



# Smooth Approximation of Lipschitz Maps and Their Subgradients

ABBAS EDALAT, Imperial College London

We derive new representations for the generalised Jacobian of a locally Lipschitz map between finite dimensional real Euclidean spaces as the lower limit (i.e., limit inferior) of the classical derivative of the map where it exists. The new representations lead to significantly shorter proofs for the basic properties of the subgradient and the generalised Jacobian including the chain rule. We establish that a sequence of locally Lipschitz maps between finite dimensional Euclidean spaces converges to a given locally Lipschitz map in the L-topology—that is, the weakest refinement of the sup norm topology on the space of locally Lipschitz maps that makes the generalised Jacobian a continuous functional—if and only if the limit superior of the sequence of directional derivatives of the maps in a given vector direction coincides with the generalised directional derivative of the given map in that direction, with the convergence to the limit superior being uniform for all unit vectors. We then prove our main result that the subspace of Lipschitz  $C^\infty$  maps between finite dimensional Euclidean spaces is dense in the space of Lipschitz maps equipped with the L-topology, and, for a given Lipschitz map, we explicitly construct a sequence of Lipschitz  $C^\infty$  maps converging to it in the L-topology, allowing global smooth approximation of a Lipschitz map and its differential properties. As an application, we obtain a short proof of the extension of Green's theorem to interval-valued vector fields. For infinite dimensions, we show that the subgradient of a Lipschitz map on a Banach space is upper continuous, and, for a given real-valued Lipschitz map on a separable Banach space, we construct a sequence of Gateaux differentiable functions that converges to the map in the sup norm topology such that the limit superior of the directional derivatives in any direction coincides with the generalised directional derivative of the Lipschitz map in that direction.

CCS Concepts: • **Mathematics of computing** → **Nonconvex optimization**; *Functional analysis*; *Approximation*; **Point-set topology**; • **Theory of computation** → **Constructive mathematics**;

Additional Key Words and Phrases: Non-smooth optimisation, generalised Jacobian, subgradient, Lipschitz maps, L-topology

## ACM Reference format:

Abbas Edalat. 2021. Smooth Approximation of Lipschitz Maps and Their Subgradients. *J. ACM* 69, 1, Article 8 (December 2021), 32 pages.  
<https://doi.org/10.1145/3481805>

## 1 INTRODUCTION

Lipschitz maps between metric spaces (i.e., maps named after Rudolph Lipschitz that increase distances by at most a given factor) provide a fundamental class of functions in pure and applied

Author's address: A. Edalat, Department of Computing, Imperial College London, 180 Queen's Gate, London, SW72RH; email: ae@ic.ac.uk.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2021 Copyright held by the owner/author(s).  
0004-5411/2021/12-ART8 \$15.00  
<https://doi.org/10.1145/3481805>

mathematics as well as a variety of areas of computation including optimisation, control theory, geometric modelling and machine learning.

In all these areas, we often need to deal with functions that are indeed locally Lipschitz but not differentiable. Lipschitz maps naturally arise as any composition of functions consisting of piecewise continuously differentiable functions, the absolute value function or the maximum or minimum of a finite set of such functions [11]. In machine learning, the least absolute deviation method or  $\ell_1$  loss gives a non-differentiable but Lipschitz map as an objective function. In deep learning, currently the most widely used activation function in feed forward neural networks is the Rectified Linear Unit (ReLU), which like the absolute value function fails to be differentiable at 0 but is Lipschitz with Lipschitz constant 1 [39].

Lipschitz maps, closed under composition, contain the important class of piecewise polynomial functions that are supported in basic mathematical software such as MatLab [14] and are widely used in geometric modelling, approximation and interpolation [14]. They are uniformly continuous and, in contrast to differentiable maps, are closed under the fundamental min, max operations and absolute value on functions. Lipschitz maps with uniformly bounded Lipschitz constants are also closed under convergence with respect to the sup norm topology. In addition, there is the distinguished property in the theory and applications of differential equations that a Lipschitz vector field in  $\mathbb{R}^n$  has a unique solution to the initial value problem [13].

On a more theoretical level, Lipschitz maps between finite dimensional Euclidean spaces are, by Rademacher's theorem, differentiable almost everywhere [12, p. 148]. By Kirszbraun's theorem [32, p. 202], a Lipschitz map from a subset of a finite dimensional Euclidean space to another such space can be extended to the whole space with the same Lipschitz constant. Lipschitz maps are also at the very foundation of non-linear functional analysis [9].

It is against this background that Frank Clarke introduced, for locally Lipschitz maps, the notion of generalised directional derivative that he used to define the subgradient, as a set-valued derivative with values that are non-empty, compact and convex sets; the subgradient extends a similar concept for convex functions [11]. A closely related notion of generalised Jacobian was introduced by him for vector Lipschitz maps. The emerging subject of set-valued analysis and non-smooth optimisation has grown to impact many areas of computation in engineering and applied mathematics [3]. The subgradient, also called *subdifferential*, plays a key role in several areas of mathematical computation which deal with Lipschitz maps including non-smooth dynamical systems, differential inclusions, calculus of variations and optimal control [2, 4, 11].

A well-established method to tackle non-smooth optimisation problems is to find a suitable smooth approximation to the minimisation problem and then use standard techniques such as the gradient descent algorithm [45]. A major challenge in optimisation is therefore to find suitable smooth approximations to non-smooth minimisation problems [5]. There is a large literature to design smooth approximations to the problem when dealing with convex optimisation that use the notion of convex conjugate of the underlying function [7].

In this article, adopting a new inter-disciplinary approach, several new representations for the subgradient and the generalised Jacobian of locally Lipschitz maps are introduced which are the basis of a framework for smooth approximation of Lipschitz maps and their differential properties. From an optimisation point of view, we develop a smooth approximation method for problems that are Lipschitz but may not be smooth or convex. We focus here on the foundational mathematical results and leave the algorithmic consequences in various areas of optimisation to future work.

We first apply domain theory, a branch of order theory, to derive some new representations and basic results about the subgradient and the generalised Jacobian of locally Lipschitz maps. Domain theory was developed, on the one hand, by Dana Scott as a mathematical model of computation for the denotational semantics of programming languages [42, 43], particularly functional

programming languages, and, on the other hand, independently by several groups of mathematicians, in various mathematical contexts, including Karl Hofmann, Jimmie Lawson, Mike Mislove, Al Stralka, Klaus Keimel, Gerhard Gierz and Yuri Ershove (see [33]). It was later used to build new computational models in several other areas including exact real number computation, computational geometry, measure and integration theory and solutions of ODEs [16–18, 23, 24, 28–30]. In addition, employing domain-theoretic methods, the so-called L-derivative was introduced for real and complex Lipschitz maps [22, 25], and it was subsequently shown that the L-derivative coincides with the subgradient for real-valued functions on finite dimensional Euclidean spaces [19]. This latter result has later been extended by Hertling to real-valued Lipschitz maps on Banach spaces [34].

Domains are partially ordered sets equipped with notions of completeness and approximation. For example, the *upper space* of a finite dimensional Euclidean space, namely the collection of non-empty compact subsets of the space partially ordered with reverse inclusion, is a domain, as is its subcollection of convex subsets, the fundamental domain under consideration in the present work. Domains are  $T_0$ -topological spaces endowed with their Scott topology. This topology for the upper space coincides with the upper topology, which is widely used in optimisation theory: set-valued functions continuous with respect to this topology are called *upper semicontinuous multifunctions* [11, p. 29]. Elements of domain theory required in this article are outlined in Section 2.

Here, we employ the notion of the lower limit or limit inferior of a map from a dense subset of a topological space without isolated points into a bounded complete domain. It is closely related to the notion of lower envelope of such a map as introduced in the work of Gierz et al. [33, Exercise II-3.19]) in connection to bounded complete domains as densely injective spaces [33, p. 182]). We show in Section 3 that the subgradient and the generalised Jacobian of locally Lipschitz maps are the lower limits, equivalently the lower envelopes, of the classical derivative of the map which exists on a dense subset. This new representation leads to significantly shorter proofs for several basic properties of the subgradient and the generalised Jacobian.

Our key results, however, use the L-topology on the space of locally Lipschitz maps which was introduced in the work of Edalat [21] and is defined using the Scott topology on the space of Scott (equivalently upper) continuous functions from a finite dimensional Euclidean space to the domain of non-empty, convex and compact subsets of the space. In fact, this function space inherits the partial order of the ambient domain, by pointwise ordering of functions, which makes the function space itself into a bounded complete domain. Given that the subgradient of a locally Lipschitz map belongs to this function space, the L-topology is defined as the weakest refinement of the sup norm topology on the space of locally Lipschitz maps that makes the subgradient a continuous functional into this function space equipped with its Scott topology. The L-topology admits a complete metric and is weaker than the Lipschitz norm topology [21].

We derive simple necessary and sufficient conditions for a sequence of locally Lipschitz maps between finite dimensional Euclidean spaces to converge in the L-topology to a given locally Lipschitz map: the limit superior of the directional derivatives of the maps in the sequence in a given vector direction must coincide with the generalised directional derivative of the given map in that direction, with the convergence to limit superior being uniform for all unit vectors. We will then prove that in  $C^\infty$  maps are dense in the space of Lipschitz maps between finite dimensional Euclidean spaces with respect to the L-topology.

It is well known that convolution with a smooth function can make a given function smooth [10]. It is also well known that given a Lipschitz map  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , with Lipschitz constant  $c$ , one can construct a sequence of  $C^\infty$  functions with Lipschitz constant  $c$  such that the sequence converges in the sup norm to  $f$ . In fact, by taking a sequence of  $C^\infty$  test functions  $\phi_n : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $\phi_n(x) = 0$  for  $\|x\| > 1/n$  and  $\int_{\mathbb{R}^n} \phi_n(x) dx = 1$  for all  $n \geq 1$ , the sequence of convolutions  $f_n = f * \phi_n$  with

$f_n(x) = \int_{\mathbb{R}^n} f(y) \phi_n(x - y) dy$  converges to  $f$  in the sup norm and each function  $f_n$  is  $C^\infty$  with Lipschitz constant  $c$  [6, p. 1]. It is easy to see that this result also holds for Lipschitz maps of type  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . In this article, we will extend these results and show that the limit superior of the sequence of directional derivatives of the maps  $f_n$  in a given vector direction  $v$  always coincides with the generalised directional derivative of the Lipschitz map in that direction  $v$ , with the convergence to the limit superior being uniform for all unit vectors. This will therefore give us a notion of approximation of Lipschitz maps by a sequence of  $C^1$  maps which entails convergence in the sup norm topology to the Lipschitz map and, in addition, preserves the differential properties of the sequence in the limit. This result then can be viewed as the basis for a method for smooth global approximation of a Lipschitz map. We will provide an elementary self-contained account of these results for the reader.

To this end, given a Lipschitz map between finite dimensional Euclidean spaces, we take its convolution with a sequence of Gaussian probability distributions, as our test functions, and explicitly construct a sequence of  $C^\infty$  maps convergent to it in the L-topology, implying that the limit superior of the directional derivatives of the functions in the sequence in a given vector direction coincide with the generalised directional derivative of the Lipschitz map in that direction, with the convergence to limit superior being uniform for all unit vectors. This result does not depend on the particular choice of the sequence of Gaussian probability distributions we use as our test functions. We first present the proof of the preceding property for scalar Lipschitz maps in Section 5, which is elementary, and formulate the more sophisticated proof in the case of vector Lipschitz maps in Section 7, which uses Imbert's expression for the generalised Jacobian [36]. These approximation properties of Lipschitz maps give credence to the claim that the subgradient of a Lipschitz map and the generalised Jacobian of a Lipschitz vector map are the generalisations of the classical derivative for the respective Lipschitz maps.

As an application of the preceding results, we present a short proof of the extension of Green's theorem to interval-valued vector fields in Section 6.

Last, in Section 8, we consider real-valued Lipschitz maps on Banach spaces, which include all  $L^p$  function spaces ( $0 < p \leq \infty$ ) used widely in optimisation. The subgradient at a point of the Banach space is now a non-empty, convex and compact set of the dual of the Banach space equipped with its weak\* topology. Using the equivalence of the L-derivative with the subgradient, we provide a short proof that the subgradient of a Lipschitz map is upper continuous. Then, for a given Lipschitz map on a separable Banach space, we use a non-degenerate Gaussian measure on the space to construct a sequence of Gateaux differentiable maps that converges to the Lipschitz map in the sup norm topology such that the limit superior of the sequence of directional derivatives of the maps in a given direction coincides with the generalised directional derivative of the Lipschitz map in that direction. We conclude the article in Section 9.

Since the L-topology is defined using the Scott topology of a function space, which is only  $T_0$  with no classical counterpart, we argue that the results in this article are based fundamentally on an interdisciplinary approach bridging domain theory as developed in the Scott theory of computation on the one hand and non-smooth optimisation on the other. An attempt has been made to use more elementary mathematical notions and results in the earlier sections related to real-valued Lipschitz maps on finite dimensional Euclidean spaces and employ more advanced and recent results in analysis only in the later sections relating to vector Lipschitz maps and Lipschitz maps on Banach spaces.

### 1.1 Notation and Terminology

We first recall that a map  $f : X \rightarrow Y$  of two metric spaces is *Lipschitz* if there exists  $c \geq 0$  such that  $d_Y(f(x_1), f(x_2)) \leq cd_X(x_1, x_2)$  for all  $x_1, x_2 \in X$ ; the map  $f$  is *locally Lipschitz* if any point in  $X$

has an open neighbourhood in which  $f$  is Lipschitz. For any positive integer  $n$ , we equip  $\mathbb{R}^n$  with the Euclidean norm  $\|v\| = \sqrt{\sum_{i=1}^n |v_i|^2}$ . The usual inner product of vectors  $x, y \in \mathbb{R}^n$  is denoted by  $\langle x, y \rangle := \sum_{i=1}^n x_i y_i$  in the standard Cartesian coordinate system of  $\mathbb{R}^n$ . If  $C \subseteq \mathbb{R}^n$  is a non-empty compact convex set and  $x \in \mathbb{R}^n$ , we write  $\langle C, x \rangle = \langle x, C \rangle := \{\langle y, x \rangle : y \in C\}$ , which is a compact interval of  $\mathbb{R}$ . For a vector  $v \in \mathbb{R}^n$  and a linear map of type  $\mathbb{R}^n \rightarrow \mathbb{R}^m$ , represented in the standard coordinate system by the matrix  $A \in \mathbb{R}^{m \times n}$ , the value of the linear map  $A$  at  $v$  is written as usual by  $Av \in \mathbb{R}^m$  with  $(Av)_i = \sum_{j=1}^n A_{ij} v_j$ . The transpose of a matrix  $M \in \mathbb{R}^{m \times n}$  is denoted by  $M^T$ , the unit sphere centred at the origin in  $\mathbb{R}^n$  by  $S^{n-1}$  and the closed ball centred at  $x \in \mathbb{R}^n$  and radius  $c > 0$  by  $B_c(x)$ . The closure, respectively, the interior, complement and boundary of a subset  $S \subseteq X$  of a topological space  $X$  are denoted by  $\bar{S}$ , respectively,  $S^\circ$ ,  $S^c$  and  $S^b$ . For two topological spaces  $X$  and  $Y$ , the set of continuous functions from  $X$  to  $Y$  is denoted by  $(X \rightarrow Y)$ . When it is convenient and there is no ambiguity, we identify  $x$  and  $\{x\}$  for  $x \in X$ . For a non-empty, compact convex set  $C \subseteq \mathbb{R}^n$ , the  $\epsilon$ -open neighbourhood of  $C$  is denoted by  $C_\epsilon := \{x \in \mathbb{R}^n : \exists y \in C. \|x - y\| < \epsilon\}$ , which is convex. For subsets  $A, B \subseteq \mathbb{R}^n$ , the *Minkowski* sum and difference are defined by  $A \pm B = \{a \pm b : a \in A, b \in B\}$ .

The collection of Borel subsets of a topological space (i.e., the smallest  $\sigma$ -algebra containing the open sets of  $X$ ) is denoted by  $\mathcal{B}_X$ . Given a map  $f : X \rightarrow Y$  of spaces  $X$  and  $Y$ , we denote by  $f[A]$  the image of a subset  $A \subseteq X$  under  $f$ . If  $f$  is continuous and  $\mu$  is a Borel measure on  $X$ , then the induced pushforward measure  $\mu_f$  on  $Y$  is given by  $\mu_f(B) = \mu(f^{-1}(B))$  for  $B \in \mathcal{B}_Y$ .

We equip the vector space  $\mathbb{R}^{m \times n}$  of  $m \times n$  matrices over  $\mathbb{R}$  with the Frobenius norm  $\|A\|_F$  given by

$$\|A\|_F^2 = \sum_{1 \leq i \leq m} \sum_{1 \leq j \leq n} |A_{ij}|^2$$

for  $A \in \mathbb{R}^{m \times n}$ . Note that the Frobenius norm, which for vectors coincides with the Euclidean norm, is subordinate to the Euclidean norm of vectors (i.e.,  $\|Av\| \leq \|A\|_F \|v\|$ ). For convenience, we write  $\|M\|_F$  as simply  $\|M\|$ . Recall also that for matrices of type  $\mathbb{R}^{m \times n}$ , the Frobenius norm, like any other matrix norm, induces the Euclidean topology on  $\mathbb{R}^{mn}$ , which in this article is taken as the topology of linear maps of type  $\mathbb{R}^n \rightarrow \mathbb{R}^m$ .

Let  $X$  and  $Y$  be normed vector spaces and  $L(X, Y)$  the normed vector space of bounded linear operators from  $X$  to  $Y$  with the operator norm. The *one-sided directional derivative*  $f'(x; v)$  of  $f : U \subseteq X \rightarrow Y$  at  $x \in U$  in the direction  $v \in X$  is given by

$$f'(x; v) = \lim_{t \rightarrow 0^+} \frac{f(x + tv) - f(x)}{t}$$

when the limit exists. The *directional derivative*  $df(x; v)$  of  $f : U \subseteq X \rightarrow Y$  at  $x \in U$  in the direction  $v \in \mathbb{R}^n$  is given by

$$df(x; v) = \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t}$$

when the limit exists, in which case  $df(x; v) = f'(x; v)$ . The Gateaux derivative of  $f$  at a point  $x \in X$ , if it exists, is a bounded linear operator  $df(x) : X \rightarrow Y$  such that for all  $v \in X$  we have  $(df(x))(v) = df(x; v)$  [9]. The *Fréchet derivative* [9, 48] of  $f$  at  $x \in U$ , if it exists, is a bounded linear map  $L : X \rightarrow Y$  with

$$\lim_{\|h\| \rightarrow 0} \frac{\|f(x + h) - f(x) - L(h)\|}{\|h\|} = 0.$$

If  $f$  is Fréchet differentiable, then it is Gateaux differentiable and the two linear maps coincide.

The convex hull of a subset  $A$  of a topological vector space is denoted by  $\text{co}\{A\}$ , or simply  $\text{co } A$  if no ambiguity arises. We write the closure of  $\text{co}\{A\}$  by  $\overline{\text{co}}\{A\} := \overline{\text{co}\{A\}}$ , which is called the *closed convex hull* of  $A$  and is equal to the intersection of closed convex sets that contain  $A$  [35, p. 31].



Recall that the *Hausdorff metric*  $d_{\mathcal{H}}$  on the set of non-empty compact subsets of  $\mathbb{R}^n$  is defined by  $d_{\mathcal{H}}(C, D) = \inf\{r > 0 : C \subseteq D_r \text{ and } D \subseteq C_r\}$ . The *support function*  $\sigma_C : \mathbb{R}^n \rightarrow \mathbb{R}$  of a non-empty convex set  $C \subseteq \mathbb{R}^n$  is defined by  $\sigma_C(x) = \sup\{\langle y, x \rangle : y \in C\}$ . We state some of its basic properties here [41, 1.7.1 and 1.8.14]: The support function is convex, thus continuous, and, for non-empty compact and convex sets  $C_1$  and  $C_2$ , we have

$$\|\sigma_{C_1} - \sigma_{C_2}\|_{\infty} = d_{\mathcal{H}}(C_1, C_2), \quad (1)$$

where  $\|\cdot\|_{\infty}$  is the uniform norm on the sphere  $S^{n-1}$ ; moreover,  $C_1 \subseteq C_2$  iff  $\sigma_{C_1} \leq \sigma_{C_2}$ —that is, any non-empty convex compact set is completely determined by its support function.

## 2 DOMAIN THEORY

Domain theory was introduced by Dana Scott in computer science as a mathematical model of computation in particular for developing denotational semantics for programming languages [42, 43]. The so-called algebraic domains were used to develop mathematical models of  $\lambda$ -calculus for the denotational semantics of functional programming language [47]. It was later shown that non-algebraic domains, the so-called continuous domains which will be employed in this work, can model mathematical computation in a variety of areas, including exact real number computation [30], computational geometry [25, 26], measure and integration theory [16, 17, 20], differential calculus [25], solution of ODEs [29] and hybrid systems [28].

