \frac{\partial G}{\partial \lambda}(\lambda, \theta), n(\theta) \right\rangle > 0.$$

We deduce that for any fixed  $\theta$ , in the basis  $(n(\theta), \tau(\theta))$ , the Jacobian of  $G$ , denoted  $J_G$ , is triangular with positive diagonal entries. More precisely fix  $\lambda, \theta$  and set  $x = G(\lambda, \theta)$  such that  $\lambda = f(x)$ ,  $\theta = \theta(x)$ . In the basis  $(n(\theta), \tau(\theta))$ , we deduce from (8) that the Jacobian of  $G$  is of the form

$$J_G(\lambda, \theta) = \begin{pmatrix} \alpha & 0 \\ \gamma & \beta \end{pmatrix},$$

where

$$\begin{aligned} \alpha &= \left\langle \frac{\partial G}{\partial \lambda}(\lambda, \theta), n(\theta) \right\rangle > 0 \\ \beta &= \left\langle \frac{\partial G}{\partial \theta}(\lambda, \theta), \tau(\theta) \right\rangle = a(\lambda)\rho_-(\theta) + b(\lambda)\rho_+(\theta) > 0 \\ \gamma &= \left\langle \frac{\partial G}{\partial \lambda}(\lambda, \theta), \tau(\theta) \right\rangle. \end{aligned}$$

It is thus invertible and we have a local diffeomorphism. We deduce that

$$J_G(\lambda, \theta)^{-1} = \begin{pmatrix} \alpha^{-1} & 0 \\ \frac{-\gamma}{\alpha\beta} & \beta^{-1} \end{pmatrix}.$$

We have  $J_G(\lambda, \theta)^{-1} = J_{G^{-1}}(x)$  so that the first line is  $\nabla f(x)$  and second line is  $\nabla \theta(x)$ , which proves the claimed expressions for gradients.

We also have  $dn(\theta)/d\theta = \tau(\theta)$  so that

$$J_{n \circ \theta}(x) = \tau(\theta(x))\nabla \theta(x)^T.$$

Differentiating the gradient expression, we obtain ( $\nabla$  denotes gradient with respect to  $x$ ):

$$\begin{aligned}
& \nabla^2 f(x) \\
&= \frac{\partial}{\partial x} \left( \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} n(\theta(x)) \right) \\
&= \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} J_{n \circ \theta}(x) \\
&\quad + n(\theta(x)) \nabla \left( \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} \right)^T \\
&= \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} J_{n \circ \theta}(x) \\
&\quad - n(\theta(x)) \nabla \left( \left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle \right)^T \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle^2} \\
&= \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} \left( \tau(\theta) \nabla \theta(x)^T - \nabla f(x) \nabla \left( \left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle \right)^T \right).
\end{aligned}$$

We have

$$\begin{aligned}
&= \nabla \left( \left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle \right)^T \\
&= \frac{\partial G}{\partial \lambda}(f(x), \theta(x))^T J_{n \circ \theta}(x) + n(\theta(x))^T J_{\frac{\partial G}{\partial \lambda}}(f(x), \theta(x)) J_{G^{-1}}(x) \\
&= \left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), \tau(\theta(x)) \right\rangle \nabla \theta(x)^T + \left\langle n(\theta(x)), \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta) \right\rangle \nabla f(x)^T,
\end{aligned}$$

where, for the last identity, we have used the fact that

$$\begin{aligned}
n(\theta(x))^T J_{\frac{\partial G}{\partial \lambda}}(f(x), \theta(x)) &= \left\langle n(\theta(x)), \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta) \right\rangle n(\theta(x))^T + \left\langle n(\theta(x)), \frac{\partial^2 G}{\partial \lambda \partial \theta}(\lambda, \theta) \right\rangle \tau(\theta(x))^T \\
&= \left\langle n(\theta(x)), \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta) \right\rangle n(\theta(x))^T \\
n(\theta(x))^T J_{G^{-1}}(x) &= n(\theta(x))^T J_G(f(x), \theta(x))^{-1} \\
&= \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} n(\theta(x))^T = \nabla f(x)^T.
\end{aligned}$$

We deduce that

$$\begin{aligned}
& \nabla^2 f(x) \\
&= \frac{1}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} \\
&\quad \left( \tau(\theta(x)) \nabla \theta(x)^T - \left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), \tau(\theta(x)) \right\rangle \nabla f(x) \nabla \theta(x)^T \right. \\
&\quad \left. - \nabla f(x) \nabla f(x)^T \left\langle n(\theta(x)), \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta) \right\rangle \right).
\end{aligned}$$

We have

$$\begin{aligned}
& \tau(\theta(x))\nabla\theta(x)^T - \left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), \tau(\theta(x)) \right\rangle \nabla f(x)\nabla\theta(x)^T \\
&= \left( \tau(\theta(x)) - \frac{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), \tau(\theta(x)) \right\rangle}{\left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle} n(\theta(x)) \right) \nabla\theta(x)^T \\
&= (a(\lambda)\rho_-(\theta) + b(\lambda)\rho_+(\theta))\nabla\theta(x)\nabla\theta(x)^T
\end{aligned}$$

So that we actually get

$$\begin{aligned}
& \nabla^2 f(x) \left\langle \frac{\partial G}{\partial \lambda}(f(x), \theta(x)), n(\theta(x)) \right\rangle \\
&= (a(\lambda)\rho_-(\theta) + b(\lambda)\rho_+(\theta))\nabla\theta(x)\nabla\theta(x)^T - \left\langle n(\theta(x)), \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta) \right\rangle \nabla f(x)\nabla f(x)^T \\
&= \left\langle \frac{\partial G}{\partial \theta}(\lambda, \theta), \tau(\theta(x)) \right\rangle \nabla\theta(x)\nabla\theta(x)^T - \left\langle n(\theta(x)), \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta) \right\rangle \nabla f(x)\nabla f(x)^T.
\end{aligned}$$

This concludes the proof.  $\square$

### 3.2 Smooth convex interpolation of smooth positively curved convex sequences

In this section, we consider an indexing set  $I$  with either  $I = \mathbb{N}$  or  $I = \mathbb{Z}$ , and an increasing sequence of compact convex sets  $(T_i)_{i \in I}$  such that for any  $i$  in  $I$ , the couple  $T_+ := T_{i+1}$ ,  $T_- := T_i$  satisfies Assumption 1. In particular, for each  $i$  in  $I$ ,  $T_i$  is compact, convex with  $C^k$  boundary and positive curvature. We denote by  $c_i$  the corresponding parametrization by the normal,  $T_i \subset \text{int } T_{i+1}$ . With no loss of generality we assume  $0 \in \cap_{i \in I} T_i$ .

This is our main theoretical result.

**Theorem 2** (Smooth convex interpolation). *Let  $I = \mathbb{N}$  or  $I = \mathbb{Z}$  and  $(T_i)_{i \in I}$  such that for any  $i \in I$ ,  $T_i \subset \mathbb{R}^2$  and the couple  $T_+ := T_{i+1}$ ,  $T_- := T_i$  satisfies Assumption 1. Then there exists a  $C^k$  convex function*

$$f: \mathcal{T} := \text{int} \left( \bigcup_{i \in I} T_i \right) \mapsto \mathbb{R}$$

such that

(i)  $T_i$  is a sublevel set of  $f$  for all  $i$  in  $I$ .

(ii) We have

$$\text{argmin } f = \begin{cases} \bigcap_I T_i & \text{if } I = \mathbb{Z} \\ \{0\} & \text{if } I = \mathbb{N}. \end{cases}$$

(iii)  $\nabla^2 f$  is positive definite on  $\mathcal{T} \setminus \text{argmin } f$ , and if  $I = \mathbb{N}$ , it is positive definite throughout  $\mathcal{T}$ .



**Proof :**

*Preconditionning.* We have that  $0 \in \cap_{i \in I} \text{int}(T_i)$ . Hence for any  $i$  in  $I$  and  $0 \leq \alpha < 1$ ,  $\alpha T_i \in \text{int}(T_i)$ . Furthermore, for  $\alpha > 0$ , small enough,  $(1 + \alpha)T_i \subset (1 - \alpha)\text{int}(T_{i+1})$ . Set  $\alpha_0$  such that  $(1 + \alpha_0)T_0 \subset (1 - \alpha_0)\text{int}(T_1)$ . By forward (and backward if  $I = \mathbb{Z}$ ) induction, for all  $i$  in  $I$ , we obtain  $\alpha_i > 0$  such that

$$(1 + \alpha_i)T_i \subset (1 - \alpha_i)\text{int}(T_{i+1}).$$

Setting for all  $i$  in  $I$ ,  $\epsilon_{i+1} = \min\{\alpha_i, \alpha_{i+1}\}$  ( $\epsilon_0 = \alpha_0$  if  $I = \mathbb{N}$ ), we have for all  $i$  in  $I$

$$(1 + \epsilon_i)T_i \subset (1 + \alpha_i)T_i \subset (1 - \alpha_i)\text{int}(T_{i+1}) \subset (1 - \epsilon_{i+1})\text{int}(T_{i+1}).$$

For all  $i$  in  $I$ , we introduce

$$\begin{aligned} S_{3i} &= T_i, \\ S_{3i+1} &= (1 + \epsilon_i)T_i \\ S_{3i+2} &= (1 - \epsilon_{i+1})T_{i+1}. \end{aligned}$$

We have a new sequence of strictly increasing compact convex sets  $(S_i)_{i \in I}$ .

*Value assignation.* For each  $i$  in  $I$ , we set

$$K_i = \max_{\|x\|=1} \frac{\sigma_{S_{i+1}}(x) - \sigma_{S_i}(x)}{\sigma_{S_i}(x) - \sigma_{S_{i-1}}(x)} \in (0, +\infty).$$

Note that for all  $i$  in  $I$ ,  $K_{3i} = 1$ . We choose  $\lambda_1 = 2$ ,  $\lambda_0 = 1$  and for all  $i$  in  $I$ ,

$$\lambda_{i+1} = \lambda_i + K_i(\lambda_i - \lambda_{i-1}). \quad (9)$$

By construction, we have for all  $i$  in  $I$  and all  $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ ,

$$\frac{\sigma_{S_{i+1}}(n(\theta)) - \sigma_{S_i}(n(\theta))}{\lambda_{i+1} - \lambda_i} \leq \frac{\sigma_{S_i}(n(\theta)) - \sigma_{S_{i-1}}(n(\theta))}{\lambda_i - \lambda_{i-1}}.$$

If  $I = \mathbb{Z}$ , this entails

$$0 < \lambda_i - \lambda_{i-1} \leq \frac{\lambda_1 - \lambda_0}{\sigma_{S_1}(n(\theta)) - \sigma_{S_0}(n(\theta))} (\sigma_{S_i}(n(\theta)) - \sigma_{S_{i-1}}(n(\theta))),$$

and the right-hand side is summable over negative indices  $i \leq 0$ , so that  $\lambda_i \rightarrow \underline{\lambda} \in \mathbb{R}$  as  $i \rightarrow -\infty$ . In all cases  $(\lambda_i)_{i \in I}$  is an increasing sequence bounded from below.

*Local interpolation.* We fix  $i$  in  $I$  and consider the function  $G_i$  described in Proposition 1 with  $T_+ = S_{3i+3} = T_{i+1}$ ,  $T_- = S_{3i} = T_i$ ,  $\lambda_+ = \lambda_{3i+3}$ ,  $\lambda_1 = \lambda_{3i+2}$ ,  $\lambda_0 = \lambda_{3i+1}$ ,  $\lambda_- = \lambda_{3i}$ ,  $e_0 = \epsilon_i$ ,  $e_1 = \epsilon_{i+1}$ . By linearity, we have for any  $(\lambda, \theta) \in [\lambda_-, \lambda_+] \times \mathbb{R}/2\pi\mathbb{Z}$ ,

$$\langle G_i(\lambda, \theta), n(\theta) \rangle = \tilde{\gamma}_\Theta(\lambda)$$

where  $\tilde{\gamma}_\Theta$  is as in Lemma 1 with input data  $q_0 = \langle c_{3i}(\theta), n(\theta) \rangle = \sigma_{S_{3i}}(n(\theta))$ ,  $q_1 = \langle c_{3i+3}(\theta), n(\theta) \rangle = \sigma_{S_{3i+3}}(n(\theta))$ , and  $\lambda_-, \lambda_0, \lambda_1, \lambda_+, e_1, e_0$  as already described. This corresponds to the Bernstein approximation of the piecewise affine interpolation between the points

$$\begin{aligned} & (\lambda_{3i}, \sigma_{S_{3i}}(n(\theta))) \\ & (\lambda_{3i+1}, \sigma_{S_{3i+1}}(n(\theta))), \\ & (\lambda_{3i+2}, \sigma_{S_{3i+2}}(n(\theta))), \\ & (\lambda_{3i+3}, \sigma_{S_{3i+3}}(n(\theta))), \end{aligned} \tag{10}$$

By construction of  $(K_i)_{i \in I}$ , we have for all  $\theta$ ,

$$0 < \frac{\sigma_{S_{3i+3}}(n(\theta)) - \sigma_{S_{3i+2}}(n(\theta))}{\lambda_{3i+3} - \lambda_{3i+2}} \leq \frac{\sigma_{S_{3i+2}}(n(\theta)) - \sigma_{S_{3i+1}}(n(\theta))}{\lambda_{3i+2} - \lambda_{3i+1}} \leq \frac{\sigma_{S_{3i+1}}(n(\theta)) - \sigma_{S_{3i}}(n(\theta))}{\lambda_{3i+1} - \lambda_{3i}}.$$

Whence the affine interpolant between points in (10) is strictly increasing and concave, and by using the shape preserving properties of Bernstein polynomials,  $\langle G_i(\lambda, \theta), n(\theta) \rangle$  has strictly positive derivative. As a consequence  $G_i$  is a diffeomorphism and its derivatives are as in Proposition 1. Furthermore

$$\lambda \mapsto \langle G_i(\lambda, \theta), n(\theta) \rangle$$

is a  $C^k$  concave function of  $\lambda$ .

*Global interpolation.* Recall that  $\underline{\lambda} = \inf_{i \in I} \lambda_i > -\infty$  and set  $\bar{\lambda} = \sup_{i \in I} \lambda_i \in (-\infty, +\infty]$ . For any  $\lambda \in (\underline{\lambda}, \bar{\lambda})$ , there exists a unique  $i_\lambda \in I$  such that  $\lambda \in [\lambda_{3i_\lambda}, \lambda_{3i_\lambda+3})$ . Define

$$\begin{aligned} G: (\underline{\lambda}, \bar{\lambda}) \times \mathbb{R}/2\pi\mathbb{Z} &\mapsto \mathbb{R}^2 \\ (\lambda, \theta) &\mapsto G_{i_\lambda}(\lambda, \theta). \end{aligned}$$

Fix  $i$  in  $I$ . The boundary of  $T_{i+1}$  is given by  $G_{i+1}(\lambda_{3i+3}, \mathbb{R}/2\pi\mathbb{Z}) = G_i(\lambda_{3i+3}, \mathbb{R}/2\pi\mathbb{Z})$  with actually

$$G_{i+1}(\lambda_{3i+3}, \theta) = G_i(\lambda_{3i+3}, \theta), \text{ for all } \theta \text{ in } \mathbb{R}/2\pi\mathbb{Z}. \tag{11}$$

Since  $K_{3i} = 1$ , we have

$$\lambda_{3i+1} - \lambda_{3i} = \lambda_{3i} - \lambda_{3i-1}.$$

The expressions of the derivatives in Proposition 1 and (11) ensure that the derivatives of  $G_{i+1}$  and  $G_i$  agree on  $\lambda_{3i+3} \times \mathbb{R}/2\pi\mathbb{Z}$  up to order  $k$ . Hence  $G$  is a local diffeomorphism. Bijectivity of each  $G_i$  ensure that  $G$  is also bijective and thus  $G$  is a diffeomorphism. Furthermore

$$\lambda \mapsto \langle G(\lambda, \theta), n(\theta) \rangle$$

is  $C^k$  piecewise concave and thus concave.

