Journal of Convex Analysis
Volume 28 (2021), No. 4, 1223–1248
https://doi.org/10.68381/jca28072

Convex Analysis in Normed Spaces and Metric Projections onto Convex Bodies

Vitor Balestro*
Inst. de Matemática e Estatística, Universidade Federal Fluminense, 24210201 Niterói, Brazil
vitorbalestro@id.uff.br

Horst Martini
Fakultät für Mathematik, Technische Universität, 09107 Chemnitz, Germany
martini@mathematik.tu-chemnitz.de

Ralph Teixeira
Inst. de Matemática e Estatística, Universidade Federal Fluminense, 24210201 Niterói, Brazil
ralphct@id.uff.br

Received: April 4, 2020
Accepted: December 9, 2020

We investigate metric projections and distance functions referring to convex bodies in finite-dimensional normed spaces. For this purpose we identify the vector space with its dual space by using, instead of the usual identification via the standard inner product, the Legendre transform associated with the given norm. This approach yields re-interpretations of various properties of convex functions, and new relations between such functions and geometric properties of the studied norm are also derived. Here the concept of Birkhoff orthogonality, mostly common in finite-dimensional normed spaces, will play an important role.

Keywords: Birkhoff orthogonality, convex body, convex functions, (differentiability of) distance functions, Legendre transform, subgradient.

2010 Mathematics Subject Classification: 41A50, 41A65, 46B20, 46G05, 52A20, 52A21, 53C23, 58C20.

1. Introduction

Let (M, d) be a metric space, and let $C \subseteq M$ be a subset of M. The *distance* of a given point $x \in M$ to C is defined to be the number

$$\operatorname{dist}(x, C) = \inf\{d(x, y) : y \in C\}.$$

If this number is finite, and if $y_0 \in C$ is such that $d(x, y_0) = \operatorname{dist}(x, C)$, then we say that y_0 is a *metric projection* of x onto C.

Metric projections have been widely studied through the past decades and in a number of contexts (see, e.g., [1], [4], [11] and [36], further references will be given throughout the text). A very special case is when one considers the metric projections onto a given *convex body* (i.e., a compact, convex set with non-empty interior) K in $\mathbb{R}^n$ endowed with the standard distance given for the Euclidean norm (which we denote by $|\cdot|$). For this case we know that, for example,

* Corresponding author

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

(i) K is a *Chebyshev set*, meaning that a metric projection onto K exists and is unique for any $x \in \mathbb{R}^n$. Actually, every Chebyshev set of $\mathbb{R}^n$ is convex, see [9] and [12]. For infinite-dimensional vector spaces this is more complicated, and we refer the reader to [32] and [31],

(ii) the map $p_K \colon \mathbb{R}^n \to K$ which takes each $x \in \mathbb{R}^n$ to its metric projection onto K is *contracting*, meaning that $|p_K(x) - p_K(y)| \leq |x - y|$ for any $x, y \in \mathbb{R}^n$,

(iii) the distance function $\operatorname{dist}(\cdot, K) \colon \mathbb{R}^n \to \mathbb{R}$ is convex, and differentiable at $\mathbb{R}^n \setminus K$, even if K is not *smooth* (a convex body is said to be *smooth* if it has a unique supporting hyperplane at each boundary point),

(iv) the gradient of the distance function can be described in terms of the outer normals of K.

We originally wanted to understand which of these (and further) properties remain true if we consider the metric given by an arbitrary norm in $\mathbb{R}^n$, instead of the usual norm given by the standard inner product. In this paper, we give precise answers to all the cases above except for **(ii)**, describing what properties the norm and/or the convex body have to satisfy such that the desired holds. For normed spaces, property **(ii)** was discussed in [18]. We mention that regularity properties for distance functions have also been extensively investigated in various contexts (see, e.g, [20], [22], [24] and [34]). In the context of normed (or *gauge*) spaces we mention the papers [13], [15], [14], [26] and [33]. However, to our best knowledge the following question was not explicitly answered: can we guarantee differentiability for the distance function to a convex body only assuming that the ambient norm is smooth and strictly convex (in the "geometric" sense that its unit ball is smooth and strictly convex)? We answer this question positively, also relating the so-called *norm gradient* of these functions to the Birkhoff orthogonality relation given by the norm.