We review the elements of domain theory we need in this article [1, 33]. A directed complete partial order (dcpo) is a partial order  $(D, \sqsubseteq)$  in which every directed subset has a supremum. The partial order relation  $x \sqsubseteq y$  is considered to imply that  $y$  has more information than  $x$ . A subset  $O \subseteq D$  is an open subset of the *Scott topology* of  $D$  if  $O$  is an upper set (i.e.,  $y \sqsupseteq x \in O \Rightarrow y \in O$ ) and inaccessible by suprema of directed sets (i.e., if  $A \subseteq D$  is directed then  $\sup A \in O \Rightarrow \exists x \in A. x \in O$ ). This topology is  $T_0$ . A map  $f : D \rightarrow E$  of dcpo's is Scott continuous iff it is monotone (i.e.,  $x \sqsubseteq y$  implies  $f(x) \sqsubseteq f(y)$ ) and preserves suprema of directed sets (i.e., if  $A \subseteq D$  is directed then  $f(\sup A) = \sup f[A]$ ).

For  $x, y \in D$ , we have  $x$  *way-below*  $y$ , denoted  $x \ll y$ , if for every directed subset  $A \subseteq D$  the relation  $y \sqsubseteq \sup A$  implies there exists  $a \in A$  with  $x \sqsubseteq a$ . The way-below relation refines the partial order (i.e.,  $x \ll y$  implies  $x \sqsubseteq y$ ). The idea is that when  $x \ll y$ ,  $x$  is a *finitary approximation* to  $y$  in the sense that if the subset  $A$  represents a set of finitary information whose total aggregate provides more information than  $y$ , then there is already a piece of information in  $A$  that exceeds the information in  $x$ . We say  $B \subseteq D$  is a *basis* of  $D$  if the set of elements in  $B$  way-below any element  $y \in D$  is directed with supremum  $y$ —that is,  $B_y := \{x \in B : x \ll y\}$  is directed with  $y = \sup B_y$ . A (countably based) *domain* is a dcpo with a (countable) basis. If  $D$  is a domain, then  $D$  itself is a basis for  $D$ . Two given elements  $x, y \in D$  of a domain have the following simple property widely used in this work: if for all  $z \in D$  the relation  $z \ll x$  implies  $z \sqsubseteq y$ , then we have  $x \sqsubseteq y$  (since in fact  $x$  is the supremum of all such  $z$ ). The way-below relation in a domain  $D$  has the so-called *interpolation property*, very useful in practice: if  $x \ll y$ , then there exists  $z \in B$ , where  $B \subseteq D$  is any given basis, such that  $x \ll z \ll y$ . A domain is *bounded complete* if every bounded set of elements has a least upper bound. Since the empty set is trivially bounded, a bounded complete domain has a least element, denoted  $\perp$ . The Scott topology of a domain with basis  $B$  has basic open sets given by  $\uparrow x := \{y \in D : x \ll y\}$  for  $x \in B$ .

Some of the basic examples of countably based domains in mathematical analysis and computation are related to finite dimensional Euclidean spaces. The lattice  $\Omega\mathbb{R}^n$  of open subsets of  $\mathbb{R}^n$ , ordered by subset inclusion, is a bounded complete domain with a greatest and a least element (i.e., a continuous lattice) in which  $O_1 \ll O_2$  iff  $\overline{O_1}$  is compact with  $\overline{O_1} \subseteq O_2$  (i.e., iff an  $\epsilon$ -neighbourhood of  $O_1$  is contained in  $O_2$ ). The interiors of rational polytopes provide a countable basis for  $\Omega\mathbb{R}^n$ .

The upper space  $\mathbf{UR}^n$  of  $\mathbb{R}^n$ , consisting of non-empty compact subsets of  $\mathbb{R}^n$  ordered by reverse inclusion and augmented with  $\mathbb{R}^n$  as the least element, is a bounded complete countably based domain with  $C_1 \ll C_2$  iff  $(C_1)^\circ \supseteq C_2$ —that is, iff an  $\epsilon$ -neighbourhood of  $C_2$  is contained in  $C_1$  [17]. A countable basis is given by rational polytopes. The Scott topology on  $\mathbf{UR}^n$  coincides with the *upper* topology which has basic open subsets of the form  $\{C \in \mathbf{UR}^n : C \subseteq O\}$ , for any open set  $O \subseteq \mathbb{R}^n$ . The singleton map  $e : \mathbb{R}^n \rightarrow \mathbf{UR}^n$  with  $e(x) = \{x\}$  is an embedding of  $\mathbb{R}^n$  onto the set of maximal elements of  $\mathbf{UR}^n$  equipped with the Scott topology. For convenience, we identify  $x \in \mathbb{R}^n$  with  $\{x\} \in \mathbf{UR}^n$  (i.e., we consider  $\mathbb{R}^n \subseteq \mathbf{UR}^n$ ). The upper space has several sub-domains relevant to our work: (i)  $\mathcal{CR}^n$  consisting of the convex subsets in  $\mathbf{UR}^n$ , which is the fundamental domain for the results of this work; (ii)  $\mathbf{I}[0, 1]^n$  consisting of axes-aligned hyper-rectangles  $\prod_{i=1}^n [a_i^-, a_i^+] \subseteq [0, 1]^n$  with  $a_i^- \leq a_i^+$  for  $1 \leq i \leq n$ , where a non-empty compact interval  $a \subseteq \mathbb{R}$  is written as  $a = [a^-, a^+]$ ; and (iii)  $\mathbf{IA}$ , for an open set  $A \subseteq \mathbb{R}^n$ , consisting of axes-aligned hyper-rectangles contained in  $A$ . We will widely use the interpolation property of the way-below relation  $O_1 \ll O_2$  in the lattice of open sets of  $\mathbb{R}^n$  and that of the way-below relation  $C_1 \ll C_2$  in  $\mathcal{CR}^n$  in this work; the reader unfamiliar with domain theory can simply regard the way-below relation on such pairs of subsets as a shorthand notation for the preceding relation between open sets or between compact sets.

From basic domains, one can construct higher-order domains, including domains of functions. In particular, if  $X$  is a topological space with a continuous lattice of open sets and  $(D, \sqsubseteq_D)$  is a bounded complete domain with basis  $B \subseteq D$ , then the function space  $(X \rightarrow D)$  consists of the collection of Scott continuous functions  $f : X \rightarrow D$  partially ordered by pointwise ordering (i.e.,  $f_1 \sqsubseteq f_2$  if  $\forall x \in X. f_1(x) \sqsubseteq_D f_2(x)$ ). Then  $(X \rightarrow D)$  is itself a bounded complete domain with a basis of step functions which we will now define [33, p. 200]. A *single-step function*  $b\chi_O : X \rightarrow D$  is given by  $b \in D$  and a basic open subset  $O \subseteq X$  with  $b\chi_O(x) = b$  if  $x \in O$  and  $\perp$  otherwise. A *step function* is the least upper bound of a finite bounded set of single-step functions. We note that a basis of  $(X \rightarrow D)$  can be obtained by taking single-step functions constructed from the basis elements  $b \in B$  and basic open subsets  $O \subseteq X$ . We have  $b\chi_O \ll f$  iff  $O \ll f^{-1}(\uparrow b)$  [33, Proposition II-4.20(iv)], which characterises the way-below relation in  $(X \rightarrow D)$ . The fundamental function space we will deal with in this article consists of Scott continuous functions of type  $f : U \subseteq \mathbb{R}^n \rightarrow \mathcal{CR}^n$ , where  $U$  is an open subset.

Finally, consider a real Banach space  $X$  and recall that the dual  $X^*$  of  $X$  is defined as the set  $L(X, \mathbb{R})$  of bounded real-valued linear functionals on  $X$ . The weak\* topology on  $L(X, \mathbb{R})$  is the weakest topology on  $L(X, \mathbb{R})$  that makes all functionals  $\hat{x} : L(X, \mathbb{R}) \rightarrow \mathbb{R}$ , where  $x \in X$  with  $\hat{x}(f) = f(x)$ , continuous. This topology is Hausdorff and by Banach-Alaoglu's theorem, the closed ball  $X_c^* := B_c(0)$  of radius  $c$  centred at the origin is compact with respect to the weak\* topology for any  $c > 0$ . Let  $\mathcal{C}(X_c^*)$  denote the set of non-empty weak\* compact and convex subsets of  $X_c^*$ , augmented with  $X^*$ , with partial ordering induced by reverse inclusion. Then,  $\mathcal{C}(X_c^*)$  is a bounded complete domain in which  $C_1 \ll C_2$  iff  $C_2^\circ \subseteq C_1$ , where the interior is with respect to the relative weak\* subspace topology on  $X_c^*$ . Given  $x, y \in X$ , the closed line segment between them is denoted by  $s(x, y) := \{\lambda x + (1 - \lambda)y : 0 \leq \lambda \leq 1\}$ .

### 3 LOWER EXTENSION, LOWER ENVELOPE AND LOWER LIMIT OF DOMAIN MAPS

In any bounded complete dcpo  $D$ , in particular in any bounded complete domain, any non-empty set  $A \subseteq D$  has an infimum given by  $\inf A = \sup\{x : \forall a \in A. x \sqsubseteq a\}$  as the latter set is directed. In addition, in such a dcpo any net  $(x_j)_{j \in J}$  has a limit inferior or liminf defined by Gierz et al. [33, p. 133]:

$$\liminf_j x_j = \sup_j \inf_{i \geq j} x_i.$$

This can be used to define the notion of the lower limit of a map as well as its lower envelope, which is in fact the construction given in the work of Gierz et al. [33, Exercise II-3.19]. Recall that an isolated point of a topological space is a point  $x$  for which  $\{x\}$  is open.

*Definition 3.1.* Let  $f : X \rightarrow Z$  where  $X$  is a dense subset of a topological space  $Y$  and  $Z$  is a bounded complete domain. The *lower envelope* (cf. [33, Exercise II-3.19]) of  $f$  is given by the map  $f^* : Y \rightarrow Z$  with

$$f^*(y) = \sup\{\inf f[O \cap X] : O \text{ open set with } y \in O\}.$$

A continuous map  $g : Y \rightarrow Z$  is a *lower extension* of  $f$  if  $g \sqsubseteq f$  on  $X$ . If  $Y$  has no isolated points, then the *lower limit* of  $f$  is given by the map  $f^\dagger := \liminf f : Y \rightarrow Z$  with

$$f^\dagger(y) = \liminf f(y) = \sup\{\inf f[(O \setminus \{y\}) \cap X] : O \text{ open set with } y \in O\}.$$

Note that the lower limit or  $\liminf$  and the lower envelope of a real-valued function on Euclidean spaces are well-known constructions; see, for example, the work of Yeh [49, p. 144–145].

**PROPOSITION 3.2.** *Let  $f : X \rightarrow Z$ , where  $X$  is a dense subset of a topological space  $Y$  and  $Z$  is a bounded complete domain:*

- (i) *The map  $f^*$  is a lower extension of  $f$  with  $g \sqsubseteq f^*$  for any lower extension  $g$  of  $f$ .*
- (ii) *If  $f$  is continuous at  $x \in X$ , then  $f(x) = f^*(x)$ .*
- (iii) *If  $Y$  has no isolated points, then  $f^* \sqsubseteq f^\dagger$ . If  $Y$  is also  $T_1$ , in particular Hausdorff, then  $f^\dagger$  is continuous.*

**PROOF.** (i) From the definition, it follows directly that  $f^* \sqsubseteq f$ . Let  $y_0 \in Y$  and  $u \ll f^*(y_0)$ . By the interpolation property, take  $v \in Z$  with  $u \ll v \ll f^*(y_0)$ . Then, by the definition of  $f^*$ , there exists an open set  $O \subseteq Y$  with  $y_0 \in O$  such that  $v \sqsubseteq \inf f[O \cap X]$ . Thus, for all  $y \in O$ , we have  $u \ll f^*(y)$ , which shows that  $f^*$  is continuous. To show that  $g(y) \sqsubseteq f^*(y)$  for any  $y \in Y$ , let  $u \ll g(y)$ . Then by the continuity of  $g$  at  $y$ , there exists an open set  $O \subseteq Y$  with  $y \in O$  such that  $u \sqsubseteq \inf g[O]$  and thus  $u \sqsubseteq f^*(y)$ .

(ii) Let  $u \ll f(x)$ . By continuity, there exists open set  $O \subseteq Y$  such that  $y \in O \cap X$  implies  $u \ll f(y)$  and thus  $u \sqsubseteq \inf f[O \cap X] \sqsubseteq f^*(x)$ . Since  $u \ll f(x)$  is arbitrary, it follows that  $f(x) \sqsubseteq f^*(x)$  and thus  $f(x) = f^*(x)$  by (i).

(iii) The relation  $f^* \sqsubseteq f^\dagger$  follows from the definitions since for any open set  $O$  with  $y \in O$ , we have  $\inf f[O \cap X] \sqsubseteq \inf f[(O \setminus \{y\}) \cap X]$ , where the latter infimum exists because, by assumption,  $Y$  has no isolated points. Suppose  $Y$  is in addition  $T_1$ . Let  $y_0 \in Y$  and  $u \ll v \ll f^\dagger(y_0)$ . Then there exists an open subset  $O_0 \subseteq Y$  with  $y_0 \in O_0$  such that  $v \sqsubseteq \inf f[(O_0 \setminus \{y_0\}) \cap X]$ . We claim that  $y \in O_0$  implies  $u \ll f^\dagger(y)$ . Let  $y_0 \neq y \in O_0$ . By the separation property, there exists an open set  $O \subseteq Y$  with  $y \in O$  and  $y_0 \notin O$ . Therefore,  $O \cap O_0 \subseteq O_0$  and thus  $u \ll \inf f[(O_0 \setminus \{y\}) \cap X] \sqsubseteq \inf f[(O \cap O_0) \setminus \{y\}) \cap X] \sqsubseteq f^\dagger(y)$ , which proves the continuity of  $f^\dagger$  at  $y_0$ .  $\square$

The properties of the lower envelope then give us the following.

**COROLLARY 3.3.** *Suppose  $Z$  and  $U$  are bounded complete domains with  $f : X \rightarrow Z$  where  $X$  is a dense subset of  $Y$  and  $g : V \rightarrow U$ , where  $V$  is a dense subset of  $Z$  with  $f[X] \subseteq V$ . Then  $g^* \circ f^* \sqsubseteq (g \circ f)^*$  in the partial ordering of the function space  $Y \rightarrow U$ .*

**PROOF.** The map  $g^* \circ f^*$  is continuous as it is the composition of two continuous functions. Moreover,  $f^* \sqsubseteq f$  and  $g^* \sqsubseteq g$  and thus  $g^* \circ f^* \sqsubseteq g \circ f$ . Hence,  $g^* \circ f^*$  is a lower extension of  $g \circ f$ , and the result follows from Proposition 3.2(i).  $\square$

We can also deduce the following property of the lower limit and lower envelope.



**PROPOSITION 3.4.** *Suppose  $X$  is a dense subset of  $Y$  and  $f : X \rightarrow Z$  a map where  $Z$  is a bounded complete domain. If  $U$  is a bounded complete domain and  $g : Z \rightarrow U$  is Scott continuous and preserves non-empty infima, then  $(g \circ f)^* = g \circ f^*$  and, if  $Y$  has no isolated points,  $(g \circ f)^\dagger = g \circ f^\dagger$ .*

**PROOF.** We give the proof for the lower envelope as the proof for the lower limit is similar. For  $y \in Y$  and any open set  $O \subseteq Y$  with  $y \in O$ , we have  $\inf g[f[O \cap X]] = g(\inf f[O \cap X])$  since  $g$  preserves non-empty infima. Thus, by Scott continuity of  $g$ , we obtain

$$(g \circ f)^*(y) = \sup_{y \in O} \inf g[f[O \cap X]] = \sup_{y \in O} g(\inf f[O \cap X]) = g(\sup_{y \in O} \inf f[O \cap X]) = g \circ f^*(y).$$

□

There are many examples of lower limits and lower envelopes of maps with sets of discontinuities of various cardinality in analysis. We start with examples related to the domain of non-empty compact intervals  $\mathbb{IR}$  of  $\mathbb{R}$  partially ordered with reverse inclusion. The first shows that the lower envelope and the lower limit can be different, and that the lower limit need not be a lower extension.

**Example 3.5.** Consider the step function  $S : \mathbb{R} \rightarrow \mathbb{IR}$  defined by  $S(x) = 0$  if  $x < 0$ ,  $S(x) = 1$  if  $x > 0$  and  $S(0) = \perp$ . Then,  $S^\dagger(x) = S^*(x)$  for  $x \neq 0$  but  $S^\dagger(0) = [0, 1] \neq \perp = S^*(0)$ .

**Example 3.6.** Consider the periodic sawtooth wave  $S : \mathbb{R} \rightarrow \mathbb{R} \subseteq \mathbb{IR}$  defined by  $S(x) = x - \lfloor x \rfloor$ . Clearly,  $S$  has a discontinuity at each  $n \in \mathbb{Z}$ . We have  $S^\dagger : \mathbb{R} \rightarrow \mathbb{I}[0, 1]$  is given by  $f^\dagger(x) = x - \lfloor x \rfloor$  if  $x \notin \mathbb{Z}$  and  $S^\dagger(x) = [0, 1]$  for  $x \in \mathbb{Z}$ . Moreover,  $S^* = S^\dagger$ .

**Example 3.7.** Consider any function  $f : [0, 1] \rightarrow \mathbb{R}$  that is bounded (i.e., from below and above). Note that  $[0, 1] \subseteq \mathbb{I}[0, 1]$  is dense with respect to the Scott topology: in fact, given  $x \in [0, 1]$ , any open interval  $(a, b)$  containing  $x$  generates the Scott (equivalently upper) open set  $\{C \in \mathbb{I}[0, 1] : C \subseteq (a, b)\}$  which contains  $\{x\}$ . It follows that  $f^\dagger, f^* : \mathbb{I}[0, 1] \rightarrow \mathbb{IR}$ . It is easily seen that  $f^\dagger([a, b]) = f^*([a, b]) = \overline{\text{co}}\{f(x) : x \in [a, b]\}$ —that is, the smallest compact interval that contains  $f(x)$  for all  $x \in [a, b]$ .

**Example 3.8.** More generally, consider any bounded function  $f : [0, 1]^n \rightarrow \mathbb{R}^m \subseteq \mathbb{C}\mathbb{R}^m$ —that is, there exists  $k \geq 0$  such that  $\|f(x)\| \leq k$  for all  $x \in [0, 1]^n$ . Since  $[0, 1]^n$  is dense in  $\mathbb{I}[0, 1]^n$ , the lower limit and lower envelope  $f^\dagger, f^* : \mathbb{I}[0, 1]^n \rightarrow \mathbb{C}\mathbb{R}^m$  exist and  $f^\dagger(A) = f^*(A) = \overline{\text{co}}\{f(x) : x \in A\}$ —that is, the smallest convex and compact set containing  $f[A]$ .

**Example 3.9.** We now give an example, presented in [25, Example 6.6], of a map  $f : [-1, 1] \rightarrow [0, 1] \subseteq \mathbb{I}[0, 1]$  whose set of discontinuities is uncountable with positive Lebesgue measure  $\mu$ . Take any positive real number  $a < 1$ . The construction in the work of Edalat and Lieutier [25] gives two disjoint open subsets  $B, C \subseteq [-1, 1]$  with  $\mu(B) = \mu(C) = a$  and  $B = (C^c)^\circ$ —that is, the interior of the complement  $C^c$  of  $C$ . Define  $f : [-1, 1] \rightarrow \mathbb{I}[0, 1]$  with  $f(x) = 0$  if  $x \in \overline{B}$  and  $f(x) = 1$  if  $x \in C$ . Then  $D := [-1, 1] \setminus (B \cup C)$  is the set of discontinuities of  $f$  and has Lebesgue measure  $\mu(D) = 2 - 2a$ . The lower limit and the lower envelope  $f^\dagger, f^* : [-1, 1] \rightarrow \mathbb{I}[0, 1]$  coincide and  $f^\dagger(x) = 0$  if  $x \in B$ ,  $f^\dagger(x) = 1$  if  $x \in C$  and  $f^\dagger(x) = [0, 1]$  if  $x \in D$ .

Furthermore, we note the following well-known fact.