*Extending  $G$ .* If  $I = \mathbb{N}$ , we may assume without loss of generality that  $S_0 = B$  the Euclidean ball and  $S_1 = 5/3S_0$ , which corresponds to  $\epsilon_0 = 2/3$ , eventually after adding

a set in the list and rescaling. Let  $\phi$  denote the function described in Lemma 10 and  $G_{-1}$  be described as in Lemma 11. This allows to extend  $G$  for  $\lambda \in [0, 1]$ ,  $G$  is then  $C^k$  on  $(0, \bar{\lambda}) \times \mathbb{R}/2\pi\mathbb{Z}$  by using Lemma 11 and Proposition 1. This does not affect the differentiability, monotonicity and concavity properties of  $G$ .

*Defining the interpolant  $f$ .* We assume without loss of generality that  $\underline{\lambda} = 0$ . We set  $f$  to be the first component of the inverse of  $G$  so that it is defined on  $G^{-1}((0, \bar{\lambda}) \times \mathbb{R}/2\pi\mathbb{Z})$ . We extend  $f$  as follows:

- $f(0) = 0$  if  $I = \mathbb{N}$ ,
- $f = 0$  on  $\cap_{i \in I} T_i$  if  $I = \mathbb{Z}$ .

Since  $G$  is  $C^k$  and non-singular on  $(0, \bar{\lambda}) \times \mathbb{R}/2\pi\mathbb{Z}$ , the inverse mapping theorem ensures that  $f$  is  $C^k$  on  $\text{int}(\mathcal{T}) \setminus \arg \min_{\mathcal{T}} f$ .

*Convexity of  $f$ .* For any  $\theta$  in  $\mathbb{R}/2\pi\mathbb{Z}$ ,

$$(0, \bar{\lambda}) \mapsto \mathbb{R}_+$$

$$\lambda \mapsto \sup_{z \in [f \leq \lambda]} n(\theta)^T z$$

is equal to  $\langle G(\lambda, \theta), n(\theta) \rangle$  which is concave. It can be extended at  $\lambda = 0$  by continuity. This preserves concavity hence, using Theorem 1, we have proved that  $f$  is convex and  $C^k$  on  $\mathcal{T} \setminus \arg \min_{\mathcal{T}} f$ .

*Smoothness around the argmin and Hessian positivity.* If  $I = \mathbb{N}$ , then the interpolant defined in Lemma 11 ensures that  $f$  is proportional to the norm squared around 0. Hence it is  $C^k$  around 0 with positive definite Hessian. We may compose  $f$  with the function  $g: t \mapsto \sqrt{t^2 + 1} + t$  which is increasing and has positive second derivative. This ensures that the resulting Hessian is positive definite outside  $\arg \min f$  and thus everywhere since

$$\nabla^2 g \circ f = g' \nabla^2 f + g'' \nabla f \nabla f^T$$

is positive definite thanks to the expressions for the Hessian of  $f$  in Proposition 1

If  $I = \mathbb{Z}$ , we let all the derivatives of  $f$  vanish around the solution set. The smoothing Lemma 14 applies and provides a function  $\phi$  with positive derivative on  $(0, +\infty)$ , such that  $\phi \circ f$  is convex,  $C^k$  with prescribed sublevel sets. Furthermore, we remark that

$$\nabla^2 \phi \circ f = \phi' \nabla^2 f + \phi'' \nabla f \nabla f^T.$$

We may compose  $\phi \circ f$  with the function  $g: t \mapsto \sqrt{t^2 + 1} + t$  which is increasing with positive second derivative, the expressions for the Hessian of  $f$  in Proposition 1 ensure that the resulting Hessian is positive definite out of  $\arg \min f$ .  $\square$

**Remark 3** (Interpolation of symmetric rings). In view of Remark 1, if we have  $T_{i+1} = \alpha T_i$  for some  $0 < \alpha < 1$  and  $i$  in  $\mathbb{Z}$ , then the interpolated level sets between  $T_i$  and  $T_{i+1}$  are all of the form  $sT_i$  for  $\alpha \leq s \leq 1$ .

**Remark 4** (Strict convexity). Recall that strict convexity of a differentiable function amounts to the injectivity of its gradient. In Theorem 2 if there is a unique minimizer, then the invertibility of the Hessian outside  $\arg \min f$  ensures that our interpolant is strictly convex (note that this is automatically the case if  $I = \mathbb{N}$ ).

### 3.3 Considerations on Legendre functions

The following proposition provides some interpolant with additional properties as global Lipschitz continuity and finiteness properties for the dual function. At this stage these results appear as merely technical but they happen to be decisive in the construction of counterexamples involving Legendre functions. The properties of Legendre functions can be found in [37, Chapter 6]. We simply recall here that, given a convex body  $C$  of  $\mathbb{R}^p$ , a convex function  $h : C \rightarrow \mathbb{R}$  is *Legendre* if it is differentiable on  $\text{int } C$  and if  $\nabla h$  defines a bijection from  $\text{int } C$  to  $\nabla h(\text{int } C)$  with in addition

$$\lim_{\substack{x \in \text{int } C \\ x \rightarrow z}} \|\nabla h(x)\| = +\infty,$$

for all  $z$  in  $\text{bd } C$ . We also assume that  $\text{epi } f := \{(x, \lambda) : f(x) \leq \lambda\}$  is closed in  $\mathbb{R}^{p+1}$ . The *Legendre conjugate* or *dual function* of  $h$  is defined through

$$h^*(z) = \sup \{ \langle z, x \rangle - h(x) : x \in C \},$$

for  $z$  in  $\mathbb{R}^p$ , and its domain is  $D := \{z \in \mathbb{R}^p : h^*(z) < +\infty\}$ . The function  $h^*$  is differentiable on the interior of  $D$ , and the inverse of  $\nabla h : \text{int } C \rightarrow \text{int } D$  is  $\nabla h^* : \text{int } D \rightarrow \text{int } C$ .

We start with a simple technical lemma on the compactness of the domain of a Legendre function.

**Lemma 3.** *Let  $h : \mathbb{R}^2 \mapsto \mathbb{R}$  be a globally Lipschitz continuous Legendre function, and set  $D = \text{int}(\text{dom}(h^*))$  where  $h^* : \mathbb{R}^2 \mapsto \mathbb{R}$  is the conjugate of  $h$ . For each  $\lambda \geq \min_{\mathbb{R}^2} h$  let  $\sigma_\lambda$  be the support function associated to the set  $\{z \in \mathbb{R}^2, h(z) \leq \lambda\}$ . The following are equivalent*

- (i)  $h^*(x) \leq 0$  for all  $x \in D$ .
- (ii) For all  $y \in \mathbb{R}^2$ ,  $\sigma_{h(y)}(\nabla h(y)) \leq h(y)$ .

In both cases  $h^*$  has compact domain.

**Proof :** Let us establish beforehand the following formula

$$h^*(z) = \sigma_{h(y)}(\nabla h(y)) - h(y), \tag{12}$$

with  $y = \nabla h^*(z)$  and  $y \in \mathbb{R}^2$ . Since  $h^*$  is Legendre, we have for all  $y$  in  $D$ ,

$$h^*(z) = \sup_{y \in \mathbb{R}^2} \langle z, y \rangle - h(y) = \langle z, \nabla h^*(z) \rangle - h(\nabla h^*(z)).$$

We have, setting  $y = \nabla h^*(z)$

$$\langle z, \nabla h^*(z) \rangle = \langle \nabla h(y), y \rangle = \sigma_{h(y)}(\nabla h(y))$$

because  $\nabla h(y)$  is normal to the sublevel set of  $h$  which contains  $y$  in its boundary. Hence we have  $h^*(z) = \sigma_{h(y)}(\nabla h(y)) - h(y)$  with  $y = \nabla h^*(z)$ , that is (12) holds. Since  $\nabla h^* : D \mapsto \mathbb{R}^2$  is a bijection, the equivalence follows. In this case the domain of  $h^*$  is closed because  $h^*$  is bounded and lower semicontinuous. The domain is also bounded by the Lipschitz continuity of  $h$ , whence compact.  $\square$

**Proposition 2** (On Legendre interpolation). *Let  $(S_i)_{i \in \mathbb{N}}$  be such that for any  $i$  in  $I$ ,  $T_- = S_i$ ,  $T_+ = S_{i+1}$  satisfy Assumption 1 and there exists a sequence  $(\epsilon_i)_{i \in \mathbb{N}}$  in  $(0, 1)$  such that for all  $i \geq 1$ ,  $(1 - \epsilon_i)^{-1} S_{3i-1} = (1 + \epsilon_i)^{-1} S_{3i+1} = S_{3i}$ . Assume in addition that,*

$$\inf_{\|x\|=1} \sigma_{S_i}(x) - \sigma_{S_{i-1}}(x) = 1 + O\left(\frac{1}{i^3}\right) \quad (\text{non degeneracy}), \quad (13)$$

$$\sup_{\|x\|=1} \left| \frac{\sigma_{S_{i+1}}(x) - \sigma_{S_i}(x)}{\sigma_{S_i}(x) - \sigma_{S_{i-1}}(x)} - 1 \right| = O\left(\frac{1}{i^3}\right) \quad (\text{moderate growth}). \quad (14)$$

Then there exists a convex  $C^k$  function  $h: \mathbb{R}^2 \mapsto \mathbb{R}$ , such that

- For all  $i$  in  $\mathbb{N}$ ,  $S_{3i}$  is a sublevel set of  $h$ ,
- $h$  has positive definite Hessian,
- $h$  is globally Lipschitz continuous,
- $h^*$  has a compact domain  $D$  and is  $C^k$  and strictly convex on  $\text{int}(D)$ .

**Proof :** The construction of  $h$  follows the exact same principle as that of Theorem 2. This ensures that the first two points are valid. Note that equation (13) implies that the sets sequence grows by at least a fixed amount in each direction as  $i$  grows. Hence we have  $\mathcal{T} = \mathbb{R}^2$ .

*Global Lipschitz continuity of  $h$ :* The values of  $h$  are defined through

$$\begin{aligned} \lambda_{i+1} - \lambda_i &= K_i(\lambda_i - \lambda_{i-1}), \quad \forall i \in \mathbb{N}^* \\ K_i &= \max_{\|x\|=1} \frac{\sigma_{S_{i+1}}(x) - \sigma_{S_i}(x)}{\sigma_{S_i}(x) - \sigma_{S_{i-1}}(x)} \in (0, +\infty), \end{aligned}$$

so that  $K_i = 1 + O(1/i^3)$  thanks to equation (14). Note that the moderate growth assumption entails

$$\sup_{i \in \mathbb{N}} \sigma_{S_{i+1}}(x) - \sigma_{S_i}(x) = O\left(\prod_{i \in \mathbb{N}^*} K_i\right) = O(1). \quad (15)$$

For  $i \geq 1$ , one has

$$\lambda_{i+1} - \lambda_i = \prod_{1 \leq j \leq i} K_j(\lambda_1 - \lambda_0). \quad (16)$$

On the other hand using the bounds (14), (13) and the identity (16), there exists a constant  $\kappa > 0$  such that for all  $i \geq 1$ , all  $\theta$  in  $\mathbb{R}/2\pi\mathbb{Z}$ ,

$$\frac{\sigma_{S_{i+1}}(n(\theta)) - \sigma_{S_i}(n(\theta))}{\lambda_{i+1} - \lambda_i} = \frac{(\sigma_{S_{i+1}}(n(\theta)) - \sigma_{S_i}(n(\theta)))}{\prod_{j=1}^i K_j(\lambda_1 - \lambda_0)} \geq \kappa > 0. \quad (17)$$

By the interpolation properties described in Lemma 1 the function  $\langle G(\lambda, \theta), n(\theta) \rangle$  constructed in Theorem 2 has derivative with respect to  $\lambda$  greater than  $\kappa$ . Recalling the

expression of the gradient as given in Proposition 1 (and the concavity of  $G$  with respect to  $\lambda$ ), this shows that

$$\|\nabla h(x)\| \leq \frac{1}{\kappa}$$

for all  $x$  in  $\mathbb{R}^2$ , and by the mean value theorem,  $h$  is globally Lipschitz continuous on  $\mathbb{R}^2$ .

*Properties of the dual function:*  $h$  is Legendre, its conjugate  $h^*$  is therefore Legendre. From the definiteness of  $\nabla^2 h$  and the fact that  $\nabla h: \mathbb{R}^2 \mapsto \text{int}(D)$  is a bijection, we deduce that  $h^*$  is  $C^k$  by the inverse mapping theorem. So the only property which we need to establish is that  $h^*$  has a compact domain, in other words, using Lemma 3, it is sufficient to show that  $\sup_{x \in \text{int}D} h^*(x) \leq 0$ .

Using the notation of the proof of Theorem 2, we will show that it is possible to verify that, for all  $\lambda, \theta$  in the domain of  $G$

$$\frac{\langle n(\theta), G(\lambda, \theta) \rangle}{\left\langle \frac{\partial G}{\partial \lambda}(\lambda, \theta), n(\theta) \right\rangle} \leq \lambda. \quad (18)$$

Equation (18) is indeed the coordinate form of the characterization given in Lemma 3. Let us observe that

$$\frac{\partial}{\partial \lambda} \left( \langle n(\theta), G(\lambda, \theta) \rangle - \lambda \left\langle \frac{\partial G}{\partial \lambda}(\lambda, \theta), n(\theta) \right\rangle \right) = -\lambda \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle, \quad (19)$$

and since  $G$  is concave, the right hand side is positive.

Assume that we have proved that,

$$\lambda \mapsto \sup_{\theta} -\lambda \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle \quad (20)$$

has finite integral as  $\lambda \rightarrow \infty$ . Since the function

$$\theta \mapsto \lambda \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle \quad (21)$$

is continuous on  $\mathbb{R}/2\pi\mathbb{Z}$  for any  $\lambda$ , Lebesgue dominated convergence theorem would ensure that

$$\theta \mapsto \int_{\lambda \geq \lambda_0} -\lambda \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle d\lambda$$

is continuous in  $\theta$ , so that:

$$\sup_{\theta} \left[ \lim_{\lambda \rightarrow \infty} \langle n(\theta), G(\lambda, \theta) \rangle - \lambda \left\langle \frac{\partial G}{\partial \lambda}(\lambda, \theta), n(\theta) \right\rangle \right] < +\infty.$$

Shifting values if necessary, we could assume that this upper bound is equal to zero to obtain equation (18). The latter being the condition required in Lemma 3, we would have reached a conclusion.

Let us therefore establish that (19) is integrable over  $\mathbb{R}_+$ . Recall that  $G$  is constructed using the Bernstein interpolation given in Lemma 1 between successive values of  $\lambda$ . As a

result, for a fixed  $\theta$ , the function  $\langle n(\theta), G(\lambda, \theta) \rangle$  is the interpolation of the piecewise affine function interpolating

$$\begin{aligned} & (\lambda_{3i}, \sigma_{S_{3i}}(n(\theta))) \\ & (\lambda_{3i+1}, \sigma_{S_{3i+1}}(n(\theta))), \\ & (\lambda_{3i+2}, \sigma_{S_{3i+2}}(n(\theta))), \\ & (\lambda_{3i+3}, \sigma_{S_{3i+3}}(n(\theta))), \end{aligned} \tag{22}$$

as in equation (10). This interpolation is concave and increasing.

Assumption (14) ensures that  $K_j = 1 + O(1/j^3)$ . Then

$$\prod_{j=1}^m K_j = \bar{K} + O(1/j^2)$$

where  $\bar{K}$  is the finite, positive limit of the product (we can for example perform integral series comparison after taking the logarithm).