Seeing properties **(iii)** and **(iv)** in the context of normed spaces, we observed that the early theory of convex functions on $\mathbb{R}^n$ can be built *replacing the norm given by the standard inner product by a smooth and strictly convex norm.* Besides providing the complete understanding of differentiability of distance functions onto convex sets in normed spaces, we believe that this approach is interesting for itself. Using the Legendre transform of the norm, we define and study the notions of *norm gradients* and *norm subgradients* of convex functions, relating them also to the concept of Birkhoff orthogonality (to our best knowledge, this relation is also new).

We organize the text as follows: in Section 2 we recall some definitions and properties regarding the classical theory of convex functions, briefly discussing also the sub-linear functions. We begin to study metric projections onto convex sets in normed spaces in Section 3, where we give necessary and sufficient conditions for the uniqueness (among other results). Section 4 is devoted to introduce and investigate the so-called norm gradients and norm subgradients of convex functions. They are also used to "detect" differentiability, and this is applied in Section 5 to study the differentiability of distance functions to convex bodies.

To finish this introduction, we fix some basic concepts and notation. We say that a hyperplane $\mathbf{h} \subseteq \mathbb{R}^n$ *supports* a given convex body K at a point $x \in \partial K$ (where ∂K denotes the *boundary of* K) if K is contained in one of the (closed) half-spaces

determined when $\mathbf{h}$ is translated to pass through x. As it was previously mentioned, a convex body is *smooth* if it is supported by a unique hyperplane at each boundary point. We also say that a convex body is *strictly convex* if its boundary contains no line segment. Let $(X, ||\cdot||)$ denote a *normed or Minkowski space*, i.e., a finite dimensional, real Banach space. The *unit ball* of $(X, ||\cdot||)$ is defined as the set $B := \{x \in X : ||x|| \leq 1\}$, and it is immediate that B is a convex body which is symmetric with respect to the origin. The boundary of the unit ball is the *unit sphere* $\partial B = \{x \in X : ||x|| = 1\}$. We say that a norm is *smooth* (resp. *strictly convex*) if its unit ball B is a smooth convex body (resp. strictly convex body). Equivalently, a norm is strictly convex if and only if the triangle inequality is strict for linearly independent vectors (see [21]).

A norm $||\cdot||$ on a vector space X induces an orthogonality relation which is known as *Birkhoff orthogonality*. We say that a vector $x \in X$ is *Birkhoff left-orthogonal* to a vector $y \in X$ if $||x + ty|| \geq ||x||$ for every $t \in \mathbb{R}$. We denote this relation by $x \dashv_B y$, and in this case we also say that y is *Birkhoff right-orthogonal* to x. It is easy to see that Birkhoff orthogonality is a homogeneous relation (meaning it is a relation between directions rather than vectors). Extending this concept, we say that a vector $x \in X$ is *Birkhoff left-orthogonal to a hyperplane* $\mathbf{h}$ if $x \dashv_B z$ for every $z \in \mathbf{h}$. In this case, we also say that $\mathbf{h}$ is Birkhoff right-orthogonal to x, and this is similarly denoted by $x \dashv_B \mathbf{h}$. Birkhoff orthogonality between vectors and hyperplanes is related to the geometry of the unit ball: the norm is smooth if and only if each non-zero vector $x \in X$ admits a unique Birkhoff right-orthogonal hyperplane, and the norm is strictly convex if and only if each hyperplane $\mathbf{h} \subseteq X$ has a unique Birkhoff left-orthogonal direction. We refer the reader to [2] for more information on Birkhoff orthogonality (and other orthogonality types in normed spaces).