**LEMMA 3.10** ([35, p. 32]). *If  $A \subseteq \mathbb{R}^n$  is a bounded set, then*

$$\bigcap \{C : A \subseteq C \text{ convex and compact}\} = \overline{\text{co}}\{A\} = \text{co}\{\overline{A}\}.$$

Thus, if  $A \subseteq \mathbb{R}^n$  is a bounded set, then  $\inf e[A] \in \mathbb{C}\mathbb{R}^n$ , where  $e : \mathbb{R}^n \rightarrow \mathbb{C}\mathbb{R}^n$  is the embedding  $e(x) = \{x\}$ , and it can be computed as  $\inf e[A] = \bigcap \{C \in \mathbb{C}\mathbb{R}^n : A \subseteq C\} = \overline{\text{co}}\{A\} = \text{co}\{\overline{A}\}$ . We will also need the following result later on.

**PROPOSITION 3.11.** *Suppose  $X$  is a dense subset of a topological space  $Y$  without any isolated points and  $g : X \rightarrow \mathbb{R}^m \subset \mathbb{C}\mathbb{R}^m$  is continuous. Then, the lower limit and lower envelope of  $g$  coincide:  $(g)^\dagger = (g)^*$  as maps of type  $Y \rightarrow \mathbb{C}\mathbb{R}^m$ .*

**PROOF.** Given  $y \in O$ , for any open set  $O \subseteq Y$ , we note that whether or not  $y \in X$  we have  $\overline{g[O \cap X]} = \overline{g[(O \setminus \{y\}) \cap X]}$  since  $g$  is continuous. Thus, we have

$$\begin{aligned} \inf g[O \cap X] &= \bigcap \{C \in \mathbb{C}\mathbb{R}^m : g[O \cap X] \subseteq C\} \\ &= \text{co}\{\overline{g[O \cap X]}\} && \text{Lemma 3.10} \\ &= \text{co}\{\overline{g[(O \setminus \{y\}) \cap X]}\} \\ &= \bigcap \{C \in \mathbb{C}\mathbb{R}^m : g[(O \setminus \{y\}) \cap X] \subseteq C\} && \text{Lemma 3.10} \\ &= \inf g[(O \setminus \{y\}) \cap X]. \end{aligned}$$

The result now follows from the definitions of the lower envelope and lower limit.  $\square$

*Example 3.12.* Let  $f : (a, b) \rightarrow \mathbb{R}$  be a differentiable function. Then  $g = f' : (a, b) \rightarrow \mathbb{R}$  is continuous on a dense set but the set of discontinuities of  $f'$  can be dense, of positive Lebesgue measure or of full Lebesgue measure [8, Chapter 1.3.2, Proposition, 1.10, p. 30]. We have  $(f')^\dagger = (f')^* : (a, b) \rightarrow \mathbb{R}$  by Proposition 3.11.

#### 4 GENERALISED JACOBIAN AS LOWER LIMIT OF DERIVATIVES

The notions of lower limit and lower envelope can be applied to non-smooth analysis and optimisation. We recall the notion of generalised Jacobians of real locally Lipschitz vector functions, as introduced by Clarke and presented in that work [11, section 2.6]. First, note that by Rademacher's theorem [12, page 148], a locally Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable almost everywhere with respect to the Lebesgue measure.

Let  $\Omega_f$  be the null set where the locally Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  fails to be differentiable, and let  $S \subset U$  be any null set with respect to the Lebesgue measure. The generalised Jacobian  $\partial_S f : U \rightarrow \mathbb{C}\mathbb{R}^n$  for  $x \in U$  is defined to be

$$\partial_S f(x) = \text{co} \left\{ \lim_{j \rightarrow \infty} f'(x_j) : x_j \rightarrow x, \quad x_j \notin \Omega_f \cup S \right\}, \quad (2)$$

where  $f'(x)$  denotes the Fréchet derivative (Jacobian) of  $f$  at  $x \in U \setminus \Omega_f$  with respect to the standard Cartesian coordinates. The right-hand side of the preceding formula is to be interpreted as follows. There are many sequences  $(x_j)$  on  $U \setminus (\Omega_f \cup S)$  that converge to  $x$  such that  $f'(x_j)$  also converges to a limit; the generalised Jacobian  $\partial_S f(x)$  is the convex hull of all such limits. The first property to note is that the resulting set in Equation (2) does not depend on  $S$ .

**THEOREM 4.1** ([46, THEOREM 4]). *The set  $\partial_S f(x)$  is independent of the null set  $S$ .*

We will thus write  $\partial f(x) := \partial_S f(x)$  for any null set  $S$ . Let the vector space  $\mathbb{R}^{m \times n}$  of  $m \times n$  matrices over real numbers be equipped with the Frobenius norm.

**THEOREM 4.2** ([11, PROPOSITION 2.6.2]). *If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a locally Lipschitz map, then  $\partial f(x)$  is a non-empty convex compact subset of  $\mathbb{R}^{m \times n}$  for each  $x \in U$ , and the map  $\partial f : U \rightarrow \mathbb{C}\mathbb{R}^{m \times n}$  is upper semi-continuous.*

We have  $\partial f(x) \subseteq \partial f_1(x) \times \cdots \times \partial f_m(x)$ , where the latter denotes the set of  $m \times n$  matrices whose  $j$ th row belong to  $\partial_j f(x)$ . For a locally Lipschitz  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  (i.e.,  $m = 1$ ), the generalised Jacobian is called the *subgradient*  $\partial f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{C}\mathbb{R}^n$  which is equivalently defined by the support function  $\sigma_{\partial f(x)} : \mathbb{R}^n \rightarrow \mathbb{R}$  of  $\partial f(x)$  for  $x \in U$  (see Clarke [11, Proposition 2.1.2(b)] and

Theorem 2.5.1]):

$$\sigma_{\partial f(x)}(v) = \sup\{\langle y, v \rangle : y \in \partial f(x)\} = f^\circ(x; v) := \limsup_{y \rightarrow x, t \rightarrow 0^+} \frac{f(y + tv) - f(y)}{t}. \quad (3)$$

Here,  $f^\circ(x; v)$  is called the *generalised directional derivative* of  $f$  at  $x$  in the direction of  $v \in \mathbb{R}^n$ . We will use each of the four equivalent terms in Equation (3) as convenient in this work.

#### 4.1 Representation by Lower Limit

We will now show that for any locally Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the generalised Jacobian  $\partial f$  is the lower limit, equivalently the lower envelope, of the derivative map  $f' : U_0 \rightarrow \mathbb{R}^{m \times n}$ , where  $U_0 \subseteq U$  is the dense subset where  $f$  is differentiable. This will provide a new representation for the generalised Jacobian and for the subgradient when  $m = 1$ . Note that  $\mathbb{R}^{m \times n}$  can be considered as a subset of the maximal elements of the bounded complete domain  $\mathcal{C}\mathbb{R}^{m \times n}$ —that is, the set of non-empty compact and convex subsets of the space of  $m \times n$  real matrices  $\mathbb{R}^{m \times n}$  ordered by reverse subset inclusion and augmented with a bottom element that can be regarded as the whole space  $\mathbb{R}^{m \times n}$ . Thus, we can consider  $f'$  as a map of type  $U_0 \rightarrow \mathcal{C}\mathbb{R}^{m \times n}$ . We need the following theorem of Carathéodory on convex hulls.

**THEOREM 4.3 ([15]).** *Any point of the convex hull of a subset  $S \subseteq \mathbb{R}^p$  lies in the convex hull of at most  $p + 1$  points in  $S$ .*

By allowing some points in Carathéodory's theorem to be the same points if necessary, we can assume that any point of the convex hull of a subset  $S \subseteq \mathbb{R}^p$  lies in the convex hull of  $p + 1$  points in  $S$ .

We now have our main result in this section.

**THEOREM 4.4.** *For any locally Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the generalised Jacobian coincides with the lower envelope and the lower limit of the derivative map:*

$$\partial f = (f')^\dagger = (f')^*.$$

**PROOF.** Let  $x \in U$ . By Proposition 3.2(iii), we already know that  $(f')^*(x) \supseteq (f')^\dagger(x)$ . We will now show that  $\partial f(x) \subseteq (f')^\dagger(x)$ . By Theorem 4.1, we have  $\partial_S f(x) = \partial f(x)$  where  $S = \{x\}$ ; hence, it is sufficient to prove that  $\partial_{\{x\}} f(x) \subseteq (f')^\dagger(x)$ . Suppose we have a sequence  $x_n \in U_0 \setminus \{x\}$ ,  $n \geq 0$ , with  $\lim_{n \rightarrow \infty} x_n = x$ , such that  $y = \lim_{n \rightarrow \infty} f'(x_n)$  exists and thus  $y \in \partial_{\{x\}} f(x)$ . Let  $O \subseteq U$  be an open set with  $x \in O$  on which  $f$  is Lipschitz. Then there exists  $N \geq 0$  such that  $n \geq N$  implies  $x_n \in O \setminus \{x\}$ —that is,  $f'(x_n) \in f'[(O \setminus \{x\}) \cap U_0]$  for all  $n \geq N$ . If  $c \geq 0$  is a Lipschitz constant for  $f$  in  $O$ , then  $\|f'(u)\| \leq c$  for any  $u \in U_0 \cap O$  and thus  $f'[(O \setminus \{x\}) \cap U_0] \subseteq B_c(0)$ , where  $B_c(0)$  is the compact unit ball of radius  $c$  around the origin in  $\mathbb{R}^{m \times n}$ . Hence, by the comment after Lemma 3.10,  $\inf f'[(O \setminus \{x\}) \cap U_0] \in \mathcal{C}\mathbb{R}^{m \times n}$  is the convex hull of the closure of  $f'[(O \setminus \{x\}) \cap U_0]$ , which implies  $y \in \inf f'[(O \setminus \{x\}) \cap U_0]$ —that is,  $\inf f'[(O \setminus \{x\}) \cap U_0] \subseteq y$ . Since this holds for any sufficiently small open set  $O$  containing  $x$ , it follows that  $\sup\{\inf f'[(O \setminus \{x\}) \cap U_0] : x \in O, O \text{ open}\} \subseteq y$ —that is,  $y \in (f')^\dagger(x)$ . But  $\partial_{\{x\}} f(x)$  is the convex hull of points such as  $y$  and  $(f')^\dagger(x)$  is convex. Thus,  $\partial f(x) = \partial_{\{x\}} f(x) \subseteq (f')^\dagger(x) \subseteq (f')^*(x)$ .

Next, we will show that  $(f')^*(x) \subseteq \partial f(x)$ . Suppose  $y \in (f')^*(x)$ . For any integer  $k \geq 0$ , let  $O_k$  be the open ball of radius  $1/2^k$  centred at  $x$ . Since for  $k \geq 0$ ,  $y \in \inf f'[O_k \cap U_0] = \overline{\text{co}}\{f'[O_k \cap U_0]\}$ , it follows from Carathéodory's Theorem 4.3, applied to the point  $y$  of the  $mn$  dimensional Euclidean space  $\mathbb{R}^{m \times n}$ , that there exist  $mn + 1$  points  $y_{ik} \in f'[O_k \cap U_0]$ , for  $1 \leq i \leq mn + 1$ , such that  $y \in \text{co}\{y_{ik} : 1 \leq i \leq mn + 1\}$ . By the definition of the closure of a set, let  $x_{ik} \in O_k \cap U_0$  be such that  $\|f'(x_{ik}) - y_{ik}\| < 1/2^k$  for each  $k \geq 0$  and  $1 \leq i \leq mn + 1$ . Since the subset  $\inf f'[O_k \cap U_0]$  is compact

for each  $k \geq 0$ , there is a subsequence  $y_{ik_\ell}$  such that the limits  $y_i := \lim_{\ell \rightarrow \infty} y_{ik_\ell}$  exist for all  $i$  with  $1 \leq i \leq mn + 1$ . By continuity, we have  $y \in \text{co}\{y_i : 1 \leq i \leq mn + 1\}$ . In fact, the polyhedron  $P_\ell := \text{co}\{y_{ik_\ell} : 1 \leq i \leq mn + 1\}$  converges to  $P := \text{co}\{y_i : 1 \leq i \leq mn + 1\}$  in the Hausdorff metric  $d_{\mathcal{H}}$  on  $\mathcal{C}\mathbb{R}^{m \times n}$  which implies  $d(P, y) = \lim_{\ell \rightarrow \infty} d(P_\ell, y) = 0$  since  $d(P_\ell, y) = 0$  for each  $\ell \geq 0$ , where  $d(A, y)$  is the minimum distance from the point  $y$  to a compact set  $A$ . By construction, we have  $\lim_{\ell \rightarrow \infty} x_{ik_\ell} = x$  and  $\lim_{\ell \rightarrow \infty} f'(x_{ik_\ell}) = y_i$ . Thus,  $y_i \in \partial f(x)$  for  $1 \leq i \leq mn + 1$ , and by convexity,  $y \in \partial f(x)$ . Hence,  $(f')^*(x) \subseteq \partial f(x)$ , which completes the proof.  $\square$

Since the lower limit or the limit inferior of a map is more widely used in analysis, we will formulate our results in this article in terms of the lower limit.

The following simple property of continuous maps provides a useful tool in computation in different contexts, in particular for the extension of continuous maps, including the elementary functions of type  $\mathbb{R}^n \rightarrow \mathbb{R}$  to intervals  $\mathbb{IR}^n \rightarrow \mathbb{IR}$  which is at the basis of interval analysis [38].

**PROPOSITION 4.5.** *Let  $g : \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^m \subseteq \mathcal{C}\mathbb{R}^m$  be a continuous function, where  $k$  is a non-negative integer. Then  $g^\dagger = g^* : \mathcal{C}\mathbb{R}^n \times \mathcal{C}\mathbb{R}^k \rightarrow \mathcal{C}\mathbb{R}^m$  and  $g^\dagger((C_1, C_2)) = \text{co}\{g[C_1 \times C_2]\}$  for compact and convex subsets  $C_1 \in \mathcal{C}\mathbb{R}^n$  and  $C_2 \in \mathcal{C}\mathbb{R}^k$ .*

**PROOF.** By Proposition 3.11,  $g^\dagger = g^*$ . Their common value easily follows from the definition of the lower envelope since a continuous map preserves compact subsets.  $\square$

If  $v \in \mathbb{R}^n$ , then the map  $L_v : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^m \subseteq \mathcal{C}\mathbb{R}^m$  given by  $L_v(M) = Mv$  is linear and thus preserves compact and convex sets. By Proposition 4.5, we have  $L_v^* = L_v^\dagger : \mathcal{C}\mathbb{R}^{m \times n} \rightarrow \mathcal{C}\mathbb{R}^m$  given by  $L_v^\dagger(C) = \{Mv : M \in C\}$ . We observe that  $L_v^\dagger$  is Lipschitz on the non-bottom elements of  $\mathcal{C}\mathbb{R}^{m \times n}$  with respect to the Hausdorff metric: if  $C, D \in \mathcal{C}\mathbb{R}^{m \times n}$  and  $r > 0$ , then  $C \subseteq D + rB_1^{m \times n}$  implies  $L_v(C) = Cv \subseteq Dv + rB_1^{m \times n}v = L_v(D) + rB_1^{m \times n}v \subseteq L_v(D) + r\|v\|B_1^m$ , since for  $M \in B_1^{m \times n}$  we have  $\|Mv\| \leq \|M\|_F\|v\|$  as the Frobenius norm is subordinate to the Euclidean norm. From this relation and its symmetric counterpart, it follows easily that  $L_v^\dagger$ , applied to non-empty compact and convex subsets, has Lipschitz constant  $\|v\|$ .

**COROLLARY 4.6.** *If  $v \in \mathbb{R}^n$  and  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is Lipschitz, then  $(L_v \circ f')^\dagger = L_v^\dagger \circ \partial f$  (i.e.,  $(L_v \circ f')^\dagger(x) = \langle \partial f(x), v \rangle$ ) for all  $x \in U$ .*

**PROOF.** Since  $L_v^\dagger : \mathcal{C}\mathbb{R}^{m \times n} \setminus \{\perp\} \rightarrow \mathcal{C}\mathbb{R}^m$  is Lipschitz with respect to the Hausdorff metric, it is Scott continuous. It is also easy to check, using Theorem 4.3 that for any non-empty family  $M_i$ , where  $i \in I$ , with  $M_i \in \mathcal{C}\mathbb{R}^{m \times n}$  we have  $\text{co}\{\bigcup_{i \in I} M_i v\} = (\text{co}\{\bigcup_{i \in I} M_i\})(v)$ . Therefore,  $L_v^\dagger$  preserves non-empty infima and the result follows from Proposition 3.4 since we have  $L_v \circ f' = L_v^\dagger \circ f'$ .  $\square$

## 4.2 Basic Properties of Generalised Jacobian

A number of properties of the generalised Jacobians proved in the work of Clarke [11] now simply follow, in the light of its coincidence with the lower envelope given by Theorem 4.4, by the basic properties of the lower envelope in Proposition 3.2.

**COROLLARY 4.7.** *For any locally Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $x \in U$ , we have*

- (i) *The set  $\partial f(x)$  is non-empty, convex and compact.*
- (ii) *If  $f'(x)$  exists, then  $f'(x) \in \partial f(x)$ .*
- (iii) *If  $f$  is continuously differentiable at  $x$ , then  $\partial f(x) = f'(x)$ .*
- (iv)  *$\partial f : U \rightarrow \mathcal{C}\mathbb{R}^{m \times n}$  is upper continuous.*

*Remark 4.8.* We cannot directly use the standard extension of continuous maps as in the work of Gierz et al. [33, p. 181] to obtain Theorem 4.4. In fact, there are Lipschitz functions which are not continuously differentiable at any point, so Corollary 4.7(iii) does not apply. For example, in the work of Lebourg [37, Proposition 1.9], a Lipschitz map  $f : [0, 1] \rightarrow \mathbb{R}$  has been constructed with  $\partial f(x) = [0, 1]$  for all  $x \in [0, 1]$ . It follows that  $f$  is not continuously differentiable at any point  $x \in [0, 1]$  since at such a point we would have  $\partial f(x) = f'(x) \neq [0, 1]$ .

Next we show that the two chain rules for the generalised Jacobian derived in the work of Clarke [11, Theorem 2.6.6 and its corollary] can be deduced with a much shorter proof. We note that if  $B \in \mathcal{C}\mathbb{R}^{m \times n}$  and  $A \in \mathcal{C}\mathbb{R}^{k \times m}$ , then  $\text{co}\{AB\} := \text{co}\{PQ : Q \in B, P \in A\} \in \mathcal{C}\mathbb{R}^{k \times n}$ . In fact, it is easy to check that  $AB \subseteq \mathbb{R}^{k \times n}$  is compact, but it is, in general, not convex. As a counter example, with  $m = k = 2$  and  $n = 1$ , let  $B = \{\lambda(1, 0)^T + (1 - \lambda)(0, 1)^T : 0 \leq \lambda \leq 1\}$  and

$$A = \left\{ \mu \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (1 - \mu) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} : 0 \leq \mu \leq 1 \right\}.$$

Then,  $AB$  is not convex as  $(1, 0)^T, (-1, 0)^T \in AB$  but  $(0, 0)^T \notin AB$ . However, we have the following proposition.

**PROPOSITION 4.9.** *If  $B \subseteq \mathbb{R}^{m \times n}$  and  $A \subseteq \mathbb{R}^{k \times m}$ , then*

$$\text{co}\{AB\} = \text{co}\{A \text{co}\{B\}\} = \text{co}\{\text{co}\{A\}B\} = \text{co}\{\text{co}\{A\} \text{co}\{B\}\}.$$

**PROOF.** Since the second and third sets in the preceding equalities are contained between the first and the fourth, it is sufficient to show the equality of the latter two. Clearly,  $\text{co}\{AB\} \subseteq \text{co}\{\text{co}\{A\} \text{co}\{B\}\}$ . If  $P \in \text{co}\{A\}$  and  $Q \in \text{co}\{B\}$ , then by Carathéodory Theorem 4.3, there exists  $P_i \in A, \lambda_i \in [0, 1]$  for  $1 \leq i \leq km + 1$  with  $\sum_{i=1}^{km+1} \lambda_i = 1$  and  $Q_j \in B, \mu_j \in [0, 1]$  for  $1 \leq j \leq mn + 1$  with  $\sum_{j=1}^{mn+1} \mu_j = 1$  such that  $P = \sum_{i=1}^{km+1} \lambda_i P_i$  and  $Q = \sum_{j=1}^{mn+1} \mu_j Q_j$ . It follows that  $PQ = \sum_{i=1}^{km+1} \sum_{j=1}^{mn+1} \lambda_i \mu_j P_i Q_j$  with  $\sum_{i=1}^{km+1} \sum_{j=1}^{mn+1} \lambda_i \mu_j = 1$ . Thus,  $PQ \in \text{co}\{AB\}$  and the result follows.  $\square$

We now recall the mean value theorem for the generalised Jacobian.