The recursion on the values writes for all  $i \geq 1$

$$\lambda_{i+1} = \lambda_i + K_i(\lambda_i - \lambda_{i-1}),$$

so that

$$\lambda_{i+1} - \lambda_i = (\lambda_1 - \lambda_0) \prod_{j=1}^i K_j = (\lambda_1 - \lambda_0) \bar{K} + O(1/i^2).$$

This means that the gap between consecutive values tends to be constant. Thus by (4) in Lemma 1, see also Remark 1, the degree of the Bernstein interpolants is bounded. Using this bound together with inequality (3), providing bounds for the derivatives of Bernstein's polynomial, ensure that, for all  $\lambda$  in  $[\lambda_{3i}, \lambda_{3i+3}]$ :

$$\begin{aligned} & \left| \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle \right| \\ & = O \left( \max_{j=3i+2, 3i+1} \left| \frac{\sigma_{S_{j+1}}(n(\theta)) - \sigma_{S_j}(n(\theta))}{\lambda_{j+1} - \lambda_j} - \frac{\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta))}{\lambda_j - \lambda_{j-1}} \right| \right). \end{aligned}$$

Now for any  $j = 3i + 2, 3i + 1$ ,

$$\begin{aligned} & \left| \frac{\sigma_{S_{j+1}}(n(\theta)) - \sigma_{S_j}(n(\theta))}{\lambda_{j+1} - \lambda_j} - \frac{\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta))}{\lambda_j - \lambda_{j-1}} \right| \\ & = \frac{1}{\lambda_{j+1} - \lambda_j} \left| (\sigma_{S_{j+1}}(n(\theta)) - \sigma_{S_j}(n(\theta))) - \frac{\lambda_{j+1} - \lambda_j}{\lambda_j - \lambda_{j-1}} (\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta))) \right| \\ & = (1/((\lambda_1 - \lambda_0)\bar{K}) + O(1/i^2)) \\ & \quad \times |(\sigma_{S_{j+1}}(n(\theta)) - \sigma_{S_j}(n(\theta))) - (1 + O(1/i^2))(\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta)))| \\ & = (1/((\lambda_1 - \lambda_0)\bar{K}) + O(1/i^2)) \\ & \quad \times |(\sigma_{S_{j+1}}(n(\theta)) - \sigma_{S_j}(n(\theta))) - (\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta)))| \end{aligned}$$

where the last identity follows from the triangle inequality because using  $\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta)) = O(1)$  in (15). Hence

$$\begin{aligned}
& \left| \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle \right| \\
&= (1/((\lambda_1 - \lambda_0)\bar{K}) + O(1/i^2)) \\
&\quad \times O\left(\max_{j=3i+2, 3i+1} |\sigma_{S_{j+1}}(n(\theta)) - \sigma_{S_j}(n(\theta)) - (\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta)))|\right) \\
&= (1/((\lambda_1 - \lambda_0)\bar{K}) + O(1/i^2)) \\
&\quad \times O\left(\max_{j=3i+2, 3i+1} |\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta))| \times \left| \frac{\sigma_{S_{j+1}}(n(\theta)) - \sigma_{S_j}(n(\theta))}{\sigma_{S_j}(n(\theta)) - \sigma_{S_{j-1}}(n(\theta))} - 1 \right| \right) \\
&= O(1/i^3),
\end{aligned}$$

where the last inequality follows from (15) and (14). Now as  $i \rightarrow \infty$ ,  $\lambda_{3i} \sim \lambda_{3i+3} \sim ic$  for some constant  $c > 0$  and

$$\sup_{\lambda \in [\lambda_{3i}, \lambda_{3i+3}], \theta \in [0, 2\pi]} -\lambda \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle = O(1/i^2)$$

and

$$\sup_{\theta \in [0, 2\pi]} -\lambda \left\langle \frac{\partial^2 G}{\partial \lambda^2}(\lambda, \theta), n(\theta) \right\rangle$$

has finite integral as  $\lambda \rightarrow \infty$ . This implies (20) and it concludes the proofs.  $\square$

## 4 Smooth convex interpolation for sequences of polygons

Given a sequence of points  $A_1, \dots, A_n$ , we denote by  $A_1 \dots A_n$  the polygon obtained by joining successive points ending the loop with the segment  $[A_n, A_1]$ . In the sequel we consider mainly convex polygons, so that the vertices  $A_1, \dots, A_n$  are also the extreme points.

The purpose of this section is first to show that polygons can be approximated by smooth convex sets with prescribed normals under weak assumptions. Figure 3 illustrates the result we would like to establish: given a target polygon with prescribed normals at its vertices, we wish to construct a smooth convex set interpolating the vertices with the desired normals and whose distance to the polygon is small.

Then given a sequence of nested polygons, we provide a smooth convex function which interpolates the polygons in the sense described just above.

Given a closed nonempty convex subset  $S$  of  $\mathbb{R}^p$  and  $x$  in  $S$ , we recall that the *normal cone to  $S$  at  $x$*  is

$$N_S(x) = \{z \in \mathbb{R}^p : \langle z, y - x \rangle \leq 0, \forall y \in S\}.$$

Such vectors will often simply called normals (to  $S$ ) at  $x$ .



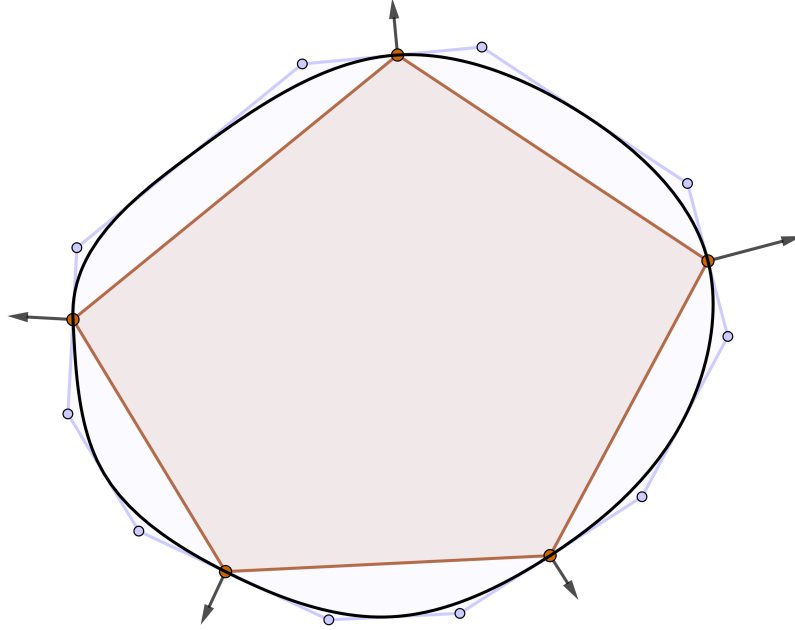


Figure 3: Arrows designate the prescribed normals. We construct a strictly convex set with smooth boundary entirely contained in the auxiliary (blue) polygon. This set interpolates the normals and the distance to the original (red) polygon can be chosen arbitrarily small. The degree of smoothness of the boundary can be chosen arbitrarily high.

## 4.1 Smooth approximations of polygons

**Lemma 4.** *For any  $r_-, r_+ > 0$ ,  $t_- > 0$ ,  $t_+ < 0$  and  $\epsilon > 0$ ,  $m \in \mathbb{N}$ ,  $m \geq 3$ , there exists a strictly concave polynomial function  $p: [0, 1] \mapsto [0, \epsilon]$  such that*

$$\begin{aligned}
 p(0) &= 0 & p(1) &= 0 \\
 p'(0) &= t_- & p'(1) &= t_+ \\
 p''(0) &= -r_- & p''(1) &= -r_+. \\
 p^{(q)}(0) &= 0 \quad q \in \{3, \dots, m\}.
 \end{aligned}$$

**Proof :** Let us begin with preliminary remarks. Consider for any  $a, b$  in  $\mathbb{R}$ , the function

$$f: t \mapsto a(t - b)^2.$$

For any  $t$  in  $\mathbb{R}$ ,  $q$  in  $\mathbb{N}$ ,  $q > 2$ , and any  $c > 0$ , we have

$$\begin{aligned}
 f(t + c) - f(t) &= \Delta_c^1 f(t) = ac(2(t - b) + c) \\
 f(t + 2c) - 2f(t + c) + f(t) &= \Delta_c^2 f(t) = ac(2(t + c - b) + c - 2(t - b) - c) = 2ac^2 \\
 \Delta_c^q f(t) &= 0
 \end{aligned} \tag{23}$$

*Choosing the degree  $d$  and constructing the polynomial.* For any  $d$  in  $\mathbb{N}$ ,  $d \geq 2m + 1$ , we set

$$\begin{aligned} a_-(d) &= \frac{-dr_-}{2(d-1)} < 0 & b_-(d) &= \frac{1}{2d} \left( 1 + 2t_- \frac{d-1}{r_-} \right) > 0 \\ a_+(d) &= \frac{-dr_+}{2(d-1)} < 0 & b_+(d) &= 1 + \frac{1}{2d} \left( -1 + 2t_+ \frac{d-1}{r_+} \right) < 1, \end{aligned} \quad (24)$$

and define the functions

$$\begin{aligned} f_d: s &\mapsto \begin{cases} a_-(d)((s - b_-(d))^2 - b_-(d)^2) & \text{if } s \leq b_-(d) \\ -a_-(d)b_-(d)^2 & \text{if } s \geq b_-(d) \end{cases} \\ g_d: t &\mapsto \begin{cases} a_+(d)((s - b_+(d))^2 - (1 - b_+(d))^2) & \text{if } s \geq b_+(d) \\ -a_+(d)(1 - b_+(d))^2 & \text{if } s \leq b_+(d). \end{cases} \end{aligned}$$

Furthermore, we set

$$\begin{aligned} f: t &\mapsto \begin{cases} \frac{r_-}{2} \left( \left( \frac{t_-}{r_-} \right)^2 - \left( s - \frac{t_-}{r_-} \right)^2 \right) & \text{if } s \leq \frac{t_-}{r_-} \\ \frac{r_-}{2} \left( \frac{t_-}{r_-} \right)^2 & \text{if } s \geq \frac{t_-}{r_-} \end{cases} \\ g: t &\mapsto \begin{cases} \frac{r_+}{2} \left( \left( \frac{t_+}{r_+} \right)^2 - \left( s - 1 - \frac{t_+}{r_+} \right)^2 \right) & \text{if } s \geq 1 + \frac{t_+}{r_+} \\ \frac{r_+}{2} \left( \frac{t_+}{r_+} \right)^2 & \text{if } s \leq 1 + \frac{t_+}{r_+}. \end{cases} \end{aligned}$$

Note that  $b_-(d) \rightarrow t_-/r_-$ ,  $b_+(d) \rightarrow 1 + t_+/r_+$ ,  $a_-(d) \rightarrow -r_-/2$  and  $a_+(d) \rightarrow -r_+/2$  as  $d \rightarrow \infty$  so that  $f_d \rightarrow f$  and  $g_d \rightarrow g$  uniformly on  $[0, 1]$ . For any  $d$ ,  $f_d$  is concave increasing and  $g_d$  is concave decreasing and all of them are Lipschitz continuous on  $[0, 1]$  with constants that do not depend on  $d$ . Note also that  $f(0) = 0 < g(0)$  and  $g(1) = 0 < f(1)$ . We choose  $d \geq 2m + 1$  such that

$$\begin{aligned} f_d \left( \frac{m}{d} \right) &\leq \min \left( \epsilon, g_d \left( \frac{m}{d} \right) \right) \\ g_d \left( 1 - \frac{m}{d} \right) &\leq \min \left( \epsilon, f_d \left( 1 - \frac{m}{d} \right) \right). \end{aligned} \quad (25)$$

Such a  $d$  always exists because in both cases, the left hand side converges to 0 and the right hand side converges to a strictly positive term as  $d$  tends to  $\infty$ . For such a  $d$ , we set  $h: s \mapsto \min \{ f_d(s), g_d(s), \epsilon \}$ . By construction,  $h$  is concave, agrees with  $f_d$  on  $[0, \frac{m}{d}] \subset [0, 1/2]$  and with  $g_d$  on  $[1 - \frac{m}{d}, 1] \subset [1/2, 1]$ . Using equation (23) with  $c = 1/d$ ,

we deduce that

$$\begin{aligned}
d(d-1)\Delta^2 h(0) &= d(d-1)\Delta^2 f_d(0) = d(d-1)\frac{2a_-(d)}{d^2} = -r_- \\
d\Delta h(0) &= d\Delta f_d(0) = a_-(d)\left(\frac{1}{d} - 2b_-(d)\right) = t_- \\
\Delta^q h(0) &= 0 = \Delta^q f_d(0) \quad \forall m \geq q \geq 3 \\
d(d-1)\Delta^2 h\left(1 - \frac{2}{d}\right) &= d(d-1)\Delta^2 g_d\left(1 - \frac{2}{d}\right) = d(d-1)\frac{2a_+(d)}{d^2} = -r_+ \\
d\Delta h\left(1 - \frac{1}{d}\right) &= d\Delta g_d\left(1 - \frac{1}{d}\right) = a_+(d)\left(\frac{-1}{d} + 2(1 - b_+(d))\right) = t_+ \\
\Delta^q h\left(1 - \frac{m}{d}\right) &= \Delta^q g_d\left(1 - \frac{m}{d}\right) = 0 \quad \forall m \geq q \geq 3.
\end{aligned}$$

From the concavity of  $h$  and the derivative formula (2), we deduce that the polynomial  $B_{h,d}$  satisfies the desired properties.  $\square$

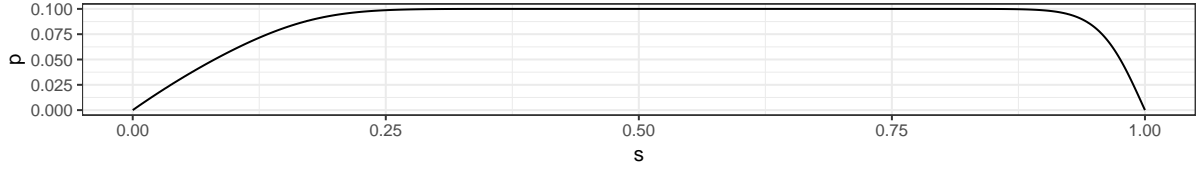


Figure 4: Illustration of the approximation result of Lemma 4 with  $\epsilon = 0.1$ ,  $m = 3$ ,  $t_- = 0.7$ ,  $t_+ = -2.2$ ,  $r_- = 2$  and  $r_+ = 0.2$ . The resulting polynomial is of degree 66. Numerical estimations of the first and second order derivatives at 0 and 1 match the required values up to 3 precision digits. The polynomial is strongly concave, however, this is barely visible because the strong concavity constant is extremely small.

We deduce the following result

**Lemma 5.** *Let  $a > 0$ ,  $r > 0$ ,  $\epsilon > 0$ , and an integer  $m \geq 3$ . Consider two unit vectors:  $v_-$  with strictly positive entries and  $v_+$  with first entry strictly positive and second entry strictly negative. Then there exists a  $C^m$  curve  $\gamma: [0, M] \mapsto \mathbb{R}^2$ , such that*

1.  $\|\gamma'\| = 1$ .
2.  $\gamma(0) = (-a, 0) := A$  and  $\gamma(1) = (0, a) := B$ .
3.  $\gamma'(0) = v_-$  and  $\gamma'(1) = v_+$ .
4.  $\|\gamma''(0)\| = \|\gamma''(-1)\| = r$ .
5.  $\det(\gamma', \gamma'') < 0$  along the curve.
6.  $\gamma^{(q)}(0) = \gamma^{(q)}(1) = 0$  for any  $3 \leq q \leq m$ .
7.  $\text{dist}(\gamma([0, M]), [A, B]) \leq \epsilon$ .