2. Convex functions

A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Equivalently, a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its *epigraph*

$$\operatorname{epi}(f) = \{(x,c) \in \mathbb{R}^{n+1} : f(x) \leq c\}$$

is a convex set. Replacing the domain by convex open subsets of $\mathbb{R}^n$, the definition can be extended, as well as the most part of our results. For the sake of simplicity, we work with functions defined over $\mathbb{R}^n$ and which only take values on $\mathbb{R}$ (and not on the extended line $\mathbb{R} \cup \{-\infty, +\infty\}$, as it is usual in the convex analysis literature).

We deal with two differentiability notions for convex functions. First, if $f\colon \mathbb{R}^n \to \mathbb{R}$ is a convex function, then we denote by f'_+ the *one-sided directional derivative*

$$f'_+(x,v) := \lim_{t \to 0^+} \frac{f(x+tv) - f(x)}{t},$$

for any $x, v \in \mathbb{R}^n$. In our context (that of convex functions whose domain is $\mathbb{R}^n$), this limit always exists. For proofs we refer the reader to [29, Chapter 1].

Letting $t \to 0^-$ in the limit above, we get the *left-sided derivative* $f'_-(x,v)$. It is not difficult to see that $f'_-(x,v) = -f'_+(x,-v)$, and for that reason we work mainly with f'_+.

A function $f\colon U \to \mathbb{R}$, where $U \subseteq \mathbb{R}^n$ is an open set, is said to be (*Fréchet*) *differentiable* at $x \in U$ if there exists a linear map $df_x\colon \mathbb{R}^n \to \mathbb{R}$ such that

$$\lim_{||v||\to 0} \frac{|f(x+v) - f(x) - df_x(v)|}{||v||} = 0,$$

where the limit is taken in the metric given by the norm. It is immediate to notice that this definition coincides with the differentiability definition given by the Euclidean norm. For convex functions, there are several characterizations of Fréchet differentiability, and next we state the ones which are important for us.

Proposition 2.1. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function, and let $x \in \mathbb{R}^n$. For simplicity, we write $c := f(x)$. The following statements are equivalent:*

(i) *f is Frechét differentiable at x,*

(ii) *the one-sided derivative $f'_+(x,\cdot)$ is a linear map, and*

(iii) $\mathrm{epi}(f)$ *is supported by a unique hyperplane at (x,c).*

In this case, we have that $f'_+(x,\cdot) = df_x(\cdot)$.

Proof. Again, for the proofs we refer to [29, Chapter 1]. □

At the points where f is not differentiable, we still can guarantee that the one-sided derivative $f'_+(x,\cdot)\colon \mathbb{R}^n \to \mathbb{R}$ is *sub-linear*, meaning that it has the following properties:

(i) (sub-additivity) $f'_+(x,u+v) \le f'_+(x,u) + f'_+(x,v)$ for every $u, v \in \mathbb{R}^n$, and

(ii) (positive homogeneity) $f'_+(x,\lambda u) = \lambda f'_+(x,u)$ for any $u \in \mathbb{R}^n$ and every $\lambda \ge 0$.

The reader may carefully notice that sub-linear functions are, in particular, convex. Hence, if $g\colon \mathbb{R}^n \to \mathbb{R}$ is a sub-linear function, then it has well-defined one-sided derivatives $g'_+(x,u)$ for any $x, u \in \mathbb{R}$, and the maps $g'_+(x,\cdot)$ are also sub-linear. A *linearity direction* of a sub-linear function g is a vector $u \in \mathbb{R}^n$ for which $g(u) = -g(-u)$. The set of linearity directions of g is denoted by $\mathrm{lin}(g)$.

The next lemma establishes some important properties of sub-linear functions that we will use later. The proof is given in [29, Chapter 1].