**PROPOSITION 4.10** ([11, PROPOSITION 2.6.5]). *If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is Lipschitz in  $U$ , then for  $x, y \in U$  we have  $f(x) - f(y) \in \text{co}\{\partial f[s(x, y)]\}(x - y)$ .*

Note that in the preceding proposition, we have  $\partial f[s(x, y)] := \bigcup\{\partial f(z) : z = \lambda x + (1 - \lambda)y, 0 \leq \lambda \leq 1\}$ , where  $s(x, y)$  is the compact line segment between the points  $x, y \in \mathbb{R}^n$ . By Proposition 4.9, we have  $(\text{co}\{\partial f[s(x, y)]\})(x - y) = \text{co}\{(\partial f[s(x, y)])(x - y)\}$ . We will actually give an alternative proof of Proposition 4.10 later in Corollary 7.6.

Observe that the composition map  $J : \mathbb{R}^{k \times m} \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{k \times n} \subseteq \mathcal{C}\mathbb{R}^{k \times n}$  with  $J((P, Q)) = PQ$  is continuous with respect to the Frobenius norm equivalently the Euclidean topology. By Proposition 4.5,  $J^* = J^\dagger : \mathcal{C}\mathbb{R}^{k \times m} \times \mathcal{C}\mathbb{R}^{m \times n} \rightarrow \mathcal{C}\mathbb{R}^{k \times n}$  with  $J^\dagger((A, B)) = \text{co}\{AB\}$ .

**THEOREM 4.11.** (Cf. [11, Corollary p. 75]) *Suppose  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is Lipschitz near  $x \in U$  and  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$  is Lipschitz near  $f(x)$ . Then for any  $v \in \mathbb{R}^n$ , we have*

$$\partial(g \circ f)(x)v \subseteq \text{co}\{\partial g(f(x))\partial f(x)\}v.$$

**PROOF.** By Corollary 4.6, it follows that

$$(L_v \circ (g \circ f)')^\dagger(x) = L_v^\dagger \circ ((g \circ f)')^\dagger(x) = \partial(g \circ f)(x)v.$$



However, the map  $x \mapsto \text{co}\{\partial g(f(x))\partial f(x)\}v = L_v^\dagger J^\dagger((g^\dagger(f(x)), f^\dagger(x)))$  of type  $U \rightarrow \mathbb{C}\mathbb{R}^k$ , is the composition of upper continuous or continuous functions and is thus upper continuous. Since by Theorem 4.4  $(L_v \circ (g \circ f)')^\dagger = (L_v \circ (g \circ f)')^*$ , the result will follow by Proposition 3.2(i) if we show that the upper continuous map  $x \mapsto \text{co}\{\partial g(f(x))\partial f(x)\}v$ , of type  $U \rightarrow \mathbb{C}\mathbb{R}^k$ , is indeed a lower extension of the map  $x \mapsto (g \circ f)'(x)v$  of the same type. Assume that the Fréchet derivative  $(g \circ f)'(x) \in \mathbb{R}^{kn}$  exists for some  $x \in U$ . (Note that the existence of  $(g \circ f)'(x)$  does not imply the existence of  $f'(x)$  or  $g'(f(x))$ .) We will show that if  $R := (g \circ f)'(x)$ , then  $Rv \in \text{co}\{\partial g(f(x))\partial f(x)\}v$  which will complete the proof. Let  $\epsilon > 0$  be given. Put  $v = \|v\|\check{v}$  where  $\check{v}$  is the unit vector in the direction of  $v$ . By the definition of the Fréchet derivative, there exists  $\delta_0 > 0$  such that  $\|v\| < \delta_0$  implies

$$\frac{\|(g \circ f)(x+v) - (g \circ f)(x) - Rv\|}{\|v\|} < \epsilon/2. \quad (4)$$

Let  $m_1 = \sup\{\|P\| : P \in \partial g(f(x))\}$  and  $m_2 = \sup\{\|Q\| : Q \in \partial f(x)\}$  and put  $m = \max\{m_1, m_2\}$ . By the upper continuity of  $\partial g$  at  $f(x)$  and that of  $\partial f$  at  $x$  as well as the continuity of  $f$  at  $x$ , there exists  $\delta_1 > 0$  such that  $\|v\| < \delta_1$  implies

$$\partial g(f(x+v)) \subseteq (\partial g(f(x)))_{\epsilon_0} \quad \text{and} \quad \partial f(x+v) \subseteq (\partial f(x))_{\epsilon_0} \quad (5)$$

with  $\epsilon_0 = \epsilon/(4m+1)$ , where  $C_{\epsilon_0}$  is, the  $\epsilon_0$ -open neighbourhood of  $C$  as defined in Section 1.1.

By Proposition 4.9 and Proposition 4.10, for  $\|v\| < \delta_1$ , we have

$$\begin{aligned} g(f(x+v)) - g(f(x)) &\in \text{co}\{\partial g[f(x+v), f(x)](f(x+v) - f(x))\} \\ &\subseteq \text{co}\{\partial g[f(x+v), f(x)]\text{co}\{\partial f[x+v, x]\}\}v \\ &= \text{co}\{\text{co}\{\partial g[f(x+v), f(x)]\}\text{co}\{\partial f[x+v, x]\}\}v \\ &\subseteq \text{co}\{(\partial g(f(x)))_{\epsilon_0}(\partial f(x))_{\epsilon_0}\}v. \end{aligned}$$

Thus, for  $\|v\| < \delta_1$ , by Carathéodory Theorem 4.3, there exist  $P_i \in (\partial g(f(x)))_{\epsilon_0}$  and  $Q_i \in (\partial f(x))_{\epsilon_0}$  together with  $P'_i \in \partial g(f(x))$  and  $Q'_i \in \partial f(x)$ , for  $1 \leq i \leq t := kn+1$ , and  $c_i \geq 0$ , with  $\sum_{i=1}^t c_i = 1$ , such that  $g(f(x+v)) - g(f(x)) = \sum_{i=1}^t c_i P_i Q_i$ ,  $\|P_i - P'_i\| < \epsilon_0$  and  $\|Q_i - Q'_i\| < \epsilon_0$ . From Relation (4), for  $\|v\| < \delta_0$ , we have  $\|(\sum_{i=1}^t c_i P_i Q_i v) - Rv\| < \epsilon\|v\|/2$ . Therefore, for  $\|v\| < \min\{\delta_0, \delta_1\}$ , we obtain

$$\begin{aligned} \|((\sum_{i=1}^t c_i P'_i Q'_i) - R)v\| &= \|(\sum_{i=1}^t c_i (P'_i Q'_i - P'_i Q_i + P'_i Q_i - P_i Q_i + P_i Q_i) v - Rv\| \\ &\leq \sum_{i=1}^t c_i (\|P'_i\|\|Q'_i - Q_i\| + \|P_i - P'_i\|\|Q_i\|) + \|\sum_{i=1}^t c_i P_i Q_i - R\|\|v\| \\ &\leq m\epsilon_0\|v\| + m\epsilon_0\|v\| + \epsilon\|v\|/2 = (2m\epsilon\|v\|)/(4m+1) + \epsilon\|v\|/2 < \epsilon\|v\|. \end{aligned}$$

It follows that  $\|\sum_{i=1}^t c_i P'_i Q'_i \check{v} - R\check{v}\| < \epsilon$ . Since  $\epsilon > 0$  is arbitrary and  $\partial g(f(x))\partial f(x)\check{v}$  is compact, we obtain  $R\check{v} \in \partial g(f(x))\partial f(x)\check{v} \subseteq \text{co}\{\partial g(f(x))\partial f(x)\}\check{v}$ . The result follows.  $\square$

COROLLARY 4.12. (Cf. [11, 2.6.2]) For  $k = 1$ , we have

$$\partial(g \circ f)(x) \subseteq \text{co}\{\partial g(f(x))\partial f(x)\}.$$

PROOF. When  $k = 1$ , the two sets  $\partial(g \circ f)(x)$  and  $\text{co}\{\partial g(f(x))\partial f(x)\}$  are compact subsets of  $\mathbb{R}^n$ . The theorem now says that the support function of  $\partial(g \circ f)(x)$  is less than that of  $\text{co}\{\partial g(f(x))\partial f(x)\}$ , from which the result follows by the comment after Equation (1).  $\square$

## 5 SMOOTH APPROXIMATION OF SUBGRADIENT

Let  $U \subseteq \mathbb{R}^n$  be an open subset, and let  $\text{Lip}(U)$  denote the set of Lipschitz maps  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , and let  $\text{Lip}_0(U)$  denote the set of locally Lipschitz maps of type  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ —that is, for each  $x \in U$ , there exists an open set  $O \subseteq U$  such that  $x \in O$  and  $f$  is Lipschitz in  $O$ . For any function

$f : U \rightarrow \mathbb{R}$ , let  $Df(x) := f'(x)$  be the derivative of  $f$  at  $x$  when it exists. In particular, we have a map

$$D : C^1(U) \rightarrow C^0(U),$$

where  $C^1(U)$  is the set of continuously differentiable functions  $U \rightarrow \mathbb{R}$  equipped with the  $C^1$  norm topology, and  $C^0(U)$  is the space of continuous functions  $U \rightarrow \mathbb{R}$  equipped with the sup norm topology.

Recall that the L-topology, the weakest refinement of the sup norm topology such that the subgradient operator  $\partial : \text{Lip}_0(U) \rightarrow \mathcal{C}\mathbb{R}^n$  is continuous, is second countable and admits a complete metric [21]. We will from now on consider  $\text{Lip}_0(U)$  equipped with its L-topology.

In this section, we characterise the L-topology in terms of sequences of locally Lipschitz maps and we will show that  $C^1(U)$ , the subset of continuously differentiable functions in  $\text{Lip}(U)$ , is dense with respect to the L-topology. Note that if  $U$  is relatively compact (i.e., has a compact closure), then any map in  $\text{Lip}(U)$  extends by continuity to the closure of  $U$  and in this case we can use the closure  $\overline{U}$  of  $U$  and our results will imply that  $C^1(\overline{U}) \subseteq \text{Lip}(\overline{U})$  is a dense subset.

We start by noting the following property.

**LEMMA 5.1.** *Suppose  $h_k \in \text{Lip}_0(U)$ ,  $k \geq 0$ , is a sequence such that  $h_k \rightarrow f$  in the sup norm topology as  $k \rightarrow \infty$  where  $f \in \text{Lip}_0(U)$ . Then, for all  $x \in U$  and  $v \in S^{n-1}$ , we have*

$$\limsup_{y \rightarrow x, k \rightarrow \infty} h_k^\circ(y; v) \geq f^\circ(x; v).$$

**PROOF.** Let  $x \in U$  and  $v \in \mathbb{R}^n$ , and let  $N > 0$ ,  $\epsilon > 0$  and  $\delta > 0$  be given. By the definition of  $f^\circ(x; v)$  in Equation (3), there exist  $y$  and  $t$  with  $s(y, y + tv) \subset U$ ,  $|x - y| < \delta$  and  $0 < t < \delta$  such that  $(f(y + tv) - f(y))/t > f^\circ(x; v) - \epsilon$ . Since  $h_k \rightarrow f$  in the sup norm topology, there exists  $\ell > N$  such that  $(h_\ell(y + tv) - h_\ell(y))/t > f^\circ(x; v) - \epsilon$ . By the compactness of  $s(y, y + tv)$ , it follows that there exists an open set  $U_0 \subseteq U$  with  $s(y, y + tv) \subset U_0$  such that  $h_\ell$  is Lipschitz in  $U_0$ . By Lebourg's mean value theorem [11, 2.3.7] applied to  $h_\ell$  in the open line segment  $(y, y + tv)$ , there exists  $z \in (y, y + tv)$  such that  $h_\ell(y + tv) - h_\ell(y) \in \langle \partial h_\ell(z), tv \rangle$  (i.e.,  $h_\ell^\circ(z; v) > f^\circ(x; v) - \epsilon$ ), and the result follows.  $\square$

## 5.1 Characterisation of Convergence in L-Topology

Observe that since  $\partial f = f'$  for  $f \in C^1(U)$ , the restriction of the L-topology on  $C^1(U)$  is precisely the  $C^1$  norm topology.

**PROPOSITION 5.2.** *The relative subspace L-topology induced on the subset  $C^1(U) \subset \text{Lip}_0(U)$  coincides with the  $C^1$  norm topology.*

We will now obtain one of our main results, which gives a characterisation of the L-topology by classical notions in mathematical analysis. First, we need to fix our terminology. For a sequence of functions  $g_k : V \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , with  $k \geq 0$ , and a subset  $A \subseteq V$ , we say  $\limsup_{k \rightarrow \infty} g_k(v) = g(v)$  uniformly for  $v \in A$  if the sequence  $G_k(v) = \sup_{m \geq k} g_m(v)$  converges (in fact decreases) to  $g(v) = \inf_{k \geq 1} G_k(v)$  as  $k \rightarrow \infty$  uniformly for  $v \in A$ .

**LEMMA 5.3.** *Suppose the sequence of maps  $h_k \in \text{Lip}_0(U)$  converges to  $f \in \text{Lip}_0(U)$  in the sup norm topology. Then, we have*

$$\limsup_{y \rightarrow x, k \rightarrow \infty} h_k^\circ(y; v) = f^\circ(x; v), \tag{6}$$

for all  $x \in U$  and  $v \in S^{n-1}$ , with the convergence being uniform for  $v \in S^{n-1}$ , iff for all  $\epsilon > 0$ , there exists  $N \geq 0$  and  $\delta > 0$  such that  $h_k^\circ(y; v) < \epsilon + f^\circ(x; v)$  for  $\|y - x\| < \delta$ ,  $k \geq N$  and all  $v \in S^{n-1}$ .

PROOF. By Lemma 5.1,  $\limsup_{y \rightarrow x, k \rightarrow \infty} h_k^\circ(y; v) \geq f^\circ(x; v)$ . Next, we note that the limsup taken over two variables is converted in the usual way to limsup over a single variable. In fact,  $\limsup_{y \rightarrow x, k \rightarrow \infty} h_k^\circ(y; v) = \inf_{k \geq 1} H_k(x, v)$  with  $H_k(x, v) = \sup_{m \geq k, y \in O_k} h_m^\circ(y; v)$ , where  $O_k$  is the open ball centred at  $x$  and of radius  $2^{-k}$  with  $x$  removed. The result now follows.  $\square$

THEOREM 5.4. *A sequence of maps  $h_k \in \text{Lip}_0(U)$  converges to  $f \in \text{Lip}_0(U)$  in the L-topology, as  $k \rightarrow \infty$ , iff  $h_k \rightarrow f$  in the sup norm topology and for all  $x \in U$  and  $v \in \mathbb{R}^n$  we have*

$$\limsup_{y \rightarrow x, k \rightarrow \infty} h_k^\circ(y; v) = f^\circ(x; v), \quad (7)$$

with the convergence being uniform for  $v \in S^{n-1}$ .

PROOF. Suppose that the sequence  $h_k \in \text{Lip}_0(U)$  converges to  $f \in \text{Lip}_0(U)$  in the L-topology. Since the L-topology is a refinement of the  $C^0$  topology,  $h_k \rightarrow f$  in the sup norm topology. To invoke Lemma 5.3 to deduce the uniform convergence to the limit superior, let  $x \in U$  and  $\epsilon > 0$ . Then, by upper continuity of  $\partial f$  at  $x$ , there exists an open set  $O_0$  with  $x \in O_0$  such that  $\partial f(y) \subseteq (\partial f(x))_{\epsilon/2}$  for  $y \in O_0$ . Let  $C \in \mathcal{C}\mathbb{R}^n$  with  $(\partial f(x))_{\epsilon/2} \subseteq C \subseteq (\partial f(x))_\epsilon$  and let  $\delta > 0$  be such that open ball  $O$  of radius  $\delta$  and centre  $x$  satisfies  $\overline{O} \subseteq O_0$ . Then,  $O \ll O_0 \subseteq (\partial f)^{-1}(\uparrow C)$  and thus  $C\chi_O \ll \partial f$  [33, Proposition II-4.20(iv)]. Since the sequence  $h_k$  converges to  $f$  in the L-topology and since  $\uparrow C\chi_O$  is a Scott open neighbourhood of  $\partial f$ , it follows that there exists  $N$  such that for  $k \geq N$ , we have  $C\chi_O \ll \partial h_k$ , and hence  $\partial h_k(y) \subset C^\circ$  for  $y \in O$ . It follows that for all  $k \geq N$ ,  $y \in O$  and  $v \in S^{n-1}$ ,  $h_k^\circ(y; v) < \sigma_C(v) < f^\circ(x; v) + \epsilon$  which yields the required uniform convergence by Lemma 5.3.

Next, assume that  $f \in \text{Lip}_0(U)$  and that  $h_k \in \text{Lip}_0(U)$ ,  $k \geq 0$ , is a sequence with  $h_k \rightarrow f$  as  $k \rightarrow \infty$  in the sup norm topology and for all  $x \in U$ , the convergence in (7) is uniform in  $v \in S^{n-1}$ . Let  $C\chi_{O_0}$  be a single-step function with  $C\chi_{O_0} \ll \partial f$  in  $(U \rightarrow \mathbb{C}\mathbb{R}^n)$  and  $\overline{O_0} \subseteq U$  compact. From  $C\chi_{O_0} \ll \partial f$ , we obtain  $O_0 \ll (\partial f)^{-1}(\uparrow C)$ . By the interpolation property, there exists an open set  $O$  with  $O_0 \ll O \ll (\partial f)^{-1}(\uparrow C)$ , which implies that for all  $x \in \overline{O_0}$  we have  $C \ll \partial f(x)$ . Thus, for each  $x \in \overline{O_0}$ , there exists  $\epsilon_x > 0$  and an open neighbourhood  $O_x \subset O$  of  $x$  such that  $(\partial f(x))_{\epsilon_x} \subseteq C^\circ$ . In addition, using the uniform convergence in (7), by Lemma 5.3, there exists, for each  $x \in \overline{O_0}$ , an open neighbourhood  $O'_x$  of  $x$  and  $N_x \geq 0$  such that  $h_k^\circ(y; v) < f^\circ(x; v) + \epsilon_x$  for all  $k \geq N_x$ ,  $v \in S^{n-1}$  and  $y \in O'_x$ . Consider the open cover of the compact set  $\overline{O_0}$  by open sets  $U_x := O_x \cap O'_x$  with  $x \in \overline{O_0}$ . Let  $U_{x_i}$  for  $i = 1, \dots, \ell$  be a finite cover and put  $\epsilon = \min\{\epsilon_{x_i} : 1 \leq i \leq \ell\}$  and  $N = \max\{N_{x_i} : 1 \leq i \leq \ell\}$ . Then for all  $k \geq N$ ,  $y \in \overline{O_0}$  and  $v \in S^{n-1}$ , we have  $y \in U_{x_i} = O_{x_i} \cap O'_{x_i}$  for some  $i = 1, \dots, \ell$  and hence:  $h_k^\circ(y; v) < f^\circ(x_i; v) + \epsilon_{x_i} < \sup\langle C, v \rangle$ . Therefore,  $\partial h_k(y) \subset C^\circ$  for all  $y \in \overline{O_0}$  and  $k \geq N$ —that is,  $\overline{O_0} \subset \partial h_k^{-1}(\uparrow C)$ , for all  $k \geq N$ . Thus,  $O_0 \ll \partial h_k^{-1}(\uparrow C)$  (i.e.,  $C\chi_{O_0} \ll \partial h_k$  for all  $k \geq N$ ), which proves the convergence of the sequence  $h_k$  to  $f$  in the L-topology.  $\square$

COROLLARY 5.5. *A sequence of maps  $h_k \in C^1(U)$ ,  $k \geq 0$ , converges to a map  $f \in \text{Lip}_0(U)$  in the L-topology iff for all  $x \in U$  and  $v \in \mathbb{R}^n$  we have*

$$\limsup_{y \rightarrow x, k \rightarrow \infty} \langle h'_k(y), v \rangle = f^\circ(x; v), \quad (8)$$

with the convergence being uniform for  $v \in S^{n-1}$ .

Recall that a Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  is *regular* at  $x \in U$  if the one-sided directional derivative of  $f$  exists at  $x$  for all  $v \in \mathbb{R}^n$  and  $f'(x; v) = f^\circ(x; v)$ . We say  $f$  is a *regular map* if it is regular for all  $x \in U$  [11, 2.3.4].

COROLLARY 5.6. A sequence of maps  $h_k \in C^1(U)$ ,  $k \geq 0$ , converges to a regular map  $f \in \text{Lip}_0(U)$  in the  $L$ -topology iff for all  $x \in U$  and  $v \in \mathbb{R}^n$  we have

$$\limsup_{y \rightarrow x, k \rightarrow \infty} \langle h'_k(y), v \rangle = f'(x; v), \quad (9)$$

with the convergence being uniform for  $v \in S^{n-1}$ .