**Proof :** Consider the graph of a polynomial as given in Lemma 4 with  $t_- = v_-[2]/v_-[1] > 0$ ,  $t_+ = v_+[2]/v_+[1] < 0$  and  $r_- = \frac{r}{2a}(1 + t_-^2)^{\frac{3}{2}}$ ,  $r_+ = \frac{r}{2a}(1 + t_+^2)^{\frac{3}{2}}$  and  $\epsilon/2a$  as an approximation parameter. This graph is parametrized by  $t$ . It is possible to reparametrize it by arclength to obtain a  $C^m$  curve  $\gamma_0$  whose tangents at 0 is  $T_-$ , at 1 is  $T_+$ , and whose curvature at 0 and 1 is  $-\frac{r}{2a}$ . Furthermore,  $\gamma_0$  has strictly negative curvature whence item 5. Consider the affine transform:  $x \mapsto 2a(x - 1/2)$ ,  $y \mapsto 2ay$ . This results in a  $C^m$  curve  $\gamma$ , parametrized by arclength which satisfies the desired assumptions.  $\square$

**Lemma 6** (Normal approximations of polygons by smooth convex sets). *Let  $S = A_1 \dots A_n$  be a convex polygon. For each  $i$ , let  $V_i$  be in  $N_S(A_i)$  such that the angle between  $V_i$  and each of the two neighboring faces is within  $(\frac{\pi}{2}, \pi)$ . Then for any  $\epsilon > 0$  and any  $m \geq 2$ , there exists a compact convex set  $C \subset \mathbb{R}^2$  such that*

- (i) *the boundary of  $C$  is  $C^m$  with non vanishing curvature,*
- (ii)  *$S \subset C$ ,*
- (iii) *for any  $i = 1, \dots, n$ ,  $A_i$  in  $\text{bd}(C)$  and the normal cone to  $C$  at  $A_i$  is given by  $V_i$ ,*
- (iv)  *$\max_{y \in C} \text{dist}(y, S) \leq \epsilon$ .*

**Proof :** We assume without loss of generality that  $A_1, \dots, A_n$  are ordered clockwise. For  $i = 1, \dots, n-1$ , and each segment  $[A_i, A_{i+1}]$ , we may perform a rotation and a translation to obtain  $A_i = -(a, 0)$  and  $A_{i+1} = (a, 0)$ . Working in this coordinate system, using the angle condition on  $V_i$ , we may choose  $v_i^-, v_{i+1}^+$  satisfying the hypotheses of Lemma 5 respectively orthogonal to  $V_i$  and  $V_{i+1}$ . Choosing  $r = 1$ , we obtain  $\gamma_i: [0, M_i] \mapsto \mathbb{R}^2$  as given by Lemma 5. Rotation and translations affect only the direction of the derivatives of curves, not their length. Hence, it is possible to concatenate curves  $(\gamma_i)_{i=1}^{n-1}$  and to preserve the  $C^m$  properties of the resulting curve. At end-points, tangents and second order derivatives coincide while higher derivatives vanish. Furthermore the curvature has constant sign and does not vanish. We obtain a closed  $C^m$  curve which defines a convex set which satisfies all the requirements of the lemma.  $\square$

**Remark 5** (Bisector). Given any polygon, choosing normal vectors as given by the direction of the bisector of each angles ensure that the above assumptions are satisfied. Hence all our approximation results hold given polygon without specifying the choice of outer normals.

## 4.2 Smooth convex interpolants of polygonal sequences

**Definition 1** (Interpolability). For  $n \geq 3$ , let  $A_1 \dots A_n$  be a convex polygon  $S$  and  $V_i$  be in  $N_S(V_i)$  for  $i = 1, \dots, n$ . We say that  $(A_i, V_i)_{i=1}^n$  is *interpolable* if for each  $i = 1, \dots, n$ , the angle between  $V_i$  and each of the two neighboring faces of the polygon is in  $(\frac{\pi}{2}, \pi)$ . The collection  $(A_i, V_i)_{i=1}^n$  is called a *polygon-normal pair*.

Let  $I = \mathbb{Z}$  or  $I = \mathbb{N}$ . Let  $(PN_i)_{i \in I}$  be a sequence of interpolable polygon-normal pairs. Setting for  $i$  in  $I$ ,  $PN_i = \left\{ (A_{j,i})_{j=1}^{n_i}, (V_{j,i})_{j=1}^{n_i} \right\}$  where  $n_j$  is in  $\mathbb{N}$  and denoting by  $T_i$  the polygon  $A_{1,i} \dots A_{n_i,i}$ , we say that the sequence  $(PN_i)_{i \in I}$  is strictly increasing if for all  $i$  in  $I$ ,  $T_i \subset \text{int}(T_{i+1})$ .

Let  $(PN_i)_{i \in I}$  be a strictly increasing sequence of interpolable polygon-normal pairs. A sequence  $(\epsilon_i)_{i \in I}$  in  $(0, 1)$  is said to be *admissible* if  $0 \in \text{int}(T_i)$  for each  $i$  in  $I$  and

$$\gamma T_i \subset \text{int} T_{i+1}$$

for all  $\gamma \in [1 - \epsilon_i, 1 + \epsilon_i]$ . We have the following corollary of Theorem 2.

**Corollary 1** (Smooth convex interpolation of polygon sequences). *Let  $I = \mathbb{Z}$  or  $I = \mathbb{N}$ . Let  $(PN_i)_{i \in I}$  be a strictly increasing sequence of interpolable polygon-normal pairs and  $(\epsilon_i)_{i \in I}$  be admissible. Set  $\mathcal{T} := \text{int}(\cup_{i \in I} T_i)$ .*

*Then for any  $k$  in  $\mathbb{N}$ ,  $k \geq 2$  there exists a  $C^k$  convex function  $f: \mathcal{T} \mapsto \mathbb{R}$ , and an increasing sequence  $(\lambda_i)_{i \in I}$ , with  $\inf_{i \in I} \lambda_i > -\infty$ , such that for each  $i$  in  $I$*

- (i)  $T_i \subset \{x, f(x) \leq \lambda_i\}$ .
- (ii)  $\text{dist}(T_i, \{x, f(x) \leq \lambda_i\}) \leq \epsilon_i$ .
- (iii) For each  $i$  in  $I$ ,  $j$  in  $\{1, \dots, n_i\}$ , we have  $f(A_{i,j}) = \lambda_i$  and  $\nabla f(x)$  is colinear to  $V_{i,j}$ .
- (iv)  $\nabla^2 f$  is positive definite outside  $\text{argmin} f$ . When there is a unique minimizer then  $\nabla^2 f$  is positive definite throughout  $\mathcal{T}$  (this is the case when  $I = \mathbb{N}$  or when  $I = \mathbb{Z}$  and  $\cap_{i \in I} T_i$  is a singleton).

We add two remarks which will be useful for directional convergence issues and the construction of Legendre functions:

- (a) If two consecutive elements of the sequence of interpolable polygon-normal pairs are homothetic with center 0 in the interior of both polytopes, then the restriction of the resulting convex function to this convex ring can be constructed such that all the sublevel sets within this ring are homothetic with the same center.
- (b) If further conditions are imposed on the elements of a strictly increasing interpolable polygon-normal pair, then the resulting function can be constructed to be Legendre and globally Lipschitz continuous (that is, its Legendre conjugate has bounded support). This is a consequence of Proposition 2 and will be detailed in the next section.

### 4.3 More on Legendre functions and a pathological function with polyhedral domain

Using intensively polygonal interpolation, we build below a finite continuous Legendre function  $h$  on an  $\ell^\infty$  square with oscillating “mirror lines”:  $t \rightarrow \nabla h^*(\nabla h(x_0) + tc)$ .

We start with the following preparation proposition related to the Legendre interpolation of Proposition 2.

**Lemma 7.** *Let  $(PN_i)_{i \in \mathbb{N}^*}$  be a strictly increasing sequence of interpolable polygon-normal pairs. Setting for  $i$  in  $\mathbb{N}^*$ ,  $PN_i = \left\{ (A_{j,i})_{j=1}^{n_i}, (V_{j,i})_{j=1}^{n_i} \right\}$  where  $n_j$  is in  $\mathbb{N}^*$  and denoting by  $T_i$  the polygon  $A_{1,i} \dots A_{n_i,i}$ , we assume that*

$$T_i = 3iP, \forall i \in \mathbb{N}^*,$$

where  $P$  is a fixed polygon which contains the unit Euclidean disk.

Then for any  $l$  in  $\mathbb{N}$ ,  $l \geq 2$ , there exists a strictly increasing sequence of sets  $(S_i)_{i \in \mathbb{N}, i \geq 2}$ , such that for  $j \geq 1$ ,

- $S_{3j}$  interpolates the normals of  $PN_j$  in the sense of Lemma 6 with  $\text{dist}(S_{3j}, T_j) \leq 1/(4(3j+2)^l)$
- $S_{3j-1} = \frac{3j-1}{3_j} S_{3j}$
- $S_{3j+1} = \frac{3j+1}{3_j} S_{3j}$

This sequence has the following properties

- there exists  $c > 0$ , such that for all  $j$  in  $\mathbb{N}$ ,  $j \geq 3$  and for all unit vector  $x$ ,

$$c \geq \sigma_{S_{j+1}}(x) - \sigma_{S_j}(x) \geq 1 - \frac{1}{(j+1)^l}. \quad (26)$$

- for all unit vector  $x$ ,

$$\left| \frac{\sigma_{S_{j+1}}(x) - \sigma_{S_j}(x)}{\sigma_{S_j}(x) - \sigma_{S_{j-1}}(x)} - 1 \right| \leq \frac{1}{j^l}, \forall j \geq 3. \quad (27)$$

**Proof :** Set for all  $j$  in  $\mathbb{N}^*$ ,  $\delta_j = \frac{1}{4(3j+2)^l}$  and let  $S_{3j}$  be the  $\delta_j$  interpolant of  $T_j = 3jP$  as given by Lemma 6 so that  $\text{dist}(S_{3j}, 3jP) \leq \delta_j$ . Since  $P$  contains the unit ball,

$$3jP \subset S_{3j} \subset (3j + \delta_j)P. \quad (28)$$

Now set

$$S_{3j-1} = \frac{3j-1}{3_j} S_{3j}$$

$$S_{3j+1} = \frac{3j+1}{3_j} S_{3j}.$$

For any  $j$  in  $\mathbb{N}^*$ , it is clear that  $S_{3j-1} \subset \text{int}(S_{3j})$  and  $S_{3j} \subset \text{int}(S_{3j+1})$ . Furthermore, by (28), we have

$$S_{3j+1} \subset \frac{3j+1}{3_j} (3j + \delta_j)P \subset ((3j+1) + 2\delta_j)P \subset \text{int}((3j+2)P) \subset \text{int}(S_{3j+2})$$

so that we indeed have a strictly increasing sequence of sets. We obtain from the construction, for any  $j$  in  $\mathbb{N}^*$ , and any unit vector  $x$ ,

$$\begin{aligned}
\sigma_{S_{3j}}(x) - \sigma_{S_{3j-1}}(x) &= \sigma_{S_{3j+1}}(x) - \sigma_{S_{3j}}(x) \\
&= \frac{1}{3j} \sigma_{S_{3j}}(x) \in \left[ \sigma_P(x), \left(1 + \frac{\delta_j}{3j}\right) \sigma_P(x) \right] \subset \left[ 1, \left(1 + \frac{\delta_j}{3j}\right) \sigma_P(x) \right] \\
\sigma_{S_{3j+2}}(x) - \sigma_{S_{3j+1}}(x) &\leq \sigma_P(x)(3j+2) \left(1 + \frac{\delta_{j+1}}{3j+3}\right) - \sigma_P(x)(3j+1) \\
\sigma_{S_{3j+2}}(x) - \sigma_{S_{3j+1}}(x) &\geq \sigma_P(x)(3j+2) - \sigma_P(x)(3j+1) \left(1 + \frac{\delta_j}{3j}\right) \\
&= \sigma_P(x) \left(1 - \delta_j \frac{3j+1}{3j}\right) \\
&\geq 1 - \delta_j \frac{4}{3} \geq 1 - \frac{1}{(3j+2)^l}
\end{aligned} \tag{29}$$

This proves (26). We deduce that for all  $j$  in  $\mathbb{N}^*$ ,

$$\begin{aligned}
\frac{\sigma_{S_{3j+1}}(x) - \sigma_{S_{3j}}(x)}{\sigma_{S_{3j}}(x) - \sigma_{S_{3j-1}}(x)} &= 1, \text{ for all nonzero vector } x, \\
\max_{\|x\|=1} \frac{\sigma_{S_{3j+2}}(x) - \sigma_{S_{3j+1}}(x)}{\sigma_{S_{3j+1}}(x) - \sigma_{S_{3j}}(x)} &\leq (3j+2) \left(1 + \frac{\delta_{j+1}}{3j+3}\right) - (3j+1) \\
&= 1 + \frac{3j+2}{3j+3} \delta_{j+1} \leq 1 + \frac{1}{(3j+1)^l}, \\
\min_{\|x\|=1} \frac{\sigma_{S_{3j+2}}(x) - \sigma_{S_{3j+1}}(x)}{\sigma_{S_{3j+1}}(x) - \sigma_{S_{3j}}(x)} &\stackrel{(29)}{\geq} \frac{1 - \delta_j \frac{3j+1}{3j}}{\left(1 + \frac{\delta_j}{3j}\right)} \\
&\geq \left(1 - \delta_j \frac{3j+1}{3j}\right) \left(1 - \frac{\delta_j}{3j}\right) \\
&\geq 1 - \delta_j \left(\frac{3j+1}{3j} + \frac{1}{3j}\right) \geq 1 - \delta_j \frac{5}{3} \geq 1 - \frac{1}{(3j+1)^l}.
\end{aligned}$$

Furthermore, using the fact that  $t \mapsto \frac{1+t}{1-t}$  is increasing on  $(-\infty, 1)$  and the fact that  $\delta_{j+1} \leq \delta_j$ ,

$$\begin{aligned}
\max_{\|x\|=1} \frac{\sigma_{S_{3j+3}}(x) - \sigma_{S_{3j+2}}(x)}{\sigma_{S_{3j+2}}(x) - \sigma_{S_{3j+1}}(x)} &\stackrel{(29)}{\leq} \frac{1 + \frac{\delta_{j+1}}{3j+3}}{1 - (3j+1) \frac{\delta_j}{3j}} \\
&\leq \frac{1 + (3j+1) \frac{\delta_j}{3j}}{1 - (3j+1) \frac{\delta_j}{3j}} \\
&\leq \frac{1 + \delta_j \frac{4}{3}}{1 - \delta_j \frac{4}{3}}.
\end{aligned} \tag{30}$$

Setting  $s(t) = (1+t)/(1-t)$ , we have, for all  $t \leq 1/2$

$$\begin{aligned} s'(t) &= \frac{2}{(1-t)^2}, \quad s(0) = 1 \\ s''(t) &= \frac{4}{(1-t)^3} \leq 24, \quad s'(0) = 2. \end{aligned}$$

Thus  $s(t) \leq 1 + 2t + 12t^2$  on  $(-\infty, 1/2]$ . Since  $\frac{4}{3}\delta_j \leq \frac{4}{75} \leq \frac{1}{2}$ , we deduce from the previous remark and (30) above:

$$\begin{aligned} \max_{\|x\|=1} \frac{\sigma_{S_{3j+3}}(x) - \sigma_{S_{3j+2}}(x)}{\sigma_{S_{3j+2}}(x) - \sigma_{S_{3j+1}}(x)} &\leq 1 + \frac{8}{3}\delta_j + \frac{64}{3}\delta_j^2 = 1 + \delta_j \left( \frac{8}{3} + \frac{64}{3}\delta_j \right) \\ &\leq 1 + \delta_j \left( 3 + \frac{64}{3 \times 25} \right) \leq 1 + 4\delta_j = 1 + \frac{1}{(3j+2)^l}. \end{aligned}$$

Finally using (29) again,

$$\begin{aligned} \min_{\|x\|=1} \frac{\sigma_{S_{3j+3}}(x) - \sigma_{S_{3j+2}}(x)}{\sigma_{S_{3j+2}}(x) - \sigma_{S_{3j+1}}(x)} &\geq \frac{1}{(3j+2) \left( 1 + \frac{\delta_{j+1}}{3j+3} \right) - (3j+1)} \\ &= \frac{1}{1 + \delta_{j+1} \frac{3j+2}{3j+3}} \\ &\geq 1 - \delta_{j+1} \frac{3j+2}{3j+3} \geq 1 - \delta_{j+1} \geq 1 - \frac{1}{(3j+2)^l}. \end{aligned}$$

This proves the desired result. □

Combining Lemma 7 and Proposition 2, we obtain the following result.