Lemma 2.2. *Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a sub-linear function. The following statements hold:*

(a) $g'_+(x,\lambda x) = \lambda g(x)$ *for any $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$,*

(b) $g'_+(x,u) \le g(u)$ *for any $x, u \in \mathbb{R}^n$,*

(c) $\mathrm{lin}(g)$ *is a vector subspace of $\mathbb{R}^n$,*

(d) $\mathrm{lin}(g'_+(x,\cdot)) \supseteq \mathrm{lin}(g) + \mathrm{span}(x)$ *for every $x \in \mathbb{R}^n$,*

(e) $\mathrm{lin}(g) = \mathbb{R}^n$ *if and only if g is linear.*

If f is a convex function, then it is easy to notice that for each $c \in \mathbb{R}$ the *sub-level set* $\{f \le c\} := \{z \in \mathbb{R}^n : f(z) \le c\}$ is convex. Despite being simple, the next proposition provides important information about the structure of the sub-level sets

of convex functions (namely, a point which is not a global minimizer lies in the boundary of some sub-level set).

Proposition 2.3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function and assume that $f(x) = c$. Then precisely one of the following statements holds:*

(i) $\{f < c\} \neq \emptyset$ *and* $x \in \partial\{f \leq c\}$,

(ii) c *is the global minimum of* f.

Proof. See, e.g., [25]. □

Proposition 2.4. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function, and assume that f is differentiable at $x \in \mathbb{R}^n$. If $f(x) = c$ is not a global minimum of f, then the sub-level set $\{f \leq c\}$ is supported by a unique hyperplane at x.*

Proof. Let $\mathbf{h}$ be a supporting hyperplane of f at x, and let $z \in \mathbf{h}$. Observe that $f(x + \varepsilon z) \geq f(x)$ for **every** $\varepsilon \in \mathbb{R}$, meaning that f grows in both directions z and $-z$. That is,

$$f'_+(x, \pm z) \geq 0.$$

On the other hand, since f is differentiable at x, we have

$$0 \leq f'_+(x, z) = df_x(z) = f'_-(x, z) = -f'_+(x, -z) \leq 0,$$

from which it follows that $df_x(z) = 0$. This shows that df_x vanishes in any direction which supports $\{f \leq c\}$ at x. Consequently, if $\{f \leq c\}$ has more than one supporting hyperplane at x, then $df_x = 0$, and this contradicts the fact that $f(x) = c$ is not a global minimum. □

Remark 2.5. In convex analysis, this result is usually stated in the following way: if $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable at x, and $df_x \neq 0$, then the normal cone of $\{f \leq f(x)\}$ at x is $\mathbb{R}_+ \nabla f(x)$.

3. The metric projection

Roughly speaking, the *metric projection* of a point x onto a convex body K is the point of ∂K where the minimum distance from x to points from K is attained. For the Euclidean case, this concept is studied in [29, Chapter 1], and in this section we extend the ideas presented there to Minkowski spaces. If $(X, ||\cdot||)$ is a normed space, then the *distance* from a point $x \in X$ to a set $C \subseteq X$ is defined as

$$\operatorname{dist}(x, C) = \inf\{||x - y|| : y \in C\}.$$

We are primarily interested in studying metric projections of points onto convex bodies. The reader must be aware of the fact that, in general, strict convexity and smoothness are not assumed, neither for the norm nor for the considered convex body.

Theorem 3.1. *Let $(X, ||\cdot||)$ be a finite-dimensional vector space, and let $K \subseteq X$ be a convex body. For each point $x \in X$, there exists a point $p_K(x) \in K$ such that*

$$\operatorname{dist}(x, K) = ||x - p_K(x)|| \leq ||x - y||$$

for any $y \in K$. Moreover, if $x \notin K$, then $p_K(x) \in \partial K$, and the vector $x - p_K(x)$ is Birkhoff (left-)orthogonal to some supporting hyperplane of K at $p_K(x)$.

Remark 3.2. The point $p_K(x)$ is called a *metric projection* of x onto K. Of course, if $x \in K$, then $p_K(x) = x$.