If  $1 \leq m \leq n$ , and  $x \in \mathbb{R}^n$ , we define  $x_{I_m} = (x_1, \dots, x_m) \in \mathbb{R}^m$  and  $x_{J_m} = (x_{m+1}, \dots, x_n) \in \mathbb{R}^{n-m}$  when  $1 \leq m < n$ . For a  $C^1$  map  $g : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , we define  $\frac{\partial g}{\partial x_{I_m}} = (\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_m})$ . Thus, when  $m = 1$ , we have  $\frac{\partial g}{\partial x_{I_1}} = \frac{\partial g}{\partial x_1}$  as usual and  $\frac{\partial g}{\partial x_{I_n}} = (\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_n}) = g'(x)$ . Consider now a Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  with  $1 \leq m < n$ . The partial subgradient of  $f$  with respect to the subspace  $\mathbb{R}^m \subseteq \mathbb{R}^n$ , denoted by  $\partial_{I_m} f(x)$ , is defined by restricting the vector  $v$  in Equation (3) to  $(v, 0) \in \mathbb{R}^m \times \mathbb{R}^{n-m}$ ; see the work of Clarke [11, p. 48].

COROLLARY 5.7. A sequence  $h_k \in C^1(U)$  converges to  $f \in \text{Lip}_0(U)$  in the  $L$ -topology, as  $k \rightarrow \infty$ , iff  $h_k \rightarrow f$  in the sup norm topology and for all  $x \in U$ ,  $1 \leq m \leq n$  and  $v \in \mathbb{R}^m$  we have

$$\limsup_{k \rightarrow \infty, y_{I_m} \rightarrow x_{I_m}} \left\langle \frac{\partial h_k(y)}{\partial y_{I_m}}, v \right\rangle = \sigma_{\partial_{I_m} f(x)}(v) \quad (10)$$

where the convergence is uniform for  $v \in S^{m-1}$ .

Let  $\pi_{I_m} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be the projection to  $\mathbb{R}^m$  with  $\pi_{I_m}(x_1, \dots, x_n) = (x_1, \dots, x_m)$ . Since for  $v \in \mathbb{R}^m$ ,

$$\limsup_{k \rightarrow \infty, y_{I_m} \rightarrow x_{I_m}} \left\langle \frac{\partial h_k(y_{I_m}, x_{J_m})}{\partial y_{I_m}}, v \right\rangle \leq \limsup_{k \rightarrow \infty, y_{I_n} \rightarrow x_{I_n}} \left\langle \frac{\partial h_k(y_{I_n})}{\partial y_{I_n}}, v \right\rangle = \limsup_{k \rightarrow \infty, y \rightarrow x} \langle h'_k(y), v \rangle,$$

from Corollary 5.7, we obtain an alternative short proof of a corresponding result in Clarke [11, Proposition 2.3.16].

COROLLARY 5.8. If  $f : U \subseteq \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$  is locally Lipschitz, then for all  $x \in U$ :

$$\partial_{I_m} f(x) \subseteq \pi_{I_m} [\partial f(x)].$$

Recall that, given any metric space  $(X, d)$ , the collection  $\text{Lip}(X, d)$  of bounded real-valued Lipschitz functions on  $X$  is equipped with its Lipschitz norm  $\| \cdot \|_{\text{Lip}}$  defined as

$$\|f\|_{\text{Lip}} = \|f\| + \|f\|_d \quad (11)$$

where  $\|f\| = \sup\{|f(x)| : x \in X\}$  is the sup norm and

$$\|f\|_d = \sup\{|f(x) - f(y)|/d(x, y) : x, y \in X, x \neq y\}.$$

If  $(X, d)$  is complete, then so is the Lipschitz norm [44]. We now present a simple example of a sequence of  $C^1$  functions that converges in the sup norm topology and in the  $L$ -topology, but not in the Lipschitz norm, to a Lipschitz map.

Example 5.9. Consider the sequence of  $C^1$  and Lipschitz functions  $h_k : [-1, 1] \rightarrow \mathbb{R}$ , for  $k \geq 1$ , with

$$h_k(x) = \begin{cases} |x| & \text{if } |x| \geq 1/k \\ \frac{kx^2}{2} + \frac{1}{2k} & \text{if } |x| < 1/k \end{cases}$$

and the Lipschitz map  $f : [-1, 1] \rightarrow \mathbb{R}$  with  $f(x) = |x|$ . Clearly,  $\lim_{k \rightarrow \infty} h_k = f$  in the sup norm topology. It is easily checked that

$$\|h_k - f\|_d \geq \left| \frac{(h_k(1/k) - f(1/k)) - (h_k(0) - f(0))}{(1/k)} \right| = \left| \frac{(1/k - 1/k) - ((1/2k) - 0)}{1/k} \right| = 1/2$$

for all  $k \geq 1$  and thus  $h_k$  does not tend to  $f$  as  $k \rightarrow \infty$  in the Lipschitz norm topology. However, for  $x, v \in \mathbb{R}$ ,

$$\limsup_{y \rightarrow x, k \rightarrow \infty} h'_k(y)v = |v| = \limsup_{y \rightarrow x, t \rightarrow 0^+} \frac{f(y + tv) - f(y)}{t},$$

and, trivially, for the two unit vectors  $v = 1$  and  $v = -1$  the convergence is uniform. Hence, by Theorem 5.5,  $h_k \rightarrow f$  in the L-topology. This therefore gives a simple application of our new results in basic mathematical analysis.

## 5.2 Construction of $C^\infty$ Approximations to a Lipschitz Map

Let  $\mathcal{G} : \mathbb{R}^n \rightarrow \mathbb{R}$  with

$$\mathcal{G}(x) = \frac{1}{(2\pi)^{n/2}} \exp(-(\|x\|^2/2)), \quad (12)$$

where  $x = (x_1, \dots, x_n)$ , be the standard multivariate Gaussian (normal) probability density distribution—that is, the product of  $n$  independent standard Gaussian distributions each along an axis of  $\mathbb{R}^n$ . For any positive integer  $k$ , let  $\mathcal{G}_k : \mathbb{R}^n \rightarrow \mathbb{R}$  be given by  $\mathcal{G}_k(x) = k^n \mathcal{G}(kx)$ . Then  $\mathcal{G}_k$  is the multivariate Gaussian probability density distribution with mean  $E(x_i) = 0$  and variance  $E(x_i^2) = 1/k^2$  for  $i = 1, \dots, n$ , and thus  $\int_{\mathbb{R}^n} \mathcal{G}_k(u) du = 1$ . Then  $\mathcal{G}_k$  is a sequence of test functions. For definiteness, we will use this particular sequence of test functions in deriving Theorem 5.10 presented in the following, but it is easy to see that this theorem follows for any sequence of test functions. For a Lipschitz map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , let  $h_k = f * \mathcal{G}_k : U \rightarrow \mathbb{R}$  be the convolution  $h_k(x) = \int_U f(y) \mathcal{G}_k(x - y) dy$  for any positive integer  $k$ .

**THEOREM 5.10.** *For any map  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  with Lipschitz constant  $c$ , the sequence  $h_k = f * \mathcal{G}_k$  is a sequence of  $C^\infty$  functions with Lipschitz constant  $c$  such that  $\lim_{k \rightarrow \infty} h_k = f$  in the L-topology.*

**PROOF.** Since  $\mathcal{G}_k$  is  $C^\infty$ , it follows from the derivative properties of convolutions [10, p. 119] that all partial derivatives

$$\frac{\partial h_k}{\partial x_i} = \frac{\partial(f * \mathcal{G}_k)}{\partial x_i} = f * \frac{\partial \mathcal{G}_k}{\partial x_i}$$

for  $i = 1, \dots, n$  exist as do all higher-order partial derivatives and therefore  $h_k$  is  $C^\infty$ . Let  $u = y - x$ , for  $x, y \in U$ , so that  $y = u + x$ . Then  $u \in U - U$ , where for subsets  $X, Y \subseteq \mathbb{R}^n$ , recall that the Minkowski sum and difference are defined by  $X \pm Y = \{x \pm y : x \in X, y \in Y\}$ . Thus, by the preceding change of variable in the integral, we obtain

$$h_k(x) = \int_{U-U} f(u + x) \mathcal{G}_k(u) du. \quad (13)$$

Then, we have

$$|h_k(x_1) - h_k(x_2)| \leq \int_{U-U} |f(u + x_1) - f(u + x_2)| \mathcal{G}_k(u) du \leq c \|x_1 - x_2\| \int_{U-U} \mathcal{G}_k(u) du \leq c \|x_1 - x_2\|.$$

Therefore,  $h_k$  has Lipschitz constant  $c$ , and we have

$$|h_k(x) - f(x)| \leq \int_{U-U} |f(u + x) - f(x)| \mathcal{G}_k(u) du \leq c \int_{U-U} \|u\| \mathcal{G}_k(u) du \leq \frac{c}{k} \int_{\mathbb{R}^n} \|u\| \mathcal{G}(u) du, \quad (14)$$

showing that  $\lim_{k \rightarrow \infty} h_k = f$  in the sup norm topology. By Theorem (5.5), it remains to show that for  $v \in \mathbb{R}^n$ :  $\limsup_{y \rightarrow x, k \rightarrow \infty} \langle h'_k(y), v \rangle = f^\circ(x; v)$ , where the convergence is uniform for  $v \in S^{n-1}$ . To use Lemma 5.3 to deduce the uniform convergence, let  $x \in U$ ,  $v \in S^{n-1}$  and  $\epsilon > 0$ . Since the map  $v \mapsto f^\circ(x; v) = \sigma_{\partial f(x)}(v)$  is uniformly continuous on the compact set  $S^{n-1}$ , there exists  $\delta_0 > 0$  such that for  $\|v - w\| < \delta_0$  with  $w \in S^{n-1}$ , we have  $|f^\circ(x; v) - f^\circ(x; w)| < \epsilon/6$ . Furthermore, since  $(x, v) \mapsto f^\circ(x; v)$  is upper continuous [11, Proposition 2.1.1(b)], there exist  $\delta_v, \alpha_v > 0$  such



that for  $\|y - x\| < \delta_v$  and  $0 < t < \alpha_v$ , we have  $(f(y + tv) - f(y))/t < f^\circ(x; v) + \epsilon/6$ . Thus, for  $\|v - w\| < \beta := \min\{\epsilon/6c, \delta_0\}$  with  $w \in S^{n-1}$ ,  $\|y - x\| < \delta_v$  and  $0 < t < \alpha_v$ , we have

$$\begin{aligned} \frac{f(y + tw) - f(y)}{t} &= \frac{f(y + tw) - f(y + tv) + f(y + tv) - f(y)}{t} < c\|v - w\| + \frac{f(y + tv) - f(y)}{t} \\ &< \frac{\epsilon}{6} + f^\circ(x; v) + \frac{\epsilon}{6} \leq \frac{\epsilon}{6} + f^\circ(x; w) + \frac{\epsilon}{6} + \frac{\epsilon}{6} = f^\circ(x; w) + \epsilon/2. \end{aligned} \quad (15)$$

Let  $O_r$  be the open ball of radius  $r > 0$  centred at the origin, and let  $N_v \geq 1$  be such that for  $k \geq N_v$ ,

$$c \int_{(U-U) \setminus O_{\delta_v/2}} \mathcal{G}_k(u) du < \frac{\epsilon}{2}. \quad (16)$$

For  $\|y - x\| < \delta_v/2$  and  $\|u\| < \delta_v/2$ , we have  $\|u + y - x\| < \|u\| + \|y - x\| \leq \delta_v/2 + \delta_v/2 = \delta_v$ . Thus, using Equation (13) to compute  $h_k(y)$  and  $h_k(y + tw)$ , applying Inequality (15), in which we replace  $y$  with  $y + u$ , and employing Inequality (16), we obtain for  $\|y - x\| < \delta_v/2$ ,  $0 < t < \alpha_v$ ,  $\|v - w\| < \beta$ ,  $w \in S^{n-1}$ , and  $k \geq N_v$ :

$$\begin{aligned} \frac{h_k(y + tw) - h_k(y)}{t} &= \int_{U-U} \left( \frac{f(u + y + tw) - f(u + y)}{t} \right) \mathcal{G}_k(u) du \\ &= \int_{(U-U) \cap O_{\delta_v/2}} \left( \frac{f(u + y + tw) - f(u + y)}{t} \right) \mathcal{G}_k(u) du \\ &\quad + \int_{(U-U) \setminus O_{\delta_v/2}} \left( \frac{f(u + y + tw) - f(u + y)}{t} \right) \mathcal{G}_k(u) du \\ &\leq (f^\circ(x; w) + \epsilon/2) \int_{(U-U) \cap O_{\delta_v/2}} \mathcal{G}_k(u) du + \int_{(U-U) \setminus O_{\delta_v/2}} \left| \frac{f(u + y + tw) - f(u + y)}{t} \right| \mathcal{G}_k(u) du \\ &\leq f^\circ(x; w) + \frac{\epsilon}{2} + \frac{\epsilon}{2} = f^\circ(x; w) + \epsilon. \end{aligned}$$

Hence, for  $k \geq N_v$ ,  $\|y - x\| < \delta_v/2$  and  $\|v - w\| < \beta$  with  $w \in S^{n-1}$ , we have  $\langle h'_k(y), w \rangle \leq f^\circ(x; w) + \epsilon$ . Now the open cover of the compact set  $S^{n-1}$  with balls of centre  $v \in S^{n-1}$  and radius  $\beta$  has a finite subcover with, say,  $v_i \in S^{n-1}$  for  $i = 1, \dots, \ell$ . We now put  $\delta = \min\{\delta_{v_i} : i = 1, \dots, \ell\}$  and also  $N = \max\{N_{v_i} : i = 1, \dots, \ell\}$ . Then, for  $k \geq N$  we obtain  $\langle h'_k(y), w \rangle \leq f^\circ(x; w) + \epsilon$  for all  $\|y - x\| < \delta$  and all  $w \in S^{n-1}$ , which completes the proof by Lemma 5.3.  $\square$

The following main result characterises the L-topology on Lipschitz maps in terms of the density of the subspace of  $C^1$  and  $C^\infty$  maps.

**COROLLARY 5.11.** *The subspace  $C^\infty(U) \cap \text{Lip}(U)$ , and thus  $C^1(U) \cap \text{Lip}(U)$ , is dense in  $\text{Lip}(U)$  with respect to the L-topology.*

Corollary 5.11 shows that the L-topology is the appropriate topology for Lipschitz maps when approximating these maps by sequences of  $C^1$  functions.

### 5.3 Subgradient Operator as Lower Limit of Derivative Operator

Now we are able to prove our final result in this section. Note that the differential operator  $D : C^1(U) \cap \text{Lip}(U) \rightarrow C^0(U)$  with  $D(f) = f'$  can be regarded as having type  $D : C^1(U) \cap \text{Lip}(U) \rightarrow (U \rightarrow \mathbb{C}\mathbb{R}^n)$  since  $C^0(U)$  can be identified as a subset of the maximal elements of  $(U \rightarrow \mathbb{C}\mathbb{R}^n)$ . For convenience, let  $F := C^1(U)$  in this section. As we have seen,  $F \cap \text{Lip}(U)$  is, by Corollary 5.11, dense in  $\text{Lip}(U)$ , and moreover, since the restriction of the L-topology on  $F \cap \text{Lip}(U)$  is the  $C^1$  norm topology,  $D$  is continuous on  $F \cap \text{Lip}(U)$ . In addition,  $\text{Lip}(U)$  equipped with the L-topology has clearly no isolated elements since for any  $f \in \text{Lip}(U)$  we have  $f + r \in \text{Lip}(U)$  for any  $r \in \mathbb{R}$ . Therefore, the lower limit  $D^\dagger : \text{Lip}(U) \rightarrow (U \rightarrow \mathbb{C}\mathbb{R}^n)$  exists. Next note that the L-topology on  $\text{Lip}(U)$ , being the meet of the sup norm topology and the Scott topology, is itself second countable. Thus, in the definition of  $D^\dagger$ , we can use a countable set of open sets. We will now find a simple expression for the support function of  $(D^\dagger f)(x)$  for  $x \in U$ .

LEMMA 5.12. *Let  $f \in \text{Lip}(U)$ ,  $v \in \mathbb{R}^n$  and  $x \in U$ . Then we have*

$$\sup\langle (D^\dagger f)(x), v \rangle = \limsup_{\substack{g \rightarrow f \\ g \in F \cap \text{Lip}(U)}} \langle g'(x), v \rangle,$$

where  $g \rightarrow f$  in the L-topology.

PROOF. Let  $(W_k)_{k \geq 0}$  be a local basis of the L-topology for  $f$ . If  $A_k$  for  $k \geq 0$  is a shrinking sequence of non-empty compact and convex sets, then by Equation (1), it is easy to see that the support function satisfies the following equality:

$$\sigma_{\bigcap_{k \geq 0} A_k} = \inf_{k \geq 0} \sigma_{A_k}. \quad (17)$$

We thus have the following derivation:

$$\begin{aligned} \sup\langle (D^\dagger f)(x), v \rangle &= \sup\langle \sup_{k \geq 0} \inf_{g \in (W_k \setminus \{f\}) \cap F} \langle g'(x), v \rangle, v \rangle && \text{definition of } D^\dagger \\ &= \sup\langle \bigcap_{k \geq 0} \inf_{g \in (W_k \setminus \{f\}) \cap F} \langle g'(x), v \rangle, v \rangle && \text{definition of sup in } \mathbb{C}\mathbb{R}^n \\ &= \inf_{k \geq 0} \sup_{g \in (W_k \setminus \{f\}) \cap F} \langle g'(x), v \rangle && \text{using Equation (17)} \\ &= \limsup_{g \rightarrow f} \langle g'(x), v \rangle && \text{definition of limit superior.} \end{aligned}$$

□

THEOREM 5.13. *The lower limit and the lower envelope of the differential operator of the type  $D : C^1(U) \cap \text{Lip}(U) \rightarrow (U \rightarrow \mathbb{C}\mathbb{R}^n)$  coincide with the subgradient operator:  $D^\dagger = D^* = \partial$ .*

PROOF. Let  $f \in \text{Lip}(U)$  with Lipschitz constant  $c$  and  $x \in U$ . Take a shrinking sequence  $W_k$  of open subsets  $W_0 \supseteq W_1 \supseteq W_2 \supseteq \dots$  with  $f \in W_k$  for all  $k \geq 0$  that form a local basis for the L-topology at  $f$ . Using the notation  $F := C^1(U)$ , if  $f \notin F$ , then we immediately have  $\inf_{g \in W_k \cap F} g' = \inf_{g \in (W_k \setminus \{f\}) \cap F} g'$ . Otherwise, if  $f \in F$ , then there is a sequence  $g_n \in (W_k \setminus \{f\}) \cap F$  with  $\lim_{n \rightarrow \infty} g_n = f$  in the L-topology, equivalently the  $C^1$  norm topology, and again  $\inf_{g \in W_k \cap F} g' = \inf_{g \in (W_k \setminus \{f\}) \cap F} g'$ . Hence,  $D^\dagger = D^*$ . We show that  $D^\dagger(f) = \partial f$ . Since  $D^\dagger(f)(x), \partial f(x) \in \mathbb{C}\mathbb{R}^n$ , it is sufficient to show that they have the same support function—that is, for each  $v \in \mathbb{R}^n$ ,

$$\sup\langle D^\dagger(f)(x), v \rangle = \sup\langle \partial f(x), v \rangle. \quad (18)$$

Let  $O_k \subset \mathbb{R}^n$  for  $k \geq 0$  be the open ball of radius  $2^{-k}$  centred at  $x$ . Assume without loss of generality that  $W_0$  contains maps with Lipschitz constant bounded by  $c + 1$  in  $O_0$ . We have  $D^\dagger(f) : U \rightarrow \mathbb{C}\mathbb{R}^n$  given by  $D^\dagger(f) = \sup_{k \geq 0} \inf_{g \in (W_k \setminus \{f\}) \cap F} g'$ . Let  $h_k \in W_k$  for  $k \geq 0$ . Then  $h_k \rightarrow f$

as  $k \rightarrow \infty$  in the L-topology. For  $g \in W_k \cap F$  and  $y \in O_k$ , by Cauchy-Schwarz inequality, obtain  $|\langle g'(y) - g'(x), v \rangle| \leq (c+1)2^{-k}\|v\|$ , which gives us

$$-(c+1)2^{-k}\|v\| < \sup_{\substack{g \in (W_k \setminus \{f\}) \cap F \\ y \in O_k \setminus \{x\}}} \langle g'(y), v \rangle - \sup_{g \in (W_k \setminus \{f\}) \cap F} \langle g'(x), v \rangle < (c+1)2^{-k}\|v\|.$$

Taking the limit as  $k \rightarrow \infty$  of the two decreasing sequences of supremums, we obtain

$$\inf_{k \geq 0} \sup_{\substack{g \in (W_k \setminus \{f\}) \cap F \\ y \in O_k \setminus \{x\}}} \langle g'(y), v \rangle = \inf_{k \geq 0} \sup_{g \in (W_k \setminus \{f\}) \cap F} \langle g'(x), v \rangle. \quad (19)$$

Then, we have

$$\begin{aligned} \sup \langle \partial f(x), v \rangle &= \limsup_{k \rightarrow \infty, y \rightarrow x} \langle h'_k(y), v \rangle && \text{by Theorem 5.5} \\ &= \inf_{k \geq 0} \sup_{\substack{g \in (W_k \setminus \{f\}) \cap F \\ y \in O_k \setminus \{x\}}} \langle g'(y), v \rangle && \text{since } h_k \rightarrow f \text{ in L-topology} \\ &= \inf_{k \geq 0} \sup_{g \in (W_k \setminus \{f\}) \cap F} \langle g'(x), v \rangle && \text{by Equation (19)} \\ &= \sup \langle D^\dagger f(x), v \rangle && \text{by Lemma 5.12.} \end{aligned}$$

□

If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  is Lipschitz and  $1 \leq m < n$ , then by using  $\frac{\partial}{\partial x_{I_m}}$  instead of  $D$  and  $\partial_{I_m} f$  instead of  $\partial f$  in Theorem 5.13, we obtain a similar result for the partial derivatives.