**Theorem 3.** *Let  $(PN_i)_{i \in \mathbb{N}^*}$  be a strictly increasing sequence of interpolable polygon-normal pairs. Set for  $i$  in  $\mathbb{N}^*$ ,  $PN_i = \left\{ (A_{j,i})_{j=1}^{n_i}, (V_{j,i})_{j=1}^{n_i} \right\}$  where  $n_i$  is in  $\mathbb{N}^*$ , denote by  $T_i$  the polygon  $A_{1,i} \dots A_{n_i,i}$ , and assume that for all  $i$  in  $\mathbb{N}^*$ ,  $T_i = 3iP$  where  $P$  is a fixed polygon which contains the unit disk in its interior.*

*Then for any  $k$  in  $\mathbb{N}$ ,  $k \geq 2$  and all  $l \geq 3$ , there exists a  $C^k$  globally Lipschitz continuous Legendre function,  $h: \mathbb{R}^2 \mapsto \mathbb{R}$ , and an increasing sequence  $(\lambda_i)_{i \in \mathbb{N}}$ , such that for each  $i$  in  $\mathbb{N}$ :*

- $T_i \subset \{x, h(x) \leq \lambda_i\}$ ,
- $\text{dist}(T_i, \{x, h(x) \leq \lambda_i\}) \leq \frac{1}{4(3i+2)^l}$ ,
- For any  $x$  with  $h(x) = \lambda_i$  and  $\nabla h(x)$  is colinear to  $V_i$  for each vertex  $x$  of  $T_i$ ,
- $h$  has positive definite Hessian and is globally Lipschitz continuous,
- $h^*$  has compact domain and is  $C^k$  on the interior of its domain.

**Corollary 2** (Continuity on the domain). *The function  $h^*$  constructed in Theorem 3 has compact polygonal domain and is continuous on this domain.*



**Proof :** Since  $P$  is a polygon and contains the unit Euclidean disk, the gauge function of  $3P$  is polyhedral with full domain  $\mathbb{R}^2$ , call it  $\omega$ . Denote by  $P^\circ$  the polar of  $P$ . This is a polytope and since  $\omega$  is the gauge of  $P$ , we actually have  $\omega = \sigma_{P^\circ}$ , the support function of the polar of  $P$  [35, Theorem 1.7.6]. Hence the the convex conjugate of  $\omega$  is the indicator of the polytope  $P^\circ$  [37, Theorem 13.2].

It can be easily seen from the proof of Proposition 2 that  $\lambda_i = \alpha i + r_i$  with  $r(i) = O(1)$  as  $i \rightarrow \infty$ . Without loss of generality, we may suppose that  $\alpha = 1$  (this is a simple rescaling) so that there is a positive constant  $c$  such that  $|\lambda_i - i| \leq c$  for all  $i$ .

Let  $h$  be given as in Theorem 3, fix  $i \geq 1$  and  $x \in \mathbb{R}^2$  such that  $\lambda_{i-1} \leq h(x) \leq \lambda_i$ . We have in  $\mathbb{R}^2$

$$\{y : \omega(y) \leq i - 1\} \subset \{y : h(y) \leq \lambda_{i-1}\} \subset \{y : h(y) \leq \lambda_i\} \subset \{y : \omega(y) \leq i + 1\}$$

and hence

$$i - 1 \leq \omega(x) \leq i + 1$$

and we deduce that

$$\omega(x) - 2 - c \leq i - 1 - c \leq \lambda_{i-1} \leq h(x) \leq \lambda_i \leq i + c \leq \omega(x) + c + 1.$$

Since  $i$  was arbitrary, this shows that there exists a constant  $C > 0$  such that  $|h(x) - \omega(x)| \leq C$  for all  $x \in \mathbb{R}^2$ . Recall that  $z \mapsto \sup_{y \in \mathbb{R}^2} \langle y, z \rangle - \omega(y)$  is the indicator function of  $P^\circ$ , hence,

$$\begin{aligned} z \in P^\circ &\Rightarrow \sup_{y \in \mathbb{R}^2} \langle y, z \rangle - \omega(y) = 0 \Rightarrow \sup_{y \in \mathbb{R}^2} \langle y, z \rangle - h(y) \leq C < +\infty \\ z \notin P^\circ &\Rightarrow \sup_{y \in \mathbb{R}^2} \langle y, z \rangle - \omega(y) = +\infty \Rightarrow \sup_{y \in \mathbb{R}^2} \langle y, z \rangle - h(y) = +\infty \end{aligned}$$

which shows that the domain of  $h^*$  is actually  $P^\circ$  which is a polytope. Now,  $h^*$  is convex and lower semicontinuous on  $P^\circ$ , invoking the results of [23], it is also upper semicontinuous on  $B^*$  and finally it is continuous on  $B^*$ .  $\square$

**Corollary 3** (A pathological Legendre function). *For any  $\theta \in \left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$  there exists a Legendre function  $h: \mathbb{R}^2 \mapsto \mathbb{R}$  whose domain is a closed square, continuous on this domain and  $C^k$  on its interior, such that for all  $i \in \mathbb{N}^*$ ,  $\nabla h^*(i, 0)$  is proportional to  $(\cos(\theta), (-1)^i \sin(\theta))$ .*

**Proof :** For  $x = (u, v)$ , set  $\|x\|_1 = |u| + |v|$ , and let  $P = \{x \in \mathbb{R}^2, \|x\|_1 \leq 2\}$ . Let us construct a strictly increasing sequence of interpolable polygon-normal pairs  $(PN_i)_{i \in \mathbb{N}^*}$  as follows, we fix  $\theta \in \left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$  and set for all  $i \in \mathbb{N}^*$  :

- $T_i = 3iP$ , the polygon associated to the  $i$ -th term  $PN_i$  of the sequence,
- except at the rightmost corner, consider the normals given by the canonical basis vectors and their opposite,

- at the rightmost corner,  $(6i, 0)$ , one chooses the normal given by the vector

$$(\cos(\theta), (-1)^i \sin(\theta)).$$

We now invoke Theorem 3 to obtain a Lipschitz continuous Legendre function, denoted  $h^*$ , with full domain having all the  $T_i$  as sublevel sets and satisfying the hypotheses of the corollary. Rescaling by a factor 6 and setting  $h = h^{**}$  gives the result.  $\square$

## 5 Counterexamples in continuous optimization

We are now in position to apply our interpolation results to build counterexamples to classical problems in convex optimization. We worked on situations ranging from structural questions to qualitative behavior of algorithms and ODEs. Through 9 counterexamples we tried to cover a large spectrum but there are many more possibilities that are left for future research. Some example are constructed from decreasing sequences of convex sets, they can be interpolated using Theorem 2 with  $I = \mathbb{Z}$ , indexing the sequence with negative indices and adding artificially additional sets for positive indices. Nonetheless we sometimes depart from the notations of the first sections and index these sequences by  $\mathbb{N}$  even though they are decreasing for simplification purposes.

### 5.1 Kurdyka-Łojasiewicz inequality may not hold

The following result is proved in [12], it was crucial to construct a  $C^2$  convex function which does not satisfy Kurdyka-Łojasiewicz (KL) inequality.

**Lemma 8.** [12, Lemma 35] *There exists a decreasing sequence of compact convex sets  $(T_i)_{i \in \mathbb{N}}$  such that for any  $i$  in  $\mathbb{N}$ ,  $T_- = T_{i+1}$  and  $T_+ = T_i$  satisfy Assumption 1 and*

$$\sum_{i=0}^{+\infty} \text{dist}(T_i, T_{i+1}) = +\infty$$

As a corollary, we improve the counterexample in [12] and provide a  $C^k$  convex counterexamples for any  $k \geq 2$  in  $\mathbb{N}$ .

**Corollary 4** (Smooth convex functions are not KL in general). *There exists a  $C^k$  convex function  $f: \mathbb{R}^2 \mapsto \mathbb{R}$  which does not satisfy KL inequality. More precisely, for any  $r > \inf f$  and  $\varphi: [\inf f, r] \mapsto \mathbb{R}$  continuous and differentiable on  $(\inf f, r)$  with  $\varphi' > 0$  and  $\varphi(\inf f) = 0$ , we have*

$$\inf\{\|\nabla(\varphi \circ f)(x)\| : x \in \mathbb{R}^2, \inf f < f(x) < r\} = 0.$$

**Proof :** Using [35, Theorem 1.8.13], each  $T_i$  can be approximated up to arbitrary precision by a polygon. Hence we may assume that all  $T_i$  are polygonal while preserving the property

of Lemma 8 as well as Assumption 1. Furthermore, using Lemma 6 and Remark 5 each  $T_i$  can in turn be approximated with arbitrary precision by a convex set with  $C^k$  boundary and positive curvature. Hence we may also assume that all  $T_i$  satisfy both the result of Lemma 8 and have  $C^k$  boundary with nonvanishing curvature. Reversing the order of the sets and adding additional sets artificially, we are in the conditions of application of Theorem 2 with  $I = \mathbb{Z}$  and the resulting  $f$  follows from the same argument as in [12, Theorem 36].  $\square$

## 5.2 Block coordinate descent may not converge

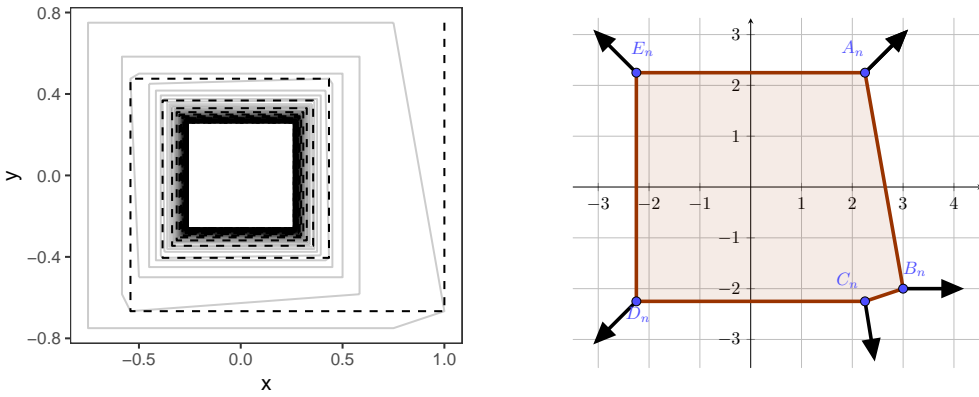


Figure 5: Illustration of the alternating minimization (resp. exact line search) example: on the left, the sublevel sets in gray and the corresponding alternating minimization (resp. exact line search) sequence in dashed lines. On the right the interpolating polygons together with their normal vectors as in Lemma 6.

The following polygonal construction is illustrated in Figure 5. For any  $n \geq 2$  in  $\mathbb{N}$ , we set

$$\begin{aligned} A_n &= \left( \frac{1}{4} + \frac{1}{n}, \frac{1}{4} + \frac{1}{n} \right) \\ B_n &= \left( \frac{1}{4} + \frac{1}{2(n-1)} + \frac{1}{2n}, \frac{1}{4} + \frac{1}{2n} + \frac{1}{2(n+1)} \right) \\ C_n &= \left( \frac{1}{4} + \frac{1}{n}, -\frac{1}{4} - \frac{1}{n} \right) \\ D_n &= \left( -\frac{1}{4} - \frac{1}{n}, -\frac{1}{4} - \frac{1}{n} \right) \\ E_n &= \left( -\frac{1}{4} - \frac{1}{n}, +\frac{1}{4} + \frac{1}{n} \right). \end{aligned}$$

This defines a convex polygon. We may choose the normals at  $A_n, C_n, D_n, E_n$  to be bisectors of the corresponding corners and the normal at  $B_n$  to be horizontal (see Figure 5). Rotating by an angle of  $-\frac{n\pi}{2}$  and repeating the process indefinitely, we obtain the

sequence of polygons depicted in Figure 5. It can be checked that the polygons form a strictly decreasing sequence of sets, as for  $n > 1$ , the polygon  $A_n B_n C_n D_n E_n$  is contained in the interior of the square  $A_{n-1} C_{n-1} D_{n-1} E_{n-1}$ . This fulfills the requirement of Corollary 1.

**Corollary 5.** *There exists a  $C^k$  convex function  $f: \mathbb{R}^2 \mapsto \mathbb{R}$  and an initialization  $x_0 = (u_0, v_0)$  such that the recursion, for  $i \geq 1$*

$$\begin{aligned} u_{i+1} &\in \underset{u}{\operatorname{argmin}} f(u, v_i) \\ v_{i+1} &\in \underset{v}{\operatorname{argmin}} f(u_{i+1}, v) \end{aligned}$$

*produces a non converging sequence  $(x_i)_{i \in \mathbb{N}} = ((u_i, v_i))_{i \in \mathbb{N}}$ .*

**Proof :** We apply Corollary 1 to the proposed decreasing sequence and by choosing  $(u_0, v_0) = B_2$  for example. This requires to shift indices (start with  $i = 2$ ) and use Theorem 2 with  $I = \mathbb{Z}$ . Note that the optimality condition for partial minimization and the fact that level sets have nonvanishing curvature ensure that the partial minima are unique.  $\square$

In the nonsmooth convex case cyclic minimization is known to fail to provide the infimum value, see e.g., [3, p. 94]. Smoothness is sufficient for establishing value convergence (see e.g. [10, 39] and references therein), whether it is enough or not for obtaining convergence was an open question. Our counterexample closes this question and shows that cyclic minimization does not yield converging sequences even for  $C^k$  convex functions. This result also closes the question for the more general nonconvex case for which we are not aware of a nontrivial counterexample for convergence of alternating minimization. Let us mention however Powell's example [34] which shows that cyclic minimization with three blocks does not converge for smooth functions.

It would also be interesting to understand how our result may impact dual methods and counterexamples in that field, as for instance the recent three blocks counterexample in [17].

### 5.3 Gradient descent with exact line search may not converge

Gradient descent with exact line search is governed by the recursion:

$$x^+ \in \operatorname{argmin} \{f(y) : y = x - t \nabla f(x), t \in \mathbb{R}\},$$

where  $x$  is a point in the plane.

Observe that the step coincides with partial minimization when the gradient  $\nabla f(x)$  is colinear to one of the axis of the canonical basis. From the previous section, we thus deduce the following.