Proof. All the claims are well-known, except for the last one. Assume that $x \notin K$ and write $\rho := \operatorname{dist}(x, K) = ||x - p_K(x)||$. Then K and $B(x, \rho)$ are convex bodies which clearly intersect along their boundaries, that is,

$$K \cap B(x, \rho) \subseteq \partial K \cap \partial B(x, \rho).$$

In fact, if this does not hold, then we would have $\operatorname{dist}(x, K) < \rho$. It follows that $\operatorname{int}(K) \cap \operatorname{int}(B(x, \rho)) = \emptyset$, and hence $\operatorname{int}(K)$ and $\operatorname{int}(B(x, \rho))$ can be (properly) separated by some hyperplane $\mathbf{h}$ (see [29, Theorem 1.3.8]). Since we clearly have that $p_K(x) \in \mathbf{h}$, we get that $\mathbf{h}$ supports both $B(x, \rho)$ and K at $p_K(x)$. The fact that $\mathbf{h}$ supports $B(x, \rho)$ at $p_K(x)$ gives that $x - p_K(x)$ is Birkhoff left-orthogonal to $\mathbf{h}$ (translated to pass through the origin). □

From now on, we sometimes will denote the distance function $\operatorname{dist}(\cdot, C)\colon X \to \mathbb{R}$ to a given non-empty compact set C by $d_C(\cdot)$. In general, we work with convex bodies, but some results are still true if we only demand compactness. It is well-known that the distance function d_K of a convex body $K \subseteq X$ is a convex function, and that the distance function d_C of a non-empty compact set $C \subseteq X$ is a *weak contraction*, meaning that

$$|d_C(x) - d_C(y)| \leq ||x - y|| \quad \text{for any } x, y \in C.$$

Next we discuss uniqueness of the metric projection of a given point. As it is explained now, we have no uniqueness only in the case that B_X is not strictly convex and ∂K contains a line segment which is parallel to some line segment of ∂B_X. The metric projection onto K is said to be *well-defined* if it is unique for each $x \in X$.

Proposition 3.3. *If $(X, ||\cdot||)$ is a normed space and $K \subseteq X$ is a convex body, then a metric projection $p_K(x)$ of an exterior point $x \notin K$ is not unique if and only if $p_K(x)$ is contained in some non-degenerate line segment of ∂K which is parallel to some non-degenerate line segment in the boundary ∂B_X of the unit ball.*

Proof. Assume that there are two distinct points $y_0, z_0 \in \partial K$ where the distance from x to K is attained, that is, with

$$||x - y_0|| = ||x - z_0|| = \operatorname{dist}(x, K) := \rho.$$

Hence $y_0, z_0 \in B(x, \rho) \cap K$. Since $B(x, \rho) \cap K \subseteq \partial K \cap \partial B(x, \rho)$, we have two distinct points in the intersection of the boundaries of two convex bodies. This is only possible if both boundaries contain the segment $\operatorname{seg}[y_0, z_0]$ connecting these points.

It follows that $\operatorname{seg}[y_0, z_0]$ is a non-degenerate line segment of ∂K. Since $\operatorname{seg}[y_0, z_0]$ is also a line segment in the boundary of $B(x, \rho)$, it corresponds to a (parallel) line segment in ∂B_X. □

The following known result can be seen as a consequence of Proposition 3.3.

Corollary 3.4. *If $(X, ||\cdot||)$ is a strictly convex normed space, then the metric projection on a given convex body is unique for any point $x \in X$.*

Let $z \in \partial K$. An *outer normal* of K at z is any outward pointing unit vector which is Birkhoff left-orthogonal to some supporting hyperplane of K at z.

Proposition 3.5. *Let $x \notin K$, and let $p_K(x)$ be a metric projection of x onto K. Then the outward pointing unit vector*

$$\eta_K(x) := \frac{x - p_K(x)}{\operatorname{dist}(x, K)}$$

is an outer normal of K at $p_K(x)$.