**COROLLARY 5.14.** *We have:  $(\frac{\partial}{\partial x_{I_m}})^\dagger = \partial_{I_m}$  i.e.,  $(\frac{\partial}{\partial x_{I_m}})^\dagger(f) = \partial_{I_m} f$ .*

## 6 EXTENSION OF FUNDAMENTAL THEOREM OF LINE INTEGRALS

We now use the results of Section 5 regarding the lower limit of the classical derivative operator to deduce a simple proof of the interval version of Green's theorem (i.e., the fundamental theorem of line integrals), which was obtained using interval valued integration in the work of Edalat et al. [27].

Let  $U \subseteq \mathbb{R}^n$  be an open set and  $p : [0, 1] \rightarrow U$  a  $C^1$  path in  $U$  from a given point  $p(0) = a$  to a point  $p(1) = b$ . If  $g : U \rightarrow \mathbb{R}$  is a  $C^1$  map, then the path integral  $\int_0^1 \langle g'(p(t)), p'(t) \rangle dt = g(p(1)) - g(p(0)) = g(a) - g(b)$  is independent of the path  $p$ . In particular, if  $a = b$ , then the path integral is always zero independent of the closed path  $p$ . We now define the operator

$$D_p : C^1(U) \cap \text{Lip}(U) \rightarrow ([0, 1] \rightarrow \mathbb{R})$$

by  $D_p(g) = \lambda t. \langle g'(p(t)), p'(t) \rangle$ , where  $\lambda t. h(t)$  stands for the function  $t \mapsto h(t)$  for any function  $h$ . Thus,  $D_p(g)$  gives the derivative of the composition  $g \circ p$  with respect to  $t \in [0, 1]$ . Since  $C^1(U) \cap \text{Lip}(U) \subseteq \text{Lip}(U)$  is dense with respect to the L-topology, we have its lower limit

$$D_p^\dagger : \text{Lip}(U) \rightarrow ([0, 1] \rightarrow \mathbb{R}), \quad D_p^\dagger(f)(t) = \lambda t. \sup_{k \geq 0} \inf_{g \in (W_k \setminus \{f\}) \cap C} \langle g'(p(t)), p'(t) \rangle, \quad (20)$$

where  $W_k$  is a shrinking sequence of open subsets for  $k \geq 0$  that form a local basis for the L-topology at  $f$ . Recall that any Scott continuous function  $f : [0, 1] \rightarrow \mathbb{R}$  is of the form  $f = [f^-, f^+]$ , where  $f^-$  and  $f^+$  are respectively lower and upper continuous function with  $f(x) = [f^-(x), f^+(x)]$ . In addition, if  $g = \sup_{i \in I} a_i \chi_{O_i}$  is a step function with  $I$  finite, then  $g \ll f$  iff  $O_i \ll f^{-1}(\uparrow a_i)$  for each  $i \in I$  [33, Proposition II-4-20].

**PROPOSITION 6.1.**

(i) *The function space  $([0, 1] \rightarrow \mathbb{R})$  is dense in  $[0, 1] \rightarrow \mathbb{R}$  with respect to the Scott topology.*

- (ii) *The lower limit of the integral operator  $\int : ([0, 1] \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$  with type  $\int^\dagger : ([0, 1] \rightarrow \mathbb{IR}) \rightarrow \mathbb{IR}$  is given by*

$$\int^\dagger(f) = \left[ \int f^- d\mu, \int f^+ d\mu \right],$$

where  $f = [f^-, f^+]$  and  $\mu$  is the Lebesgue measure on  $[0, 1]$ .

PROOF. (i) Let  $h = \sup_{i \in I} a_i \chi_{O_i} : [0, 1] \rightarrow \mathbb{IR}$ , be a step function, where  $a_i \in \mathbb{R}$  is compact,  $O_i$  is an open interval, and  $I$  is a finite indexing set with  $\uparrow h \neq \emptyset$ , which implies  $(h(x))^\circ \neq \emptyset$  for all  $x \in [0, 1]$ . Since  $h$  is a basic Scott open set in  $[0, 1] \rightarrow \mathbb{IR}$ , all we need to do is to construct a continuous function  $h_0 : [0, 1] \rightarrow \mathbb{R}$  with  $h \ll h_0$ . We have  $h = [h^-, h^+]$ , where  $h^-$  and  $h^+$  are, respectively, lower and upper semicontinuous, piecewise constant maps with  $h^-(x) < h^+(x)$  as real numbers if  $h(x) \neq \perp$ , and otherwise  $h^-(x) = -\infty$  and  $h^+(x) = +\infty$  when  $h(x) = \perp$ . The collection of open intervals  $O_i$ , for  $i \in I$ , induces a partition  $P : 0 = q_0 < q_1 < \dots, q_{t-1} < q_t = 1$  of  $[0, 1]$  into a finite number of (open, closed or half-open/half-closed) intervals on each of which  $h^-$  and  $h^+$  are constant. The Scott continuity of  $h$  at  $q_i$  for  $0 < i < t$  implies that  $h(x) \subseteq h(y)$  or  $h(y) \subseteq h(x)$  for  $x \in (q_{i-1}, q_i)$  and  $y \in (q_i, q_{i+1})$ . Let  $u_i, v_i \in \mathbb{IR}$  with  $v_i \subseteq u_i$  be the value of  $h(x)$  in the interiors of the two intervals in  $P$  with common boundary  $q_i$ . Then  $u_i$  is compact while  $v_i$  is either compact or  $v_i = \mathbb{R} = \perp$ . Put  $c_i = (u_i^- + u_i^+)/2$  for  $0 < i < t$ . If  $u_0 = h(0)$  is compact, put  $c_0 = (u_0^- + u_0^+)/2$ ; otherwise, let  $c_0 = 0$ . Similarly, if  $u_t = h(1)$  is compact, put  $c_t = (u_t^- + u_t^+)/2$ ; otherwise, let  $c_t = 0$ . Then, consider the piecewise linear map  $h_0 : [0, 1] \rightarrow \mathbb{R}$  with  $h_0(q_i) = c_i$ , linear in each interval  $[q_i, q_{i+1}]$  for  $0 \leq i \leq t-1$ . By construction, for each  $x \in [0, 1]$ , we have  $h(x) \ll h_0(x)$ . In fact, for each  $i \in I$ , there exists an open set  $O'_i$  with  $O_i \ll O'_i$  such that  $a_i \ll h_0(x)$  for  $x \in O'_i$ . It follows that  $O_i \ll h_0^{-1}(\uparrow a_i)$  for  $i \in I$  and hence  $h \ll h_0$ .

(ii) We first note that the integral  $\int : ([0, 1] \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$  is a continuous functional since if  $f_1, f_2 \in ([0, 1] \rightarrow \mathbb{R})$ , then  $|\int_0^1 f_1(x) - f_2(x) dx| \leq \int_0^1 |f_1(x) - f_2(x)| dx \leq \|f_1 - f_2\|_\infty$  where  $\|\cdot\|$  is the sup norm on  $([0, 1] \rightarrow \mathbb{R})$ . Thus, by the previous part and Proposition 3.11, we know that  $\int^\dagger$  and  $\int^*$  both exist and  $\int^\dagger = \int^*$ . Consider any Scott continuous function  $f : [0, 1] \rightarrow \mathbb{IR}$ . To compute  $\int^\dagger f$ , assume  $O \subseteq ([0, 1] \rightarrow \mathbb{IR})$  is a Scott open set with  $f \in O$ . Let  $f = [f^-, f^+]$ . Since  $f^-$  and  $f^+$  are respectively lower and upper semi-continuous functions, there exist an increasing sequence of continuous functions  $f_i^- : [0, 1] \rightarrow \mathbb{R}$  and a decreasing sequence of continuous functions  $f_i^+ : [0, 1] \rightarrow \mathbb{R}$ , where  $i \geq 0$ , such that  $\sup_{i \geq 0} f_i^- = f^-$  and  $\inf_{i \geq 0} f_i^+ = f^+$  [31, Section 1.7.15(c)]. Thus, the sequence  $f_i = [f_i^-, f_i^+]$ , for  $i \geq 0$ , is an increasing sequence of Scott continuous functions with  $f = \sup_{i \geq 0} f_i$ , and hence there exists  $i \geq 0$  such that  $f_i \in O$ . Since this holds for all open sets  $O$  containing  $f$ , it follows that  $[\int f^- d\mu, \int f^+ d\mu] \subseteq \int^\dagger f$ . If  $(a, b) \subset \mathbb{R}$  is any open interval with  $[\int f^- d\mu, \int f^+ d\mu] \subset (a, b)$ , then by the monotone convergence theorem applied to the sequences  $f_i^-$  and  $f_i^+$ , there exists  $i \geq 0$  such that  $[\int f_i^- d\mu, \int f_i^+ d\mu] \subset (a, b)$ —that is,  $\int^\dagger f = [\int f^- d\mu, \int f^+ d\mu]$ .  $\square$

We now obtain a short proof for the interval version of Green's theorem, a main result in the work of Edalat et al. [27].

THEOREM 6.2. *The composition*

$$\left( \int_{[0,1]} \right)^\dagger \circ (D_p)^\dagger : \text{Lip}(U) \rightarrow \mathbb{IR}$$

is given by

$$\left( \int_{[0,1]} \right)^\dagger \circ (D_p)^\dagger(f) = \left[ \int_0^1 \langle \partial f(p(t)), p'(t) \rangle^- dt, \int_0^1 \langle \partial f(p(t)), p'(t) \rangle^+ dt \right] \quad (21)$$

and satisfies

$$f(p(1)) - f(p(0)) \in \left( \int_{[0,1]} \right)^\dagger \circ (D_p)^\dagger(f). \quad (22)$$

PROOF. Take  $x = p(t)$  and  $v = p'(t)$  in Equation (18). Then from the proof of Theorem 5.13, the expression for  $(D_p)^\dagger$  in Equation (20) gives us  $(D_p)^\dagger(f) = \lambda t. \langle ((\partial f)(p(t)), p'(t)) \rangle$ . Equation (21) now follows from Proposition 6.1(ii). Next consider the operator

$$\int_{[0,1]} \circ D_p : C^1(U) \cap \text{Lip}(U) \rightarrow \mathbb{R},$$

with  $(\int_{[0,1]} \circ D_p)(g) = \int_0^1 g'(p(t))p'(t) dt = g(p(1)) - g(p(0))$ . Its lower limit has type

$$\left( \int_{[0,1]} \circ D_p \right)^\dagger : \text{Lip}(U) \rightarrow \mathbb{R}$$

and is continuous with respect to the L-topology by Proposition 3.2(iii) since  $C^1(U) \cap \text{Lip}(U)$  is a Hausdorff space. Let  $h_k \in C^1(U) \cap \text{Lip}(U)$ ,  $k \geq 0$ , be any sequence with  $\lim_{k \rightarrow \infty} h_k = f$  in the L-topology. Since  $(\int_{[0,1]} \circ D_p)^\dagger$  is continuous with respect to the L-topology and  $\int_{[0,1]} \circ D_p$  is continuous with respect to the relative subspace L-topology, we have

$$\begin{aligned} \left( \int_{[0,1]} \circ D_p \right)^\dagger(f) &= \lim_{k \rightarrow \infty} \left( \int_{[0,1]} \circ D_p \right)^\dagger(h_k) = \lim_{k \rightarrow \infty} \left( \int_{[0,1]} \circ D_p \right)(h_k) = \lim_{k \rightarrow \infty} h_k(p(1)) - h_k(p(0)) \\ &= f(p(1)) - f(p(0)). \end{aligned}$$

But the composition  $(\int_{[0,1]})^\dagger \circ (D_p)^\dagger$  of Scott continuous functions is Scott continuous, and if  $g : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  is  $C^1$ , then  $(\int_{[0,1]})^\dagger \circ (D_p)^\dagger(g) = g(p(1)) - g(p(0)) = (\int_{[0,1]} \circ D_p)(g)$ . Thus,  $(\int_{[0,1]})^\dagger \circ (D_p)^\dagger$  is a lower extension of  $\int_{[0,1]} \circ D_p$ . Since  $D_p$  and  $\int_{[0,1]}$  are both continuous functionals, by Proposition 3.11 we have  $(D_p)^\dagger = (D_p)^*$  as well as  $(\int_{[0,1]})^\dagger = (\int_{[0,1]})^*$ . Thus, Relation (22) follows from Proposition 3.2(i).  $\square$

Clearly, the composition in Equation (21) is in general interval valued rather than real valued. For example, let  $n = 1$ ,  $U = (-1, 2)$  and  $p(t) = t$  and consider the Lipschitz function  $f : (-1, 2) \rightarrow \mathbb{R}$  with  $\partial f(t) = [0, 1]$  for all  $t \in (-1, 2)$  as in the work of Edalat [19, Lemma 7.8]. Then,  $f(p(1)) - f(p(0)) = f(1) - f(0)$  but since  $p'(t) = 1$  and  $\partial f(t) = [0, 1]$  for all  $t \in [0, 1]$ , we obtain  $(\int_{[0,1]})^\dagger \circ (D_p)^\dagger(f) = [0, 1]$ . We have thus constructed an example in which the lower limit of the composition of two higher-order maps is not equal to the composition of the lower limits of the two maps.

## 7 SMOOTH APPROXIMATION OF GENERALISED JACOBIAN

In this section, we will extend the results of Sections 5 to vector Lipschitz maps. Let  $C^1(U, \mathbb{R}^m)$  denote, for an open set  $U \subset \mathbb{R}^n$ , the set of continuously differentiable maps of type  $U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ , and let  $\text{Lip}(U, \mathbb{R}^m)$ , respectively  $\text{Lip}_0(U, \mathbb{R}^m)$ , be the set of Lipschitz maps, respectively locally Lipschitz maps,  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ . We will use the following closed expression for the generalised Jacobian derived by Imbert [36], which can be viewed as a divergence theorem for Lipschitz vector maps. Let  $M_{n,m} \cong \mathbb{R}^{m \times n}$  denote the collection of  $m \times n$  real matrices with the inner product  $\langle A, B \rangle := \text{Tr}(A^T B) = \sum_{j=1}^n \sum_{i=1}^m A_{ij} B_{ij}$ , which induces the Frobenius norm. For  $x \in \mathbb{R}^n$  and  $\epsilon > 0$ , consider the hyper-cube  $P_\epsilon(x)$  of volume  $\epsilon^n$  with sides emanating from  $x$  in the direction of the canonical unit vectors  $e_i$  with  $1 \leq i \leq n$ —that is,

$$P_\epsilon(x) = \left\{ x + \epsilon \sum_{i=1}^n t_i e_i : 0 \leq t_i \leq 1, 1 \leq i \leq n \right\}.$$



Let  $(P_\epsilon(x))^b$  be the boundary<sup>1</sup> of  $P_\epsilon(x)$  with  $n(y)$  the outer unit normal at  $y \in (P_\epsilon(x))^b$  and  $s$  the surface Lebesgue measure on  $(P_\epsilon(x))^b$ . Consider  $f \in \text{Lip}_0(U, \mathbb{R}^m)$ . Then, the following two results are obtained in the work of Imbert [36]. For  $n \geq 2$ , the support function of  $\partial f(x_0)$  in the direction  $M \in M_{n,m}$  is given by

$$\sup \langle M, \partial f(x_0) \rangle = \limsup_{x \rightarrow x_0, \epsilon \rightarrow 0^+} \frac{1}{\epsilon^n} \int_{(P_\epsilon(x))^b} \langle f(y), Mn(y) \rangle ds. \quad (23)$$

For  $n = 1$ , note that  $\langle A, B \rangle = \langle A, B \rangle$ , for  $A, B \in \mathbb{R}^{m \times 1}$  and we have

$$\sup \langle M, \partial f(x_0) \rangle = \sup \langle M, \partial f(x_0) \rangle = g^\circ(x_0; 1), \quad (24)$$

where  $g = \langle M, f \rangle : \mathbb{R} \rightarrow \mathbb{R}$  [36].

**PROPOSITION 7.1.** *If a sequence of maps  $h_k \in \text{Lip}_0(U, \mathbb{R}^m)$ , for  $k \geq 0$ , converges uniformly to a map  $f \in \text{Lip}_0(U, \mathbb{R}^m)$ , then  $\limsup_{x \rightarrow x_0, k \rightarrow \infty} \langle M, \partial h_k(x) \rangle \geq \sup \langle M, \partial f(x_0) \rangle$  for all  $x_0 \in U$  and  $M \in M_{n,m}$ .*

**PROOF.** Assume first that  $n \geq 2$ ,  $\delta > 0$ ,  $K > 0$  and  $\alpha > 0$  are given. Then, by Equation (23), there exist  $x \in U$ , with  $\|x - x_0\| < \alpha/2$ , and  $\epsilon > 0$ , with  $\epsilon < \alpha/2n$ , such that  $P_\epsilon(x) \subseteq U$  and

$$\sup \langle \partial f(x_0), M \rangle - \frac{\delta}{2} < \frac{1}{\epsilon^n} \int_{(P_\epsilon(x))^b} \langle f(y), Mn(y) \rangle ds(y). \quad (25)$$

Since  $h_k \rightarrow f$  uniformly as  $k \rightarrow \infty$ , there exists  $k \geq K$  such that

$$\left| \frac{1}{\epsilon^n} \int_{(P_\epsilon(x))^b} \langle f(y), Mn(y) \rangle ds(y) - \frac{1}{\epsilon^n} \int_{(P_\epsilon(x))^b} \langle h_k(y), Mn(y) \rangle ds(y) \right| < \delta/2. \quad (26)$$

However, since  $h_k : U \rightarrow \mathbb{R}^m$  is differentiable almost everywhere by Rademacher's theorem [12, page 148], using Gauss-Green (divergence) theorem [40, Theorem 2.9], we obtain

$$\begin{aligned} \frac{1}{\epsilon^n} \int_{(P_\epsilon(x))^b} \langle h_k(y), Mn(y) \rangle ds(y) &= \frac{1}{\epsilon^n} \int_{(P_\epsilon(x))^b} \langle M^T h_k(y), n(y) \rangle ds(y) \\ &= \frac{1}{\epsilon^n} \int_{P_\epsilon(x)} \nabla \cdot (M^T h_k(y)) d\mu(y) = \frac{1}{\epsilon^n} \int_{P_\epsilon(x)} \langle h'_k(y), M \rangle d\mu(y), \end{aligned} \quad (27)$$

where  $\nabla \cdot g = \sum_{i=1}^{\ell} \frac{\partial g_i}{\partial x_i}$  is the divergence of the vector field  $g : \mathbb{R}^\ell \rightarrow \mathbb{R}^\ell$ . Thus, by using Relations (25), (26) and (27),

$$\sup \langle \partial f(x_0), M \rangle - \delta < \frac{1}{\epsilon^n} \int_{P_\epsilon(x)} \langle h'_k(y), M \rangle d\mu(y).$$

From the preceding inequality, it follows that  $\sup \langle \partial f(x_0), M \rangle - \delta < \langle h'_k(y_0), M \rangle \leq \langle \partial h_k(y_0), M \rangle$  for some  $y_0 \in P_\epsilon(x)$  (in fact, for  $y_0$  in a subset of  $P_\epsilon(x)$  of positive Lebesgue measure). Then  $\|y_0 - x_0\| \leq \|y_0 - x\| + \|x - x_0\| \leq \sqrt{n}\epsilon + \alpha/2 < \alpha$ . Since  $\delta > 0$ ,  $K > 0$  and  $\alpha > 0$  are arbitrary, the result follows for  $n \geq 2$ .