**Corollary 6** (Failure of gradient descent with exact line search). *There exists a  $C^k$  convex function  $f: \mathbb{R}^2 \mapsto \mathbb{R}$  and an initialization  $z_0$  in the plane such that the recursion, for  $i \geq 1$*

$$x_{i+1} \in \operatorname{argmin} \{f(y) : y = x_i - t \nabla f(x_i), t \in \mathbb{R}\}$$

*produces a well defined non converging sequence  $(x_i)_{i \in \mathbb{N}}$ .*

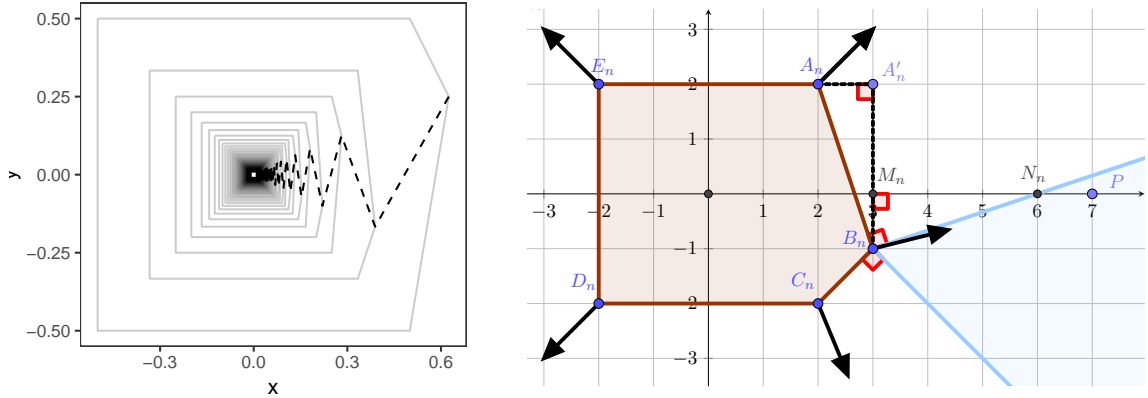


Figure 6: Illustration of the Tikhonov regularization example, on the left in gray, polygons used to build the sublevel sets of the constructed  $f$  and the corresponding solutions to (31) for some values of  $r$  (solutions are joined by dotted lines). On the right the normal to be chosen to apply Lemma 6 (for  $n = 1$ , see main text for details). The point  $P$  represents  $\bar{x}$ , it sits on the  $x$ -axis and is constantly contained in the normal cone at  $B_n$  for any  $n \geq 1$ .

Convergence failure for gradient descent with exact line search is new up to our knowledge. Let us mention that despite non convergence, the constructed sequence satisfy sublinear convergence rates in function values [10].

## 5.4 Tikhonov regularization path may have infinite length

Following [36], we consider for any  $r > 0$

$$x(r) = \operatorname{argmin} \{f(x) + r\|x - \bar{x}\|_2^2 : x \in \mathbb{R}^2\} \quad (31)$$

where  $f$  is  $C^k$  convex and where  $\bar{x}$  is any anchor point. We would like to show that the curve  $r \mapsto x(r)$  may have infinite length. Torralba provided a counterexample in his PhD Thesis for *continuous* convex functions, see [36]. This work extends his result to smooth  $C^k$  convex functions in  $\mathbb{R}^2$ .

For any  $n$  in  $\mathbb{N}^*$ , we set

$$\begin{aligned} A_n &= \left(\frac{2}{n}, \frac{2}{n}\right) \\ B_n &= \left(\frac{2}{n} + \frac{1}{n^2}, -\frac{1}{n}\right) \\ C_n &= \left(\frac{2}{n}, -\frac{2}{n}\right) \\ D_n &= \left(-\frac{2}{n}, -\frac{2}{n}\right) \\ E_n &= \left(-\frac{2}{n}, \frac{2}{n}\right). \end{aligned}$$

This is depicted in Figure 6. For all  $n \geq 1$ , denote by  $M_n$  the point on the  $x$  axis above  $B_n$  and  $N_n$ , the intersection of the normal cone at  $B_n$  and the  $x$  axis. We have

$$\frac{M_n N_n}{M_n B_n} = n \times \frac{M_n N_n}{M_n B_n} = \frac{A'_n B_n}{A_n A'_n} = \frac{3/n}{1/n^2} = 3n,$$

so that for all  $n \geq 1$ ,  $M_n N_n = 3$  and  $N_n = (3 + 2/n + 1/n^2, 0)$ . Choosing  $P = (7, 0)$ , since for  $n \geq 1$ ,  $3 + 2/n + 1/n^2 \leq 6 < 7$ , this shows that  $P$  constantly belongs to the interior of the normal cone at  $B_n$  for all  $n \geq 1$ . The sequence of level sets is constructed as in Figure 6 by considering alternating symmetries with respect to the  $x$ -axis of the sequence of polygons above. It can be checked that the polygons form a strictly decreasing sequence of sets, as for  $n > 1$ , the polygon  $A_n B_n C_n D_n E_n$  is contained in the interior of the square  $A_{n-1} C_{n-1} D_{n-1} E_{n-1}$ . We choose the normal at  $A_n, C_n, D_n, E_n$  to belong to the bisector at the corner and the normal at  $B_n$  to be proportional to the vector  $B_n P$ . Applying Corollary 1, we construct  $f$  and choose  $\bar{x} = P$  in (31) to obtain the following:

**Corollary 7** (A bounded infinite length Tikhonov path). *There exists a  $C^k$  strictly convex function  $f: \mathbb{R}^2 \mapsto \mathbb{R}$  and  $\bar{x} \in \mathbb{R}^2$  such that the curve  $x((0, 1))$  given by (31) has infinite length.*

**Proof :** We apply Corollary 1 with  $I = \mathbb{Z}$  and revert the indices set to match the sequence that we have described. For any  $n \geq 1$  there exists a value of  $\lambda_n$  such that  $f(B_n) = \lambda_n$  and  $\nabla f(B_n)$  is colinear to the vector  $B_n P$ . Set

$$r = \frac{\|\nabla f(B_n)\|}{2B_n P}$$

we have  $\nabla f(B_n) + 2r(B_n - P) = 0$  which is the optimality condition in (31) with  $\bar{x} = P$ . Hence we have shown that there exists a value of  $r$  such that  $B_n$  is the solution to (31). Since  $n$  was arbitrary this is true for all  $n$  and the curve  $r \mapsto x(r)$  has to go through a sequence of points whose second coordinate is of the form  $(-1)^n/n$  for all  $n \geq 1$ . Since this sequence is not absolutely summable, the curve has infinite length.  $\square$

This result is in contrast with the definable case for which we have finite length by the monotonicity lemma, since the whole trajectory is definable and bounded.

## 5.5 Secants of gradient curves at infinity may not converge

**Thom's gradient conjecture and Kurdyka-Mostowski-Parusinski's theorem** A theorem of Lojasiewicz [27] asserts that bounded solutions to the gradient system

$$\dot{x}(t) = -\nabla f(x(t))$$

converge when  $f$  is a real analytic potential. Thom conjectured in [38] that this convergence should occur in a stronger form: trajectories converging to a given  $\bar{x}$  should admit a tangent at infinity, that is

$$\frac{x(t) - \bar{x}}{\|x(t) - \bar{x}\|} \tag{32}$$

should have a limit as  $t \rightarrow \infty$ . Lines passing through  $\bar{x}$  and having (32) as a slope are called *secants of  $x$  at  $\bar{x}$* . This conjecture was proved to be true in [26]. In the convex world, it is well known that solutions to the gradient system converge for general potentials (this is a F  ejer monotonicity argument due to Bruck); see also the original approaches by Manselli and Pucci [30] and Daniilidis et al. [19]. It is then natural to wonder whether this convergence satisfies higher order rigidity properties as in the analytic case. The answer turns out to be negative in general yielding a quite mysterious phase portrait.

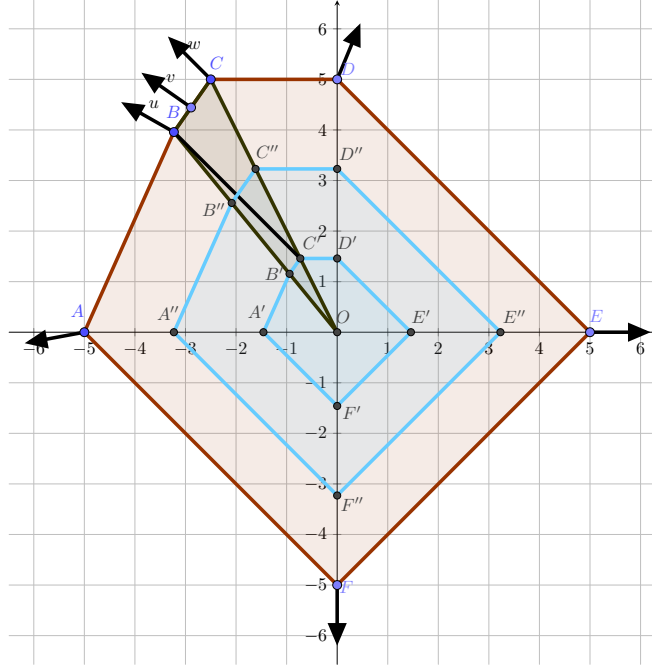


Figure 7:  $A = (-5, 0)$ ,  $B = (-\frac{5}{3} - \frac{25}{16}, \frac{10}{3} + \frac{5}{4})$ ,  $C = (-\frac{5}{2}, 5)$ ,  $D = (0, 5)$ ,  $E = (5, 0)$ ,  $F = (0, -5)$ . All normals are chosen to be bisectors except  $w$  which is parallel to the line  $(DE)$ . The vector  $v$  is orthogonal to the segment  $[BC]$ . The point  $C'$  is obtained by considering the intersection between the line  $(Bw)$  (starting from  $B$  with direction  $w$ ), and the segment  $[OC]$ . The points  $A', B', D', E', F'$  are obtained by performing a scaling of  $A, B, D, E, F$  of a factor  $\frac{OC'}{OC}$ . The polygon  $A''B''C''D''E''F''$  is  $ABCDEF$  scaled by a factor  $\frac{OC'+OC}{2OC}$ .

**Absence of tangential convergence for convex potentials** The construction given in this paragraph is more complex than the previous ones, we start with a technical lemma which will be the basic building block for our counterexample.

**Lemma 9.** *Let  $S$  be a convex set with  $C^k$  boundary interpolating  $ABCDEF$  in Figure 7 and let  $g$  be the gauge function associated to  $S$ . The function  $g$  is differentiable outside the origin. Consider any initialization  $x_0$  in  $[BC]$  with corresponding trajectory solution*

to the equation

$$\begin{aligned}\dot{x}(t) &= -\nabla g(x(t)), \quad t \geq 0, \\ x(0) &= x_0.\end{aligned}$$

Set  $\bar{t} = \sup_{x(t) \in OBC} t$ , we have  $\bar{t} < +\infty$  and  $x(\bar{t})$  in  $[CC']$ .

**Proof :** The fact that  $g$  is differentiable comes from the fact that its subgradient is uniquely determined by the normal cone to  $S$  which has dimension one because of the smoothness of the boundary of  $S$ . Since  $S$  is interpolating the polygon, we have  $g(B) = g(C) = 1$ . Furthermore, we have for all  $t$ ,  $\frac{d}{dt}g(x(t)) = -\|\nabla g(x(t))\|^2 = -1$ , thence  $\bar{t} \leq 1 - g(C')$ . By homogeneity, for any  $x \neq 0$  and  $s > 0$ ,  $\nabla g(sx) = \nabla g(x)$ . For any  $x$  in  $[BC]$ , by convexity

$$0 \leq \langle C - x, \nabla g(C) - \nabla g(x) \rangle = \langle C - B, \nabla g(C) - \nabla g(x) \rangle \frac{\|C - x\|}{\|C - B\|},$$

and therefore

$$-\langle C - B, \nabla g(x) \rangle \geq -\langle C - B, \nabla g(C) \rangle \quad (33)$$

By homogeneity of  $g$ , (33) is true for any  $x$  in the triangle  $OCB$  (different from 0) and thus in the triangle  $C'CB$ . Denote by  $y$  the solution to the equation

$$\begin{aligned}\dot{y} &= -\nabla g(C) \\ y(0) &= B,\end{aligned}$$

which integrates to  $y(t) = B - tw$  for all  $t$ . Equation (33) ensures that for any  $0 \leq t \leq \bar{t}$

$$\frac{d}{dt}(\langle C - B, x(t) \rangle) \geq \frac{d}{dt}(\langle C - B, y(t) \rangle)$$

Hence, we have for any  $0 \leq t \leq \bar{t}$ , integrating on  $[0, t]$

$$\begin{aligned}\langle C - B, x(t) \rangle &\geq \langle C - B, y(t) \rangle + \langle C - B, x_0 - B \rangle \\ &\geq \langle C - B, y(t) \rangle.\end{aligned} \quad (34)$$

Furthermore, for all  $x$  in  $[BC]$ , we have

$$1 = \|\nabla g(x)\|^2 = \frac{1}{\|C - B\|^2} \langle C - B, \nabla g(x) \rangle^2 + \langle v, \nabla g(x) \rangle^2,$$

because  $v$  is orthogonal to  $C - B$ . The first term is maximal for  $x = C$  and thus the second term is minimal for  $x = C$ , we have thus for all  $x$  in  $[BC]$

$$0 < \langle \nabla g(C), v \rangle = \langle -\nabla g(C), -v \rangle \leq \langle \nabla g(x), v \rangle = \langle -\nabla g(x), -v \rangle \leq 1. \quad (35)$$

Equation (35) holds for all  $x$  in  $OCB$  different from  $O$  by homogeneity. We deduce that for all  $0 \leq t \leq \bar{t}$ , we have

$$\frac{d}{dt}(\langle -v, x(t) \rangle) \geq \frac{d}{dt}(\langle -v, y(t) \rangle)$$



and by integration

$$\begin{aligned}\langle -v, x(t) \rangle &\geq \langle -v, y(t) \rangle + \langle -v, x_0 - B \rangle \\ &= \langle -v, y(t) \rangle.\end{aligned}\tag{36}$$

Hence, in the coordinate system  $(C - B, -v)$ , which is orthogonal, for all  $t$  in  $[0, \bar{t}]$ ,  $x(t)$  has larger coordinates than  $y(t)$ .

The trajectory  $y(t)$ , of equation  $t \mapsto B - tw$  is the line going from  $B$  to  $C'$ . From equations (34) and (36), we may write for all  $t$  in  $[0, \bar{t}]$ ,  $x(t) = y(t) + \alpha(t)(C - B) + \beta(t)(-v)$  where  $\alpha$  and  $\beta$  are positive functions. Since  $y(t)$  belongs to the line  $(BC')$ , this shows that  $x(t)$  has to be above this line for all  $t \geq 0$ ,  $t \leq \bar{t}$  and actually,  $x(\bar{t})$  in  $BCC'$ . Hence at time  $\bar{t}$ , we have  $x(\bar{t})$  in  $[CC']$ . This holds true because  $x(\bar{t})$  is on the boundary of  $OCB$  and on the boundary of  $BCC'$ . Hence either  $x(\bar{t})$  in  $[CC']$ , either  $x(\bar{t})$  in  $[BC]$ . Equation (35) ensures that if  $x(\bar{t})$  in  $[BC]$  then  $x(\bar{t}) = C$  which concludes the proof.  $\square$

**Corollary 8** (Secants of gradient curves may all fail to converge). *There exists a  $C^k$  strictly convex function on  $\mathbb{R}^2$  with a unique minimizer  $\bar{x}$ , such that any nonconstant solution to the gradient flow equation*

$$\dot{x}(t) = -\nabla f(x(t))$$

is such that

$$\frac{x(t) - \bar{x}}{\|x(t) - \bar{x}\|}$$

does not have a limit as  $t \rightarrow \infty$ .

The function  $f$  has a positive definite Hessian everywhere except at 0.

**Proof :** We assume without loss of generality that  $\bar{x} = O$  is the origin. Writting  $x(t) = (r(t), \theta(t))$  in polar coordinate, we will construct a function  $f$  such that each solution to the ODE produces nonconverging trajectories  $\theta(t)$ .

We start with an interpolating set  $S_0 = ABCDE$  as in Lemma 9 and let  $S_1 = A'B'C'D'E'$  be its scaled version as described in Figure 7.

Let  $\alpha$  be the value of the angle  $\widehat{BOC}$  and  $m = \lceil \frac{2\pi}{\alpha} \rceil + 1$ . We have

$$\frac{2\pi}{m} < \alpha.$$

To obtain  $S_2$ , we rotate  $S_0$  by an angle  $2\pi/m$ , we denote  $S'_0$  the resulting set. We rescale  $S'_0$  by a factor  $\beta$  in  $(0, 1)$  so that  $\beta S'_0$  lies in the interior of  $S_1$ . Call the resulting set  $S_2$  and  $S_3$  is obtained from  $S_2$  exactly the same way as  $S_1$  is obtained from  $S_0$ . We repeat the same process indefinitely to obtain a strictly decreasing sequence of  $C^k$  sets. Note that for any  $k$  in  $\mathbb{N}$ ,  $S_{2km}$  and  $S_{2km+1}$  are homothetic to  $S_0$ .

We now invoke Corollary 1 (with  $I = \mathbb{Z}$  and revert the indices) to obtain a  $C^k$  function  $f$  with those prescribed level sets. Using Remark 3 it turns out that the level sets of  $f$  between  $S_0$  and  $S_1$  are simple scalings of  $S_0$ . Hence the gradient curves of  $f$  and those of the gauge function of  $S$  are the same between  $S_0$  and  $S_1$ , up to time reparametrization.

Using Lemma 9 any trajectory crossing  $[BC]$  in Figure 7, must also be crossing  $[CC']$  and leave the triangle  $BOC$  in finite time. The same statement holds after scaling the level sets and since for all  $k$  in  $\mathbb{N}$ ,  $S_{2km}$  and  $S_{2km+1}$  are homothetic to  $S_0$ , this shows that no solution stays indefinitely in the triangle  $BOC$ .

Lemma 9 still holds after rotations and by our construction, for any triangle  $T$  obtained by rotating  $BOC$  by a multiple of  $2\pi/m$ , no trajectory stays indefinitely within  $T$ . Since  $2\pi/m < \alpha$ , the union of these triangles  $U$  contains  $O$  in its interior.

Note first that any gradient curve converges to  $\bar{x}$ . Let us argue by contradiction and assume that there exists a continuous gradient curve  $t \mapsto z(t)$  distinct from the stationary solution  $\bar{x}$ , such that

$$\frac{z(t) - \bar{x}}{\|z(t) - \bar{x}\|}$$

converges. This exactly means that the angle  $\theta(t)$  of the curve has a limit in  $[0, 2\pi)$  as  $t$  goes to infinity. There is a rotation of  $BOC$  by a multiple of  $2\pi/m$  whose interior intersects the half line given by the direction  $\theta$ , call this triangle  $T$ . The directional convergence entails that there exists  $t_0 \geq 0$  such that  $z(t)$  belongs to  $T$  for all  $t \geq t_0$ . Hence  $z$  can not be a gradient curve. To complete the proof, we may add disks of increasing size to the list of sets to obtain a full domain function and invoke Theorem 2 with  $I = \mathbb{Z}$ .  $\square$

## 5.6 Newton's flow may not converge

Given a twice differentiable convex function  $f$ , we define the open set  $\Omega := \{x \in \mathbb{R}^2 : \nabla^2 f \text{ is invertible}\}$  and we consider maximal solutions to the differential equation

$$\dot{x}(t) = -\nabla^2 f(x(t))^{-1} \nabla f(x(t)), \quad (37)$$

on  $\Omega$ . This is the continuous counterpart of Newton's method, it has been studied in [5] and [2]. Let  $x_0$  be in  $\Omega$ , there exists a unique maximal nontrivial interval  $I$  containing 0 and a unique solution  $x$  to (37) on  $I$  with  $x(0) = x_0$ . Equation (37) may be rewritten as

$$\frac{d}{dt} \nabla f(x(t)) = -\nabla f(x(t))$$

and thus for all  $t$  in  $I$ , we have

$$\nabla f(x(t)) = e^{-t} \nabla f(x_0). \quad (38)$$

If we could ensure that  $I = \mathbb{R}$  and  $f$  has oscillating gradients close to its minimum, then (38) entails that the direction of the gradient is constant along the solution, which requires oscillations in space to compensate for gradient oscillations.

**Corollary 9** (A bounded Newton's curve that oscillates at infinity). *For any  $k \geq 2$ , there exists a  $C^k$  convex coercive function  $f: \mathbb{R}^2 \mapsto \mathbb{R}$  and an initial condition  $x_0$  in  $\mathbb{R}^2$  such that the solution to (37) is bounded, defined on  $\mathbb{R}$  and has at least two distinct accumulation points.*

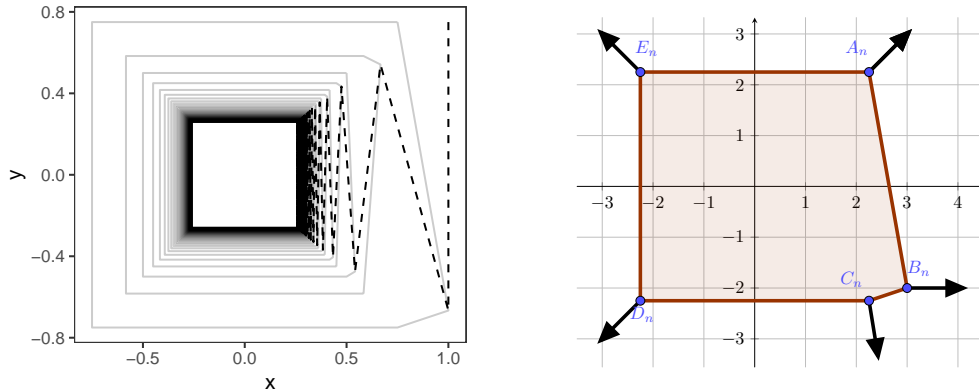


Figure 8: Illustration of the continuous time Newton’s dynamics. On the left, the “skeletons” of the sublevel sets in gray and a sketch of the corresponding curve. On the right, the normals to be chosen in order to apply Lemma 6.

The counterexample is sketched in Figure 8, the construction is the same as for Corollary 5 but instead of doing quarter rotations, we use symmetry with respect to the first axis. We can then call for Corollary 1 to construct the function  $f$  and equation (38) ensures that the solution interval is unbounded.

## 5.7 Bregman descent (mirror descent) may not converge

The mirror descent algorithm was introduced in [31] as an efficient method to solve constrained convex problems. In [9], this method is shown to be equivalent to a projected subgradient method, using non-Euclidean projections. It plays an important role for some categories of constrained optimization problem; see e.g., [6] for recent developments and [20] for a surprising example.

Let us recall beforehand some definitions. Given a Legendre function  $h$  with domain  $\text{dom } h$ , define the *Bregman distance*<sup>4</sup> associated to  $h$  as  $D_h(u, v) = h(u) - h(v) - \langle \nabla h(v), u - v \rangle$  where  $u$  is in  $\text{dom } h$  and  $v$  is in the interior of  $\text{dom } h$ .

Given a smooth convex function  $f$  that we wish to minimize on  $\overline{\text{dom } h}$ , we consider the Bregman method

$$x_{i+1} = \operatorname{argmin} \{ \langle \nabla f(x_i), u - x_i \rangle + \lambda D_h(u, x_i) : u \in \mathbb{R}^2 \},$$

where  $x_0$  is in  $\text{int dom } h$  and  $\lambda > 0$  is a step size. When the above iteration is well defined, e.g. when  $\text{dom } h$  is bounded, it writes:

$$x_{i+1} = \nabla h^* (\nabla h(x_i) - \lambda \nabla f(x_i)).$$

In [6] the authors identified a generalized smoothness condition which confers good min-

---

<sup>4</sup>It is actually not a proper distance.

imizing properties to the above method:

$$Lh - f \text{ convex,} \quad (39)$$

$$\lambda \in (0, L). \quad (40)$$

The corollary below shows that such an algorithm may not converge, even though we assume the cost to satisfy (39), the step to satisfy (40), and the Legendre function to have a compact domain.

**Corollary 10** (Bregman or mirror descent may not converge). *There exists a Legendre function  $h: D \mapsto \mathbb{R}$ , defined on a closed square  $D$ , continuous on  $D$ , a vector  $c$  in  $\mathbb{R}^2$ , and  $x_0$  in  $\mathbb{R}^2$  such that the Bregman descent recursion*

$$x_{i+1} = \nabla h^* (\nabla h(x_i) - c),$$

*produces a bounded sequence  $(x_i)_{i \in \mathbb{N}}$  which has at least two distinct accumulation points.*

**Proof :** We fix  $\theta \in (\frac{-\pi}{4}, \frac{\pi}{4})$ ,  $\theta \neq 0$ , and consider  $h$  constructed in Corollary 3 and choose  $c = (-1, 0)$ . In this case the Bregman descent recursion writes for all  $i$  in  $\mathbb{N}$ ,

$$\nabla h(x_{i+1}) - \nabla h(x_i) = -c$$

so that we actually have  $\nabla h(x_i) - \nabla h(x_0) = \nabla h(x_i) = -ic$  and thus

$$x_i = \nabla h^*(-ic) = \nabla h^*(i, 0).$$

By Corollary 3, we have for all  $i \in \mathbb{N}$  that  $\nabla h^*(i, 0)$  proportional to  $(\cos(\theta), (-1)^i \sin(\theta))$ . Since the norm of the gradient of  $h^*$  cannot vanish at infinity (no flat direction) and is bounded, this proves that the sequence  $(x_i)_{i \in \mathbb{N}}$  has at least two accumulation points which is the desired result.  $\square$

## 5.8 Central paths of Legendre barriers may not converge

Consider the problem

$$\min_{x \in D} \langle c, x \rangle \quad (41)$$

where  $D$  is a subset of  $\mathbb{R}^2$  and  $c$ . Given a Legendre function  $h$  on  $D$ , we introduce the  $h$  central path through

$$x(r) = \operatorname{argmin} \{ \langle c, x \rangle + rh(x) : x \in \mathbb{R}^2 \} \quad (42)$$

where  $r > 0$  is meant to tend to 0. Central paths are one of the essential tools behind interior point methods, see e.g., [33, 4] and references therein.

Note that the accumulation points of  $x(r)$  as  $r \rightarrow 0$ , have to be in the the solution set of (41). It is even tempting to think that the convergence of the path to some specific minimizer could occur, as it is the case for many barriers, see e.g. [4]. We have however:

**Corollary 11** (A nonconverging central path). *There exists a Legendre function  $h: D \mapsto \mathbb{R}$ , defined on a closed square  $D$ , continuous on  $D$ , a vector  $c$  in  $\mathbb{R}^2$ , such that the  $h$  central path  $r \mapsto x(r)$  has two distinct accumulation points.*

**Proof :** The optimality condition which characterizes  $x(r)$  for any  $r > 0$  writes,

$$x(r) = \nabla h^* \left( \frac{c}{r} \right),$$

and the construction is the same as in Corollary 10. □

## 5.9 Hessian Riemannian gradient dynamics may not converge

The construction of this paragraph is similar to the two previous paragraphs. Consider a  $C^k$  ( $k \geq 2$ ) Legendre function  $h: D \mapsto \mathbb{R}$  and the continuous time dynamics

$$\dot{x}(t) = -\nabla_H f(x(t)), t \geq 0, \quad (43)$$

where  $H = \nabla^2 h$  is the Hessian of  $h$  and  $\nabla_H f = H^{-1} \nabla f$  is the gradient of some differentiable function  $f$  in the Riemannian metric induced by  $H$  on  $\text{int } D$ . Such dynamics were considered in [14, 1].

We have the following result:

**Corollary 12** (Nonconverging Hessian Riemannian gradient dynamics). *There exists a Legendre function  $h: D \mapsto \mathbb{R}$ , defined on a closed square  $D$ , continuous on  $D$ , a vector  $c$  in  $\mathbb{R}^2$ , and  $x_0$  in  $\mathbb{R}^2$  such that the solution to (43) with  $f = \langle c, \cdot \rangle$  has two distinct accumulation points.*

**Proof :** Equation (43) may be rewritten

$$\frac{d}{dt} \nabla h(x(t)) = -\nabla f(x(t)),$$

so choosing  $c = (-1, 0)$ , we have for all  $t \in \mathbb{R}$ ,  $\nabla h(x(t)) = \nabla h(x(0)) + (t, 0) = (t, 0)$  and the construction is the same as in Corollary 10. □

## 6 Appendix

**Lemma 10** (Smooth concave interpolation: in between square root and affine). *There exists a  $C^\infty$  strictly increasing concave function  $\phi: [0, 1] \mapsto [0, 1]$  such that*

$$\begin{aligned} \phi(t) &= \sqrt{2t/3} \quad \forall t \leq 1/6 \\ \phi(1) &= 1 \\ \phi'(1) &= 2/3 \\ \phi^{(m)}(1) &= 0, \quad \forall m \geq 2 \end{aligned}$$

**Proof :** Consider a  $C^\infty$  function  $g_0: \mathbb{R} \mapsto [0, 1]$  such that  $g_0 = 1$  on  $(-\infty, -1)$ ,  $g_0 = 0$  on  $(1, +\infty)$  (for example convoluting the step function with a smooth bump function). Set  $g(t) = \frac{1}{2}(g_0(t) + 1 - g_0(-t))$  we have that  $g$  is  $C^\infty$ ,  $g = 1$  on  $(-\infty, -1)$ ,  $g = 0$  on  $(1, +\infty)$  and  $g(t) + g(-t) = 1$  for all  $t$ . We have

$$\int_{-1}^1 g(s) ds = 1$$

$$\int_{-1}^1 \left( \int_{-1}^t g(s) ds \right) dt = 1$$

Set  $\phi_0: [-3, 3] \mapsto \mathbb{R}$ , such that

$$\phi_0(t) = \int_{-3}^t \left( \int_{-3}^r g(s) ds \right) dr.$$

For all  $r$  in  $[-3, 3]$ , we have

$$\int_{-3}^r g(s) ds = \begin{cases} r + 3 & \text{if } r \leq -1 \\ 2 + \int_{-1}^r g(s) ds & \text{if } -1 \leq r \leq 1 \\ 3 & \text{if } r \geq 1 \end{cases}$$

and thus

$$\phi_0(t) = \begin{cases} \frac{t^2}{2} - 9/2 + 3(t + 3) & \text{if } t \leq -1 \\ 2 + 2(t + 1) + \int_{-1}^t \left( \int_{-1}^r g(s) ds \right) dr & \text{if } -1 \leq t \leq 1 \\ 6 + 3(t - 1) & \text{if } 1 \geq t \end{cases}$$

and in particular  $\phi_0(3) = 12$  and  $\phi_0'(3) = 3$ . Set  $\phi_1(s) = \phi_0(6s - 3)/12$ .

$$\phi_1(0) = 0$$

$$\phi_1(t) = \left( \frac{(6t - 3)^2}{2} - 9/2 + 2(3t) \right) / 12 = 3t^2/2 = \text{if } t \leq 1/3$$

$$\phi_1(1) = 1$$

$$\phi_1'(1) = 3/2.$$

$\phi_1$  is stricly increasing, let  $\phi: [0, 1] \mapsto [0, 1]$  denote the inverse of  $\phi_1$ , we have

$$\phi(1) = 1$$

$$\phi'(1) = 2/3$$

$$\phi(t) = \sqrt{2t/3} \text{ if } t \leq 1/6.$$

□

**Lemma 11** (Interpolation inside a sublevel set). *Consider any strictly increasing  $C^k$  function  $\phi: (0, 2) \mapsto \mathbb{R}$  such that  $\phi(1) = 1$  and  $\phi^{(m)}(1) = 0$ ,  $m = 2, \dots, k$ . Then the function*

$$G: (0, 2) \times \mathbb{R}/2\pi\mathbb{Z} \mapsto \mathbb{R}^2$$

$$(s, \theta) \mapsto \phi(s)n(\theta)$$

is diffeomorphism which satisfies for any  $m = 1 \dots, k$  and  $l = 2, \dots, k$ ,

$$\begin{aligned}\frac{\partial^m G}{\partial \theta^m}(1, \theta) &= n^{(m)}(\theta) \\ \frac{\partial^{m+1} G}{\partial \lambda \partial \theta^m}(1, \theta) &= \phi'(1) n^{(m)}(\theta) \\ \frac{\partial^{l+m} G}{\partial \lambda^l \partial \theta^m}(\lambda_-, \theta) &= 0.\end{aligned}$$

**Lemma 12. : Combinatorial Arbogast-Faà di Bruno Formula (from [29]).** Let  $g: \mathbb{R} \mapsto \mathbb{R}$  and  $f: \mathbb{R}^p \mapsto [0, +\infty)$  be  $C^k$  functions. Then we have for any  $m \leq k$  and any indices  $i_1, \dots, i_m \in \{1, \dots, p\}$ .

$$\frac{\partial^m}{\prod_{l=1}^m \partial x_{i_l}} g \circ f(x) = \sum_{\pi \in \mathcal{P}} g^{(|\pi|)}(f(x)) \prod_{B \in \pi} \frac{\partial^{|B|} f}{\prod_{l \in B} \partial x_{i_l}}(x),$$

where  $\mathcal{P}$  denotes all partitions of  $\{1, \dots, m\}$ , the product is over subsets of  $\{1, \dots, m\}$  given by the partition  $\pi$  and  $|\cdot|$  denotes the number of elements of a set. We rewrite this as follows

$$\frac{\partial^m}{\prod_{l=1}^m \partial x_{i_l}} g \circ f(x) = \sum_{k=1}^m \sum_{\pi \in \mathcal{P}_k} g^{(k)}(f(x)) \prod_{B \in \pi} \frac{\partial^{|B|} f}{\prod_{l=1}^m \partial x_{i_l}}(x),$$

where  $\mathcal{P}_k$  denotes all partitions of size  $k$  of  $\{1, \dots, m\}$ .