For  $n = 1$ , let  $\delta > 0$ ,  $K > 0$  and  $\alpha > 0$  be given. There exist, by Equation (24),  $t > 0$  and  $x \in U$ , with  $t < \alpha$  and  $|x - x_0| < \alpha$ , such that  $(x - t, x + t) \subseteq U$  and  $\sup \langle M, \partial f(x) \rangle - \delta/2 < (g(x + t) - g(x))/t$ ,

<sup>1</sup>We avoid the usual notation  $\partial A$  for the boundary of a set  $A$  as it can be confused here with the generalised Jacobian.

where  $g = \langle M, f \rangle$ . Put  $g_k = \langle M, h_k \rangle$ . Since  $g_k \rightarrow g$  uniformly as  $k \rightarrow \infty$ , there exists  $N \geq K$  such that for  $k \geq N$ , we have

$$\left| \frac{g_k(x+t) - g_k(x)}{t} - \frac{g(x+t) - g(x)}{t} \right| < \delta/2.$$

Thus, for  $k \geq N$ , we have

$$\sup \langle M, \partial f(x) \rangle - \delta < (g_k(x+t) - g_k(x))/t.$$

By Lebourg's mean value theorem, there exists  $y \in (x, x+t)$  such that

$$\sup \langle M, \partial f(x) \rangle - \delta < (g_k(x+t) - g_k(x))/t \leq \partial g_k(y).$$

Since  $\delta > 0$ ,  $K > 0$  and  $\alpha > 0$  are arbitrary, the result follows.  $\square$

The L-topology for locally Lipschitz vector maps  $\text{Lip}_0(U, \mathbb{R}^m)$  is defined similarly as for locally Lipschitz scalar maps in Section 5: it is the weakest refinement of the sup norm topology on  $\text{Lip}_0(U, \mathbb{R}^m)$  that makes the generalised Jacobian  $\partial : \text{Lip}_0(U, \mathbb{R}^m) \rightarrow \mathcal{C}\mathbb{R}^{m \times n}$  continuous with respect to the Scott topology on  $\mathcal{C}\mathbb{R}^{m \times n}$ . The following theorem can now be deduced with a proof similar to that of Theorem 5.4. Since the inner product  $\langle\langle A, B \rangle\rangle$ , when  $n = 1$ , is reduced to  $\langle A, B \rangle$ , we can state the result below uniformly for all  $n \geq 1$ .

**THEOREM 7.2.** *A sequence  $h_k \in \text{Lip}_0(U, \mathbb{R}^m)$ ,  $k \geq 0$ , converges to  $f \in \text{Lip}_0(U, \mathbb{R}^m)$  in the L-topology iff  $h_k$  converges to  $f$  in the sup norm topology and for all  $x_0 \in U$  and  $M \in M_{n,m}$ , we have*

$$\limsup_{k \rightarrow \infty, x \rightarrow x_0} \langle\langle \partial h_k(x), M \rangle\rangle = \sup \langle\langle \partial f(x_0), M \rangle\rangle$$

where the convergence is uniform for  $M \in S^{mn-1}$ .

**COROLLARY 7.3.** *A sequence  $h_k \in C^1(U, \mathbb{R}^m)$ ,  $k \geq 0$ , converges to  $f \in \text{Lip}_0(U, \mathbb{R}^m)$  in the L-topology iff  $h_k$  converges to  $f$  in the sup norm topology and for all  $x_0 \in U$  and  $M \in M_{n,m}$ , we have*

$$\limsup_{k \rightarrow \infty, x \rightarrow x_0} \langle\langle h'_k(x), M \rangle\rangle = \sup \langle\langle \partial f(x_0), M \rangle\rangle$$

where the convergence is uniform for  $M \in S^{mn-1}$ .

To show that  $C^1(U, \mathbb{R}^m) \cap \text{Lip}(U, \mathbb{R}^m)$  is dense in  $\text{Lip}(U, \mathbb{R}^m)$ , suppose  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  with components  $f = (f_i)_{1 \leq i \leq m}$  is a Lipschitz map with Lipschitz constant  $c > 0$  and  $\mu$  is the standard Gaussian probability measure on  $\mathbb{R}^n$ . For each positive integer  $k$ , as in Theorem 5.10 of Section 5.2, consider the convolutions  $h_k : \mathbb{R}^n \rightarrow \mathbb{R}^m$  with components  $h_k = (h_{ki})_{1 \leq i \leq m}$  given by  $h_{ki} = f_i * \mathcal{G}_k$ , where  $\mathcal{G}_k(x) = k^n \mathcal{G}(kx)$  and  $\mathcal{G}$  is the standard Gaussian probability distribution of Equation (12).

**THEOREM 7.4.** *For each  $k \geq 1$ , we have  $h_k \in C^\infty(U, \mathbb{R}^m)$  and the sequence  $h_k$  converges to  $f$  in the L-topology.*

**PROOF.** Applying Theorem 5.10 to each component  $h_{ki}$ , for  $1 \leq i \leq m$  and  $k \geq 1$ , it follows that  $h_k \in C^\infty(U, \mathbb{R}^m)$  with  $h_k \rightarrow f$  in the sup norm topology. It is now convenient to express  $h_{ki}$ , for  $1 \leq i \leq m$ , in terms of the probability measure  $\mu$  as follows:

$$h_{ki}(x) = \int_{U-U} f_i(x+y) \mathcal{G}_k(y) dy = \int_{V-V} f_i(x+k^{-1}w) \mathcal{G}(w) dw = \int_{V-V} f_i(x+k^{-1}w) d\mu(w), \quad (28)$$

with  $w = ky$  where  $V = kU = \{ku : u \in U\}$ . In addition, we express Equation (28) in vector notation—that is,

$$h_k(x) = \int_{V-V} f(x+k^{-1}w) d\mu(w). \quad (29)$$

To show that the convergence as in Theorem 7.3 is uniform for  $M \in S^{mn-1}$ , we first compute the following triple limit superior. For  $n \geq 2$ , using Formula (23) for the Jacobian  $h'_k$  of  $h_k$ , Fubini's theorem and the reverse Fatou's lemma, we obtain

$$\begin{aligned}
& \limsup_{\substack{k \rightarrow \infty \\ Q \rightarrow M \\ X \rightarrow X_0}} \langle h'_k(x), Q \rangle \\
&= \limsup_{\substack{k \rightarrow \infty \\ Q \rightarrow M \\ X \rightarrow X_0}} \limsup_{\substack{\epsilon \rightarrow 0 \\ z \rightarrow x}} \frac{1}{\epsilon^n} \int_{(P_\epsilon(z))^b} \langle h_k(u), Qn(u) \rangle ds(u) && \text{by Equation (23)} \\
&= \limsup_{\substack{k \rightarrow \infty \\ Q \rightarrow M \\ X \rightarrow X_0}} \limsup_{\substack{\epsilon \rightarrow 0 \\ z \rightarrow x}} \frac{1}{\epsilon^n} \int_{(P_\epsilon(z))^b} \int_{V-V} \langle f(u + k^{-1}w), Qn(u) \rangle d\mu(w) ds(u) && \text{by Equation (29)} \\
&= \limsup_{\substack{k \rightarrow \infty \\ Q \rightarrow M \\ X \rightarrow X_0}} \limsup_{\substack{\epsilon \rightarrow 0 \\ z \rightarrow x}} \frac{1}{\epsilon^n} \int_{V-V} \int_{(P_\epsilon(z))^b} \langle f(u + k^{-1}w), Qn(u) \rangle ds(u) d\mu(w) && \text{by Fubini's theorem} \\
&\leq \limsup_{\substack{k \rightarrow \infty \\ Q \rightarrow M \\ X \rightarrow X_0}} \int_{V-V} \limsup_{\substack{\epsilon \rightarrow 0 \\ z \rightarrow x}} \frac{1}{\epsilon^n} \int_{(P_\epsilon(z))^b} \langle f(u + k^{-1}w), Qn(u) \rangle ds(u) d\mu(w) && \text{by reverse Fatou's lemma} \\
&\leq \limsup_{\substack{k \rightarrow \infty \\ Q \rightarrow M \\ X \rightarrow X_0}} \int_{V-V} \sup \langle \partial f(x + k^{-1}w), Q \rangle d\mu(w) && \text{by Equation (23).}
\end{aligned}$$

We claim that the last term is bounded above by  $\sup \langle \partial f(x_0), M \rangle$ . Let  $\epsilon > 0$  be given. Let  $r > 0$  be large enough so that  $\int_{\|w\| \geq r} d\mu < \epsilon/2c(\|M\| + 1)$ . Since  $\partial f$  is upper semi-continuous at  $x_0$  and the inner product is continuous, it follows that there exists  $\delta > 0$  with  $\delta < 1$  such that for  $\|y - x_0\| < \delta$  and  $\|Q - M\| < \delta$  with  $Q \in S^{n-1}$ , we have  $\sup \langle \partial f(y), Q \rangle < \sup \langle \partial f(x_0), M \rangle + \epsilon/2$ . Thus, for  $\|x - x_0\| < \delta/2$ ,  $\|w\| < r$  and  $k > 2r/\delta$ , we have  $\|x + k^{-1}w - x_0\| \leq \|x - x_0\| + k^{-1}r < \delta/2 + \delta/2 = \delta$  which, denoting the open ball of radius  $r$  centred at the origin by  $O_r$ , implies

$$\begin{aligned}
& \int_{V-V} \sup \langle \partial f(x + k^{-1}w), Q \rangle d\mu(w) \\
&= \int_{(V-V) \cap O_r} \sup \langle \partial f(x + k^{-1}w), Q \rangle d\mu(w) + \int_{(V-V) \setminus O_r} \sup \langle \partial f(x + k^{-1}w), Q \rangle d\mu(w) \\
&< \sup \langle \partial f(x_0), M \rangle + \frac{\epsilon}{2} + \frac{c(\|M\| + 1)\epsilon}{2c(\|M\| + 1)} = \sup \langle \partial f(x_0), M \rangle + \epsilon.
\end{aligned}$$

Since  $\epsilon > 0$  is arbitrary, our claim follows. Thus,

$$\limsup_{\substack{k \rightarrow \infty \\ Q \rightarrow M \\ X \rightarrow X_0}} \langle h'_k(x), Q \rangle \leq \sup \langle \partial f(x_0), M \rangle. \quad (30)$$

Now, to use Lemma 5.3 to complete the proof, let  $\epsilon > 0$  be given. Since  $\partial f(x) \subseteq \mathbb{R}^{m \times n}$  is a compact convex set, the support function  $\sigma_{\partial f(x)} : M \mapsto \sup \langle \partial f(x), M \rangle : S^{mn-1} \rightarrow \mathbb{R}$  is continuous and hence uniformly continuous on  $S^{mn-1}$ . Thus, there exists  $\delta_0 > 0$  such that for  $\|M_1 - M_2\| < \delta_0$  with  $M_1, M_2 \in S^{mn-1}$ , we have  $|\sup \langle \partial f(x), M_1 \rangle - \sup \langle \partial f(x), M_2 \rangle| < \epsilon/2$ . By Relation (30), there exists  $\delta_1 > 0$ ,  $\alpha_M > 0$  and  $k_M \geq 0$  such that for  $\|Q - M\| < \delta_1$  with  $Q \in S^{mn-1}$ ,  $\|x - x_0\| < \alpha_M$  and  $k \geq k_M$ , we have  $\langle h'_k(x), Q \rangle < \sup \langle \partial f(x_0), M \rangle + \epsilon/2$ . Hence, for  $\|Q - M\| < \delta_M = \min\{\delta_0, \delta_1\}$ ,  $\|x - x_0\| < \alpha_M$  and  $k \geq k_M$  we have  $\langle h'_k(x), Q \rangle < \sup \langle \partial f(x_0), M \rangle + \epsilon/2 < \sup \langle \partial f(x_0), Q \rangle + \epsilon$ . Consider the open cover of  $S^{mn-1}$  by balls of centre  $M$  and radius  $\delta_M$  and take a finite cover given

by  $M_i$  for  $i = 1, \dots, \ell$ , say. It follows that for  $\|x - x_0\| < \alpha = \min\{\alpha_{M_i} : 1 \leq i \leq \ell\}$ ,  $k \geq \max\{k_{M_i} : 1 \leq i \leq \ell\}$  and all  $Q \in S^{mn-1}$ , we have  $\langle h'_k(x), Q \rangle < \sup\langle \partial f(x_0), Q \rangle + \epsilon$  and the proof is complete for  $n \geq 2$ .

For  $n = 1$ , with  $g_Q = \langle Q, f \rangle$ , we have

$$\langle Q, h_k(x) \rangle = \int_{V-V} \langle Q, f(x + k^{-1}y) \rangle d\mu(y) = \int_{V-V} g_Q(x + k^{-1}y) d\mu(y),$$

where  $\mu$  is now the one-dimensional standard Gaussian distribution. Put

$$p_\ell(x) := \frac{\langle Q, h_k(x + 1/\ell) \rangle - \langle Q, h_k(x + 1/\ell) \rangle}{1/\ell}$$

for integers  $\ell \geq 1$ . Since  $\left| \frac{g_Q(x+k^{-1}y+\ell^{-1})-g_Q(x+k^{-1}y)}{\ell^{-1}} \right| \leq c\|Q\|$ , we can apply the dominated convergence theorem to obtain

$$\begin{aligned} \langle Q, h'_k(x) \rangle &= \lim_{\ell \rightarrow \infty} p_\ell(x) = \int_{V-V} \lim_{\ell \rightarrow \infty} \frac{g_Q(x+k^{-1}y+\ell^{-1})-g_Q(x+k^{-1}y)}{\ell^{-1}} d\mu(y) \\ &= \int_{V-V} g'_Q(x + k^{-1}y; 1) d\mu(y), \end{aligned}$$

for  $Q \in S^{m-1}$ . Now using reverse Fatou's lemma, we deduce

$$\begin{aligned} \limsup_{\substack{Q \rightarrow M \\ \bar{x} \rightarrow x_0}} \langle Q, h'_k(x) \rangle &= \limsup_{\substack{Q \rightarrow M \\ \bar{x} \rightarrow x_0}} \int_{V-V} g'_Q(x + k^{-1}y; 1) d\mu(y) \\ &\leq \int_{V-V} \limsup_{\substack{Q \rightarrow M \\ \bar{x} \rightarrow x_0}} g'_Q(x + k^{-1}y; 1) d\mu(y) \leq g_M^\circ(x_0; 1), \end{aligned}$$

where the latter inequality easily follows from the definition of  $g_M^\circ(x_0; 1)$ . The uniform convergence for  $M \in S^{m-1}$ , in view of the compactness of  $S^{m-1}$ , now follows with a proof similar to the case of  $n \geq 2$  presented after Equation (30).  $\square$

Theorem 7.4 has finally extended Theorem 5.10 to vector Lipschitz maps, and we conclude the following.

**COROLLARY 7.5.**  $C^1(U, \mathbb{R}^m)$  is dense in  $\text{Lip}(U, \mathbb{R}^m)$  with respect to the  $L$ -topology.

We can now also obtain a simple proof of the mean value theorem for vector Lipschitz functions.

**COROLLARY 7.6.** (cf. [11, 2.6.5]) Suppose  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  is Lipschitz in an open set containing the line segment  $s(x, y) \subseteq U$ . Then,  $f(x) - f(y) \in \text{co}\{\partial f[s(x, y)]\}(x - y)$ .

**PROOF.** Let  $C \in \mathcal{C}\mathbb{R}^{m \times n}$  with  $C \ll \text{co}\{\partial f[s(x, y)]\}$ . Then,  $s(x, y) \subseteq (\partial f)^{-1}(\uparrow C)$  where the latter set is open. Take open set  $O \subseteq \mathbb{R}^n$ , on which  $f$  is Lipschitz, with  $s(x, y) \subseteq O \ll (\partial f)^{-1}(\uparrow C)$ . Then  $C\chi_O \ll \partial f$  by Gierz et al. [33, Proposition II-4-20]. Suppose  $h_k : O \rightarrow \mathbb{R}^m$  is a sequence of  $C_{n,m}^1(U)$  functions that converges to  $f$  in the  $L$ -topology. Thus, there exists  $N$  such that  $n \geq N$  implies  $C\chi_O \ll h'_k$  and in particular  $h'_k[s(x, y)] \subseteq C$ . We now apply the classical mean value theorem to obtain  $h_k(x) - h_k(y) \in (h'_k[s(x, y)])(x - y) \subseteq C(x - y)$ . Since  $h_k \rightarrow f$  in the sup norm topology, it follows that  $f(x) - f(y) \in C(x - y)$ . As  $C \ll \text{co}\{\partial f[s(x, y)]\}$  is arbitrary, we conclude that  $f(x) - f(y) \in \text{co}\{\partial f[s(x, y)]\}(x - y)$ .  $\square$

Finally, we have the counterpart of Theorem 5.13 for vector Lipschitz maps, which is proved in a similar way. Let  $D_{n,m} : C^1(U, \mathbb{R}^m) \cap \text{Lip}(U, \mathbb{R}^m) \rightarrow (U \rightarrow \mathcal{C}\mathbb{R}^{m \times n})$  denote the vector differential operator with  $D_{n,m}(f) = f'$ .

**THEOREM 7.7.** The lower limit and the lower envelope of the differential operator  $D_{n,m}$  coincide with the subgradient operator—that is,  $D_{n,m}^\dagger = D_{n,m}^* = \partial$  with  $D_{n,m}^\dagger(f) = D_{n,m}^*(f) = \partial f$ .

## 8 SUBGRADIENT ON BANACH SPACES

In this section, we consider Lipschitz maps of type  $X \rightarrow \mathbb{R}$ , where  $X$  is a real Banach space with  $\|x\|$  as the norm of  $x \in X$ . If  $C \subseteq X^*$  is a non-empty weak\* compact and convex set, then its support function  $\sigma_C : X \rightarrow \mathbb{R}$  is given by  $\sigma_C(v) = \sup\{Lv : L \in C\}$ .

Consider a Lipschitz map  $f : X \rightarrow \mathbb{R}$  with Lipschitz constant  $c > 0$  and the bounded complete domain  $\mathcal{C}(X_c^*)$  of non-empty weak\* compact and convex subsets of  $X_c^*$  as presented at the end of Section 2. Since  $X_c^*$  is a compact Hausdorff space with respect to the relative subspace weak\* topology on  $X_c^*$ , the upper topology and the Scott topology coincide on it [17]. The subgradient  $\partial f(x) \in \mathcal{C}(X_c^*)$  of  $f$  at  $x \in X$  is a compact convex subset of  $X_c^*$  which, as in the finite dimensional case, can be defined by its support function [11, p. 28]:

$$\sigma_{\partial f(x)}(v) = \limsup_{y \rightarrow x, t \rightarrow 0^+} \frac{f(y + tv) - f(y)}{t}.$$

The subgradient coincides with the so-called L-derivative as introduced in the work of Edalat [19] which we will now define. We start by observing that  $X_c^*$  can be identified with the set of maximal elements of the bounded complete domain  $\mathcal{C}(X_c^*)$ . Moreover,  $X_c^*$ , identified as usual with the subset of maximal elements of  $\mathcal{C}(X_c^*)$ , is dense in  $\mathcal{C}(X_c^*)$ . Hence, for  $x \in X$ , the linear map  $\hat{x} : X_c^* \rightarrow \mathbb{R}$ , with  $\hat{x}(L) = Lx$  has continuous extension  $\hat{x}^\dagger : \mathcal{C}(X_c^*) \rightarrow \mathbb{R}$ , with

$$\hat{x}^\dagger(C) = \{\hat{x}(L) : L \in C\} = \{Lx : L \in C\}.$$

Given an open set  $O \subseteq X$  and  $C \in \mathcal{C}(X^*)$  a weak\* compact convex set, the tie  $\delta(O, C)$  is defined to be the set of all Lipschitz maps  $f : X \rightarrow \mathbb{R}$  such that  $f(x) - f(y) \in C(x - y)$  for all  $x, y \in O$ , where  $C(x - y) := (\widehat{x - y})^\dagger(C) = \{L(x - y) : L \in C\}$ , which is a compact interval. It was shown in the work of Edalat [19] that the map  $\mathcal{L}f : X \rightarrow \mathcal{C}(X^*)$  given by

$$\mathcal{L}f(x) = \bigcap \{C : f \in \delta(O, C), x \in O\} \quad (31)$$

is well defined and upper continuous with  $\partial f(x) \subseteq \mathcal{L}f(x)$ , and in addition if  $X$  is finite dimensional, then  $\partial f = \mathcal{L}f$ . Hertling [34] showed later that in fact  $\partial f(x) = \mathcal{L}f(x)$  for any Banach space  $X$ . This leads to a short proof of the upper continuity of the subgradient.