**Lemma 13.** From [12, Lemma 45] Let  $h$  in  $C^0((0, r_0], \mathbb{R}_+^*)$  be an increasing function. Then there exists a function  $\psi$  in  $C^\infty(\mathbb{R}, \mathbb{R}_+)$  such that  $\psi = 0$  on,  $\mathbb{R}_-$  and  $0 < \psi(s) \leq h(s)$  for any  $s$  in  $(0, r_0]$  and  $\psi$  is increasing on  $\mathbb{R}$

**Lemma 14** (High-order smoothing near the solution set). Let  $D \subset \mathbb{R}^p$  be a nonempty compact convex set and  $f: D \mapsto \mathbb{R}$  convex, continuous on  $D$  and  $C^k$  on  $D \setminus \operatorname{argmin}_D f$ . Assume further that  $\operatorname{argmin}_D f \subset \operatorname{int}(D)$ ,  $k \geq 1$ , with  $\min_D f = 0$ . Then there exists  $\phi: \mathbb{R} \mapsto \mathbb{R}_+$ ,  $C^k$ , convex and increasing with positive derivative on  $(0, +\infty)$ , such that  $\phi \circ f$  is convex and  $C^k$  on  $D$ .

**Proof :** By a simple translation, we may assume that  $\min_D f = 0$  and  $\max_D f = 1$ . Any convex function is locally Lipschitz continuous on the interior of its domain so that  $f$  is globally Lipschitz continuous on  $D$  and its gradient is bounded. Hence,  $f^2$  is  $C^1$  and convex on  $D$ . We now proceed by recursion. For any  $m = 1, \dots, k$ , we let  $Q_m$  denote the  $m$ -order tensor of partial derivatives of order  $m$ . Fix  $m$  in  $\{1, \dots, k\}$ . Assume that  $f$  is  $C^m$  throughout  $D$  while it is  $C^{m+1}$  on  $D \setminus \operatorname{argmin}_D f$ . Note that all the derivatives up to order  $m$  are bounded. We wish to prove that  $f$  is globally  $C^{m+1}$ .

Consider the increasing function

$$\begin{aligned}h: (0, 1] &\mapsto \mathbb{R}_+^* \\ s &\mapsto \frac{s}{1 + \sup_{s \leq f(x) \leq 1} \|Q_{m+1}(x)\|_\infty}\end{aligned}$$

and set  $\psi$  as in Lemma 13. Recall that  $\psi$  is  $C^\infty$  and all its derivative vanish at 0 and  $\psi \leq h$  on  $(0, 1]$ . Let  $\phi$  denote the anti-derivative of  $\psi$  such that  $\phi(0) = 0$ .  $\phi$  is  $C^\infty$  and convex increasing on  $\mathbb{R}$  and, since its derivatives at 0 vanish as well, one has, for any  $q$  in  $\mathbb{N}$ ,  $\phi^{(q)}(z) = o(z)$ . Consider the function  $\phi \circ f$ . It is  $C^m$  on  $D$  and it has bounded derivatives up to order  $m$ . Furthermore, it is  $C^{m+1}$  on  $D \setminus \operatorname{argmin}_D f$ . Let  $\bar{y}$  in  $\operatorname{argmin}_D f$ . If  $\bar{y}$  in  $\operatorname{int}(\operatorname{argmin}_D f)$ , then  $f$  and  $\phi \circ f$  have derivatives of all order vanishing at  $\bar{y}$ . Assuming that  $\bar{y}$  in  $\operatorname{argmin}_D f \setminus \operatorname{int}(\operatorname{argmin}_D f)$ . By the induction assumption and Lemma 12, we have for any indices  $i_1, \dots, i_m \in \{1, \dots, p\}$  and any  $h$  in  $\mathbb{R}^p$ :

$$\begin{aligned} & \frac{\partial^m}{\prod_{l=1}^m \partial x_{i_l}} (\phi \circ f)(\bar{y} + z) - \frac{\partial^m}{\prod_{l=1}^m \partial x_{i_l}} (\phi \circ f)(\bar{y}) \\ &= \frac{\partial^m}{\prod_{l=1}^m \partial x_{i_l}} (\phi \circ f)(\bar{y} + z) \\ &= \sum_{q=1}^m \sum_{\pi \in \mathcal{P}_q} \phi^{(q)}(f(\bar{y} + z)) \prod_{B \in \pi} \frac{\partial^{|B|} f}{\prod_{l=1}^m \partial x_{i_l}}(\bar{y} + z). \end{aligned}$$

All the derivatives of  $f$  are of order less or equal to  $m$  and thus remain bounded as  $z \rightarrow 0$ . Further more  $f$  is Lipschitz continuous on  $D$  so that  $f(\bar{y} + z) = O(\|z\|)$  near 0, and, for any  $q$  in  $\mathbb{N}$ ,  $\phi^{(q)}(f(\bar{y} + z)) = o(\|z\|)$ . Hence  $\phi \circ f$  has derivative of order  $m + 1$  at  $\bar{y}$  and it is 0.

Since  $\operatorname{argmin}_D f \subset \operatorname{int}(D)$ , we may consider any sequence of point  $(y_j)_{j \in \mathbb{N}}$  in  $D \setminus \operatorname{argmin}_D f$  converging to  $\bar{y}$ . By Lemma 12, we have for any indices  $i_1, \dots, i_{m+1} \in \{1, \dots, p\}$ , and any  $j$  in  $\mathbb{N}$ ,

$$\begin{aligned} \frac{\partial^{(m+1)}}{\prod_{l=1}^{m+1} \partial x_{i_l}} (\phi \circ f)(y_j) &= \phi'(f(y_j)) \frac{\partial^{(m+1)} f}{\prod_{l=1}^m \partial x_{i_l}}(y_j) + \sum_{q=2}^{m+1} \sum_{\pi \in \Pi_q} \phi^{(q)}(f(y_j)) \prod_{B \in \pi} \frac{\partial^{|B|} f}{\prod_{l=1}^m \partial x_{i_l}}(x) \\ &\leq h(f(y_j)) \frac{\partial^{(m+1)} f}{\prod_{l=1}^m \partial x_{i_l}}(y_j) + \sum_{q=2}^{m+1} \sum_{\pi \in \Pi_q} \phi^{(q)}(f(y_j)) \prod_{B \in \pi} \frac{\partial^{|B|} f}{\prod_{l=1}^m \partial x_{i_l}}(x) \\ &= f(y_j) \frac{\frac{\partial^{(m+1)} f}{\prod_{l=1}^m \partial x_{i_l}}(y_j)}{1 + \sup_{f(y_j) \leq f(x) \leq 1} \|Q_{m+1}(x)\|_\infty} + O(f(y_j)) \\ &= O(f(y_j)), \end{aligned}$$

where the inequality follows from the construction of  $\phi$ . The third step follows using the definition of  $h$  and the fact that, for any  $q \geq 2$ ,

1. Each partition of  $\{1, \dots, m + 1\}$  of size  $q$  contains subsets of size at most  $m$ . Thus in the product, the terms  $\partial^{|B|} f$  correspond to bounded derivatives of  $f$  by the induction hypothesis.
2.  $\phi^{(q)}(a) = o(a)$  as  $a \rightarrow 0$ .

The last step stems from the fact that the ratio has absolute value less than 1. This shows that the derivatives of order  $m + 1$  of  $\phi \circ f$  are decreasing to 0 as  $j \rightarrow \infty$  and  $\phi \circ f$  is actually  $C^{m+1}$  and convex on  $D$ . The result follows by induction up to  $m = k$  and by the fact that a composition of increasing convex functions is increasing and convex.  $\square$



**Lemma 15.** *Let  $p: \mathbb{R}_+ \mapsto \mathbb{R}_+$  be concave increasing and  $C^1$  with  $p' \geq c$  for some  $c > 0$ . Assume that there exists  $A > 0$  such that for all  $x$  in  $\mathbb{R}_+$*

$$p(x) - xp'(x) \leq A.$$

*Then setting  $a = A/c$ , we have for all  $x \geq a$ ,*

$$p(x - a) - xp'(x - a) \leq 0$$

**Proof :** For all  $x \geq a$ , we have

$$f(x - a) - (x - a)f'(x - a) \leq A,$$

hence

$$f(x - a) - xf'(x - a) \leq A - af'(x - a) \leq A - ac = 0.$$

□

**Acknowledgements.** The authors acknowledge the support of AI Interdisciplinary Institute ANITI funding, through the French “Investing for the Future – PIA3” program under the Grant agreement n°ANR-19-PI3A-0004, Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant numbers FA9550-19-1-7026, FA9550-18-1-0226, and ANR MasDol. J. Bolte acknowledges the support of ANR Chess, grant ANR-17-EURE-0010 and ANR OMS.

## References

- [1] Alvarez, F., Bolte, J. and Brahic, O. (2004). Hessian Riemannian gradient flows in convex programming. *SIAM journal on control and optimization*, 43(2), 477–501.
- [2] Alvarez D. F. and Pérez C., J. M. (1998). A dynamical system associated with Newton’s method for parametric approximations of convex minimization problems. *Applied Mathematics and Optimization*, 38, 193–217.
- [3] Auslender, A., *Optimisation Méthodes Numériques*, Masson, Paris, France, 1976.
- [4] Auslender, A. (1999). Penalty and barrier methods: a unified framework. *SIAM Journal on Optimization*, 10(1), 211–230.
- [5] Aubin, J.-P. and Cellina, A. (1984). *Differential inclusions: set-valued maps and viability theory*. Springer.
- [6] Bauschke, H. H., Bolte, J. and Teboulle, M. (2016). A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research*, 42(2), 330–348.

- [7] Bauschke, H. H., and Combettes, P. L. (2011). *Convex analysis and monotone operator theory in Hilbert spaces* (Vol. 408). New York: Springer.
- [8] Beck, A. (2017). *First-order methods in optimization* (Vol. 25). SIAM.
- [9] Beck, A. and Teboulle, M. (2003). Mirror descent and nonlinear projected sub-gradient methods for convex optimization. *Operations Research Letters*, 31(3), 167–175.
- [10] Beck, A. and Tretuashvili, L. (2013). On the convergence of block coordinate descent type methods. *SIAM journal on Optimization*, 23(4), 2037-2060.
- [11] Bertsekas, D. P., Scientific, A. (2015). *Convex optimization algorithms*. Belmont: Athena Scientific.
- [12] Bolte, J., Daniilidis, A., Ley, O., and Mazet, L. (2010). Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6).
- [13] Bolte, J., Nguyen, T. P., Peypouquet, J., and Suter, B. W. (2017). From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2), 471–507.
- [14] Bolte, J., and Teboulle, M. (2003). Barrier operators and associated gradient-like dynamical systems for constrained minimization problems. *SIAM journal on control and optimization*, 42(4), 1266-1292.
- [15] Borwein, J. M., Li, G., and Yao, L. (2014). Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM Journal on Optimization*, 24(1), 498–527.
- [16] Boyd, S., and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- [17] Chen, C., He, B., Ye, Y., and Yuan, X. (2016). The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155(1-2), 57–79.
- [18] Crouzeix, J.-P. (1980). Conditions for convexity of quasiconvex functions. *Mathematics of Operations Research*, 5(1), 120-125.
- [19] Daniilidis, A., Ley, O., and Sabourau, S. (2010). Asymptotic behaviour of self-contracted planar curves and gradient orbits of convex functions. *Journal de mathématiques pures et appliquées*, 94(2), 183–199.
- [20] Dragomir, R. A., Taylor, A., d’Aspremont, A., and Bolte, J. (2019). Optimal complexity and certification of Bregman first-order methods. *arXiv preprint arXiv:1911.08510*.
- [21] Fenchel, W., *Convex Cones, Sets and Functions*, Mimeographed lecture note, Princeton University, 1951.

- [22] de Finetti, B. (1949). Sulle stratificazioni convesse. *Annali di Matematica Pura ed Applicata*, 30(1), 173–183.
- [23] Gale, D., Klee, V. and Rockafellar, R. T. (1968). Convex functions on convex polytopes. *Proceedings of the American Mathematical Society*, 19(4), 867–873.
- [24] Golub, G. H., Hansen, P. C., and O’Leary, D. P. (1999). Tikhonov regularization and total least squares. *SIAM Journal on Matrix Analysis and Applications*, 21(1), 185–194.
- [25] Kannai, Y. (1977). Concavifiability and constructions of concave utility functions. *Journal of mathematical Economics*, 4(1), 1–56.
- [26] Kurdyka, K., Mostowski, T. and Parusinski, A. (2000). Proof of the gradient conjecture of R. Thom. *Annals of Mathematics*, 152(3), 763–792.
- [27] Łojasiewicz, S. (1984), Sur les trajectoires du gradient d’une fonction analytique. *Seminari di Geometria, Bologna (1982/83)*, Università degli Studi di Bologna, Bologna (1984), 115–117.
- [28] Lorentz, G. G. (1954). Bernstein polynomials. *American Mathematical Soc.*
- [29] Ma, T. W. (2009). Higher chain formula proved by combinatorics. *The electronic journal of combinatorics*, 16(1).
- [30] Manselli, P., and Pucci, C. (1991). Maximum length of steepest descent curves for quasi-convex functions. *Geometriae Dedicata*, 38(2), 211–227.
- [31] A. S. Nemirovsky and D. B. Yudin (1983). *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience, New York.
- [32] Nesterov, Y. (2003). *Lectures on convex optimization (Vol. 137)*. Springer.
- [33] Nesterov, Y., and Nemirovskii, A. (1994). *Interior-point polynomial algorithms in convex programming (Vol. 13)*. Siam.
- [34] Powell, M. J. (1973). On search directions for minimization algorithms. *Mathematical programming*, 4(1), 193–201.
- [35] Schneider, R. (1993). *Convex bodies: the Brunn-Minkowski theory (No. 151)*. Cambridge university press.
- [36] Torralba, D., *Convergence épigraphique et changements d’échelle en analyse variationnelle et optimisation*, Ph.D. Thesis (Université Montpellier 2, 1996).
- [37] Rockafellar, R. T. (1970). *Convex analysis (Vol. 28)*. Princeton university press.
- [38] Thom, R. (1989). Problèmes rencontrés dans mon parcours mathématique : un bilan. *Publications mathématiques de l’IHES*, 70, 199–214.
- [39] Wright, S. J. (2015). Coordinate descent algorithms. *Mathematical Programming*, 151(1), 3–34.