**PROPOSITION 8.1.** *For any Lipschitz map  $f : U \subseteq X \rightarrow \mathbb{R}$  on a Banach space  $X$ , the subgradient  $\partial f$  is upper continuous.*

**PROOF.** If  $c$  is a Lipschitz constant for  $f$ , then we have  $\partial f : U \rightarrow \mathcal{C}(X_c^*)$ . We know that  $\mathcal{L}f : X \rightarrow \mathcal{C}(X_c^*)$  is Scott continuous [19] and  $\partial f = \mathcal{L}f$  [34]. But since  $\mathcal{C}(X_c^*)$  is a compact Hausdorff space, the Scott topology on  $\mathcal{C}(X_c^*)$  coincides with the upper topology [17].  $\square$

### 8.1 Approximation of Subgradient on Separable Banach Spaces

For an infinite dimensional Banach space  $X$ , the function space  $(X \rightarrow \mathcal{C}(X_c^*))$  is no longer a domain and the Scott topology on it does not have a simple representation. However, if  $X$  is separable, we show that there is a sequence of Gateaux differentiable functions converging to a given Lipschitz map in the sup norm topology such that the limit superior of the sequence the Gateaux derivatives of the functions in any given direction converges to the generalised directional derivative of the Lipschitz map in that direction. We assume from now on that  $X$  is actually a separable Banach space and  $f : U \subseteq X \rightarrow \mathbb{R}$  is Lipschitz with Lipschitz constant  $c > 0$ .

Recall that a *Gaussian measure* on the separable Banach space  $X$  is a probability Borel measure  $\mu$  on  $X$  such that for every  $L \in X^*$  the induced measure  $\mu_L$  on  $\mathbb{R}$  given by  $L$  with  $\mu_L(A) = \mu\{x : Lx \in A\}$ , for any Borel set  $A \subseteq \mathbb{R}$ , has a Gaussian distribution on  $\mathbb{R}$ ; a *non-degenerate* Gaussian measure on the separable Banach space  $X$  is one such that for every non-zero  $L \in X^*$  the induced



measure  $\mu_L$  on  $\mathbb{R}$  has a non-degenerate Gaussian distribution on  $\mathbb{R}$  [9, 6.17]. We first describe how such a Gaussian measure can be constructed on a Hilbert space (see [9]). Consider a separable real Hilbert space  $H$  with inner product  $\langle x, y \rangle$  for  $x, y \in H$ . Assume  $Q \in H^* \simeq H$  is a trace class positive operator—that is, there exists an orthonormal basis  $(e_n)$  of  $H$  with  $\lambda_n > 0$  such that  $Q(e_n) = \lambda_n e_n$  for  $n \in \mathbb{N}$ , and  $\text{tr}(Q) = \sum_{n=0}^{\infty} \lambda_n < \infty$ . If  $m \in H$  is any given point, then the probability measure  $\mu$  with the characteristic function

$$\int_H e^{i\langle x, y \rangle} d\mu(y) = e^{i\langle x, m \rangle - \langle Qx, x \rangle / 2} = \prod_{n \geq 0} e^{i v_n a_n - \lambda_n a_n^2 / 2}, \quad x \in H, \quad (32)$$

where  $a_n = \langle x, e_n \rangle$  and  $v_n = \langle m, e_n \rangle$ , is a Gaussian measure on  $H$ . Thus, the coordinates  $a_n$ , for  $n \geq 0$ , of  $x$  with respect to the basis  $\{e_n\}$  are independent Gaussian variables with mean  $v_n$  and variance  $\lambda_n$ .

We can now use the Gaussian measure  $\mu$  constructed on  $H$  to obtain a Gaussian measure on the separable Banach space  $X$ . We fix a sequence  $(x_n)$  with a dense linear span in  $X$  such that  $\sum_{n \geq 0} \|x_n\|^2 < \infty$ . Consider the linear map  $T : H \rightarrow X$  with  $T(e_n) = x_n$ . It follows that the forward measure  $\mu_T$  on  $X$  induced by  $T$  is a non-degenerate Gaussian measure on  $X$  that is constructed from a simple Gaussian measure on a Hilbert space.

The key property we need to invoke to obtain the result of this section is the following.

**THEOREM 8.2** ([9, 6.25 AND 6.42]). *If  $f : X \rightarrow \mathbb{R}$  is a Lipschitz map on the separable Banach space  $X$ , then the Gateaux derivative  $df(x)$  exists everywhere on  $X$  except for  $x$  in a null set with respect to any non-degenerate Gaussian measure on  $X$ .*

We can now deal with the construction of a sequence of Gateaux differentiable functions converging to a given Lipschitz map with similar limiting properties as in the finite dimensional case in Theorem 5.5. We first observe that the proof of Lemma 5.1 extends to the case of a sequence of Gateaux differentiable maps on a Banach space converging in the sup norm topology to a Lipschitz map. This is because the mean value theorem, invoked in that lemma, also holds for any Gateaux differentiable map  $h : X \rightarrow \mathbb{R}$ . In fact if  $x, y \in X$  with  $x \neq y$ , define  $g : [0, 1] \rightarrow \mathbb{R}$  by  $g(t) = h(x + t(y - x))$ . Then  $g$  is continuous in  $[0, 1]$  and differentiable in  $(0, 1)$ , and hence there exists  $t_0 \in (0, 1)$  such that  $h(y) - h(x) = g(1) - g(0) = g'(t_0) = dh(z)(y - x)$  with  $z = x + t_0(y - x)$ . Thus, we have the following lemma.

**LEMMA 8.3.** *Suppose  $h_k : U \subseteq X \rightarrow \mathbb{R}$  is a sequence of Gateaux differentiable such that  $h_k \rightarrow f$  in the sup norm topology as  $k \rightarrow \infty$  where  $f$  is a Lipschitz map. Then for all  $x \in U$  and  $v \in \mathbb{R}^n$ , we have*

$$\limsup_{y \rightarrow x, k \rightarrow \infty} (dh_k(y))(v) \geq f^\circ(x; v).$$

Let  $f : X \rightarrow \mathbb{R}$  be a Lipschitz map with Lipschitz constant  $c > 0$ , where we have assumed for convenience (and without loss of generality) that the domain of  $f$  is the whole space  $X$ . We consider the construction in [9, 6.43] of a sequence of maps  $h_n : X \rightarrow \mathbb{R}$  given by

$$h_n(x) = \int f(x + n^{-1}y) d\mu(y)$$

for  $n \geq 1$ , where  $\mu$  is any non-degenerate Gaussian measure on  $X$ .

**LEMMA 8.4** ([9, 6.43]). *The maps  $h_n$  are uniformly Lipschitz with Lipschitz constant  $c$  and are Gateaux differentiable for  $n \geq 1$  with the sequence  $h_n$  converging in the sup norm topology to  $f$ .*

We can now obtain our final result.

**THEOREM 8.5.** *There is a sequence  $(h_n)_{n \geq 1}$  of uniformly Lipschitz and Gateaux differentiable maps converging in the sup norm topology to  $f$  such that we have for  $z, v \in X$ ,*

$$\limsup_{n \rightarrow \infty, x \rightarrow z, u \rightarrow v} (dh_n(x))(u) = \limsup_{n \rightarrow \infty, x \rightarrow z} (dh_n(x))(v) = f^\circ(z; v).$$

PROOF. Consider the sequence  $h_n$  for  $n \geq 1$  in Lemma 8.4. By Theorem 8.2, the Gateaux derivative  $df(x)$  exists everywhere on  $X$  except for  $x$  in a null set with respect to the non-degenerate Gaussian measure  $\mu$  on  $X$ . We can therefore obtain an integral expression with respect to  $\mu$  for the Gateaux derivative of  $h_n$  as follows. We have

$$\frac{h_n(x + tv) - h_n(x)}{t} = \int_X \frac{f(x + n^{-1}y + tv) - f(x + n^{-1}y)}{t} d\mu(y).$$

Since the integrand, in absolute value, on the RHS is bounded by  $c\|v\|$ , writing  $t = 1/\ell$  for integer  $\ell \geq 1$ , with  $t \rightarrow 0$  as  $\ell \rightarrow \infty$ , the dominated convergence theorem (as in the proof of Theorem 7.4) implies that  $f'(x + n^{-1}y; v)$  is integrable with respect to  $\mu$  with

$$(dh_n(x))(v) = \int_X f'(x + n^{-1}y; v) d\mu(y).$$

By Lemma 8.3, we already have

$$\limsup_{\substack{x \rightarrow z, n \rightarrow \infty, u \rightarrow v}} (dh_n(x))(u) \geq \limsup_{x \rightarrow z, n \rightarrow \infty} (dh_n(x))(v) \geq f^\circ(z; v),$$

where the first inequality follows from the continuity of the bounded linear map  $dh_n(x) : X \rightarrow \mathbb{R}$  for each  $x \in X$  and  $n \geq 1$ .

Let  $\epsilon > 0$ . Since the map  $(z, v) \rightarrow f^\circ(z; v)$  is upper continuous in  $z$  and  $v$  [11, 2.1.1(b)], there exists  $\delta > 0$  such that for  $\|x - z\| < \delta$  and  $\|u - v\| < \delta$  we have  $f^\circ(x; u) < f^\circ(z; v) + \epsilon/2$ . Hence, in the above neighbourhoods of  $z$  and  $v$ , whenever  $f'(x; u)$  exists, we have  $f'(x; u) < f^\circ(z; v) + \epsilon/2$ , which implies that  $\limsup_{x \rightarrow z, u \rightarrow v} f'(x; u) \leq f^\circ(z; v) + \epsilon/2$ . However, we have  $|f'(x + n^{-1}y; u)| \leq c\|u\| \leq c(\|v\| + \delta)$  for  $\|u - v\| < \delta$ . Let  $R > 0$  be such that  $\int_{\|y\| \geq R} d\mu(y) < \epsilon/(2c(\delta + \|v\|))$ . Then, the reverse Fatou's lemma yields

$$\begin{aligned} \limsup_{\substack{x \rightarrow z \\ n \rightarrow \infty \\ u \rightarrow v}} (dh_n(x))(u) &= \limsup_{\substack{x \rightarrow z \\ n \rightarrow \infty \\ u \rightarrow v}} \int f'(x + n^{-1}y; u) d\mu(y) \leq \int \limsup_{\substack{x \rightarrow z \\ n \rightarrow \infty \\ u \rightarrow v}} f'(x + n^{-1}y; u) d\mu(y) \\ &= \int_{\|y\| < R} \limsup_{\substack{x \rightarrow z \\ n \rightarrow \infty \\ u \rightarrow v}} f'(x + n^{-1}y; u) d\mu(y) + \int_{\|y\| \geq R} \limsup_{\substack{x \rightarrow z \\ n \rightarrow \infty \\ u \rightarrow v}} f'(x + n^{-1}y; u) d\mu(y) \\ &< f^\circ(z; v) + \frac{\epsilon}{2} + \frac{\epsilon c(\delta + \|v\|)}{2c(\delta + \|v\|)} = f^\circ(z; v) + \epsilon. \end{aligned}$$

Therefore, since  $\epsilon > 0$  is arbitrary, we have

$$\limsup_{\substack{x \rightarrow z \\ n \rightarrow \infty \\ u \rightarrow v}} (dh_n(x))(u) \leq f^\circ(z; v),$$

and the result follows.  $\square$

Comparing Theorem 8.5 with Theorem 5.10, we see that our result for approximation of a Lipschitz map on a separable Banach space provides a sequence of Gateaux differentiable maps rather than a sequence of  $C^\infty$  maps as was the case in finite dimensions. In addition, in finite dimensions, the convergence in Theorem 5.5 and thus in Theorem 5.10 is uniform on the unit sphere, a property which is fundamentally based on compactness properties in finite dimensions.

## 9 CONCLUDING REMARKS

We have obtained new representations for the subgradient of a real-valued locally Lipschitz map, defined on a finite dimensional Euclidean space, and also for the generalised Jacobian of a locally Lipschitz vector map between finite dimensional Euclidean spaces. These results lead us to simpler proofs for some of the basic properties of these generalised derivatives. The L-topology on the space of locally Lipschitz maps has been characterized in terms of convergent sequences in this space. We have shown that the set of  $C^1$  maps is dense in the space of Lipschitz maps equipped with the L-topology and that convergence of a sequence of  $C^1$  maps to a Lipschitz map in the L-topology is equivalent to the convergence of the sequence to the Lipschitz map in the sup norm topology and the uniform convergence of the limit superior of the sequence of the derivatives of the  $C^1$  maps in a given unit vector direction with the subgradient of the Lipschitz map in the direction of that unit vector. Given a Lipschitz map between finite dimensional Euclidean spaces, we have constructed a sequence of  $C^\infty$  maps that converges to the Lipschitz map in the L-topology. This result confirms that the L-topology is the appropriate topology on the space of Lipschitz maps when approximating a Lipschitz map with  $C^1$  or  $C^\infty$  maps. For a real-valued Lipschitz map on a separable Banach space, we also explicitly derived a sequence of Gateaux differentiable Lipschitz maps converging in the sup norm to the Lipschitz map such that the limit superior of the Gateaux derivative of the maps in any direction coincides with the subgradient of the Lipschitz map in that direction.

As for future work, the question arises if any of the new representations of the subgradient of a Lipschitz map between finite dimensional Euclidean spaces can be extended to find a new representation of the subgradient of a real-valued Lipschitz map on a separable Banach space.

## REFERENCES

- [1] S. Abramsky and A. Jung. 1994. Domain theory. In *Handbook of Logic in Computer Science*, S. Abramsky, D. M. Gabbay, and T. S. E. Maibaum (Eds.), Vol. 3. Clarendon.
- [2] V. Acary and B. Brogliato. 2008. *Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics*. Springer Science & Business Media.
- [3] J.-P. Aubin and A. Cellina. 2012. *Differential Inclusions: Set-Valued Maps and Viability Theory*. Vol. 264. Springer Science & Business Media.
- [4] J.-P. Aubin and H. Frankowska. 2009. *Set-Valued Analysis*. Springer Science & Business Media.
- [5] A. Auslender and M. Teboulle. 2006. *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer Science & Business Media.
- [6] Daniel Azagra, Juan Ferrera, Fernando López-Mesas, and Yenny Rangel. 2007. Smooth approximation of Lipschitz functions on Riemannian manifolds. *Journal of Mathematical Analysis and Applications* 326, 2 (2007), 1370–1378.
- [7] A. Beck and M. Teboulle. 2012. Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization* 22, 2 (2012), 557–580.
- [8] John Benedetto. 2013. *Real Variable and Integration: With Historical Notes*. Springer-Verlag.
- [9] Y. Benyamini and J. Lindenstrauss. 2000. *Geometric Nonlinear Functional Analysis*. American Mathematical Society.
- [10] R. N. Bracewell. 1986. *The Fourier Transform and Its Applications*. McGraw-Hill, New York, NY.
- [11] F. H. Clarke. 1990. *Optimization and Nonsmooth Analysis* (2nd ed.). Classics in Applied Mathematics, Vol. 5. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- [12] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. 1998. *Nonsmooth Analysis and Control Theory*. Springer.
- [13] E. A. Coddington and N. Levinson. 1955. *Theory of Ordinary Differential Equations*. McGraw-Hill.
- [14] T. A. Davis and K. Sigmon. 2005. *MATLAB Primer* (7 ed.). CRC Press, Boca Raton, FL.
- [15] J. Eckhoff. 1993. Helly, radon, and caratheodory type theorems. In *Handbook of Convex Geometry*. Vol. B. North-Holland, 389–448.
- [16] A. Edalat. 1995. Domain theory and integration. *Theoretical Computer Science* 151 (1995), 163–193.
- [17] A. Edalat. 1995. Dynamical systems, measures and fractals via domain theory. *Information and Computation* 120, 1 (1995), 32–48.
- [18] A. Edalat. 1997. Domains for computation in mathematics, physics and exact real arithmetic. *Bulletin of Symbolic Logic* 3, 4 (1997), 401–452.

- [19] A. Edalat. 2008. A continuous derivative for real-valued functions. In *New Computational Paradigms*. Springer, 493–519.
- [20] A. Edalat. 2009. A computable approach to measure and integration theory. *Information and Computation* 207 (2009), 642–659.
- [21] A. Edalat. 2010. A differential operator and weak topology for Lipschitz maps. *Topology and Its Applications* 157, 9 (2010), 1629–1650.
- [22] A. Edalat. 2015. A derivative for complex Lipschitz maps with generalised Cauchy–Riemann equations. *Theoretical Computer Science* 564 (2015), 89–106.
- [23] A. Edalat and R. Heckmann. 1998. A computational model for metric spaces. *Theoretical Computer Science* 193, 1–2 (1998), 53–73.
- [24] A. Edalat and R. Heckmann. 2002. Computing with real numbers: (i) LFT approach to real computation, (ii) Domain-theoretic model of computational geometry. In *Applied Semantics: Advanced Lectures*, G. Barthe, P. Dybjer, L. Pinto, and J. Saraiva (Eds.). Lecture Notes in Computer Science, Vol. 2395. Springer, 193–267.
- [25] A. Edalat and A. Lieutier. 2002. Foundation of a computable solid modelling. *Theoretical Computer Science* 284, 2 (2002), 319–345.
- [26] A. Edalat, A. Lieutier, and E. Kashefi. 2001. The convex hull in a new model of computation. In *Proceedings of the 13th Canadian Conference on Computational Geometry*.
- [27] A. Edalat, A. Lieutier, and D. Pattinson. 2013. A computational model for multi-variable differential calculus. *Information and Computation* 224 (2013), 22–45.
- [28] A. Edalat and D. Pattinson. 2007. Denotational semantics of hybrid automata. *Journal of Logic and Algebraic Programming* 73, 1–2 (2007), 3–21.
- [29] A. Edalat and D. Pattinson. 2007. A domain-theoretic account of Picard’s theorem. *LMS Journal of Computation and Mathematics* 10 (2007), 83–118.
- [30] A. Edalat, P. J. Potts, and P. Sünderhauf. 1998. Lazy computation with exact real numbers. In *Proceedings of the 3rd ACM SIGPLAN International Conference on Functional Programming*. ACM, New York, NY, 185–194.
- [31] R. Engelking. 1989. *General Topology*. Sigma Series in Pure Mathematics, Vol. 6. Heldermann Verlag, Berlin.
- [32] H. Federer. 1969. *Geometric Measure Theory*. Springer.
- [33] G. Gierz, K. H. Hofmann, K. Keimel, J. D. Lawson, M. Mislove, and D. S. Scott. 2003. *Continuous Lattices and Domains*. Cambridge University Press.
- [34] P. Hertling. 2017. Clarke’s generalized gradient and Edalat’s L-derivative. *Journal of Logic and Analysis* 9 (2017).
- [35] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. 2012. *Fundamentals of Convex Analysis*. Springer Science & Business Media.
- [36] C. Imbert. 2002. Support functions of Clarke generalized Jacobian and of its plenary hull. *Nonlinear Analysis: Theory, Methods and Applications* 49, 8 (2002), 111–1125.
- [37] G. Lebourg. 1979. Generic differentiability of Lipschitzian functions. *Transactions of American Mathematical Society* 256 (1979), 125–144.
- [38] R. E. Moore. 1966. *Interval Analysis*. Prentice-Hall.
- [39] K. P. Murphy. 2012. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.
- [40] Washek F. Pfeffer. 2005. The Gauss–Green theorem in the context of Lebesgue integration. *Bulletin of the London Mathematical Society* 37, 1 (2005), 81–94.
- [41] R. Schneider. 1993. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge University Press.
- [42] D. Scott. 1970. Outline of a mathematical theory of computation. In *Proceedings of the 4th Annual Princeton Conference on Information Sciences and Systems*. 169–176.
- [43] D. Scott. 1976. Data types as lattices. *SIAM Journal on Computing* 5, 3 (1976), 522–587.
- [44] D. R. Sherbert. 1963. Banach algebras of Lipschitz functions. *Pacific Journal of Mathematics* 13, 4 (1963), 1387–1399.
- [45] N. Z. Shor. 2012. *Minimization Methods for Non-Differentiable Functions*. Vol. 3. Springer Science & Business Media.
- [46] J. Warga. 1981. Fat homeomorphisms and unbounded derivate containers. *Journal of Mathematical Analysis and Applications* 81 (1981), 545–560.
- [47] S. Winskel. 1993. *The Formal Semantics of Programming Languages: An Introduction*. MIT Press, Cambridge, MA.
- [48] S. Yamamuro. 1970. *Differential Calculus in Topological Linear Spaces*. Lecture Notes in Mathematics, Vol. 374. Springer-Verlag.
- [49] James Yeh. 2006. *Real Analysis: Theory of Measure and Integration* (2nd ed.).

Received October 2019; revised July 2021; accepted August 2021