Proof. The fact that $\eta_K(x)$ is Birkhoff left-orthogonal to some supporting hyperplane of K at $p_K(x)$ comes from Theorem 3.1. Since $\eta_K(x)$ is clearly unit and outward pointing, we get that $\eta_K(x)$ is an outer normal of K. □

In what follows, we denote the line passing through z in the direction of u as $z+\mathbb{R}u$. The ray starting at z with the direction of u will be denoted by $z + \mathbb{R}^+u$. Next we investigate the "converse" of the idea that the (normalized) vector subtraction between an exterior point and its metric projection is an outer normal. The following proposition extends a result, well-known for the Euclidean subcase, to finite-dimensional normed spaces, i.e., Euclidean geometry to Minkowski geometry. The corresponding orthogonality type is the known Birkhoff orthogonality, mostly common in Minkowski spaces; see Sections 3.2 and 3.5 of [30] and Section 6 of [21].

Proposition 3.6. *Let K be any convex body in X, and let $x \in X \setminus K$. Then $p_K(x)$ is a metric projection for any $y \in p_K(x) + \mathbb{R}^+\eta_K(x)$.*

Proof. We write $\rho := \operatorname{dist}(x, K)$. By Theorem 3.1, let $\mathbf{h}$ be a hyperplane which supports both K and $B(x, \rho)$ at $p_K(x)$. If $y \in p_K(x) + \mathbb{R}^+\eta_K(x)$, then $y - p_K(x)$ has the direction of $x - p_K(x)$, and hence it is Birkhoff (left-)orthogonal to $\mathbf{h}$. Writing $\alpha := ||y - p_K(x)||$, it follows that $\mathbf{h}$ supports $B(y, \alpha)$ at $p_K(x)$. Since $B(y, \alpha)$ clearly lies in the same half-space determined by $\mathbf{h}$ as $B(x, \rho)$, we get that $K \cap \operatorname{int}(B(y, \alpha)) = \emptyset$. Consequently, we have that $||y - z|| \geq \alpha$ for all $z \in K$, and hence $p_K(x)$ is a metric projection of y onto K (because $||y - p_K(x)|| = \alpha$). □

Remark 3.7. A set with the property proved in this proposition is often called a *sun* (see, e.g., [8, 16, 28] and references therein).

Lemma 3.8. *If $z \in \partial K$, then any point in the ray starting at z and going into the direction of an outer normal of K at z has z as a metric projection.*

Proof. Let $z \in \partial K$ and assume that u is an outer normal of K at z. Let $\mathbf{h}$ be a supporting hyperplane of K at z which is Birkhoff (right-)orthogonal to u. If $w = z + \beta u$ for some $\beta > 0$, then $z \in \partial B(w, \beta)$ and $w - z$ has the direction of u, from where $\mathbf{h}$ supports $B(w, \beta)$ at z. Clearly, K and $B(w, \beta)$ lie in distinct half-spaces determined by $\mathbf{h}$, from where we get that $K \cap \operatorname{int}(B(w, \rho)) = \emptyset$. It follows that $||w - y|| \geq \beta$ for every $y \in K$, and this shows that z is a metric projection of w onto K, because $||w - z|| = \beta$. □

We have a geometric consequence that will be useful to understand the regularity of the distance function. A *parallel set* (or *parallel body*) of a given convex body K is a set of the type

$$K + \delta B := \{x + \delta y : x \in K,\ y \in B\} = \{x \in X : \operatorname{dist}(x, K) \leq \delta\}, \quad \text{for } \delta > 0.$$

The equality above is immediate, and we will skip the proof. Also, it is easy to see that each parallel set $K + \delta B$ is a convex body.

If we assume that B is smooth (in the sense that it has a unique supporting hyperplane at each boundary point), then $K + \delta B$ is smooth for any $\delta > 0$, even if K is not smooth (see, e.g., [19]).

Proposition 3.9. *Let K be a convex body, and let $z \in \partial K$ be a boundary point. Let $u_K(z)$ be an outer normal vector of K at z, and let $\delta > 0$. If $x = z + \delta u_K(z)$, then $u_K(z)$ is a Birkhoff normal vector of the parallel body $K + \delta B$ at x.*

Proof. Denote by $\mathbf{h}$ the hyperplane which is Birkhoff right-orthogonal to the vector $u_K(z)$. It suffices to prove that $\mathbf{h}$ supports $K + \delta B$ at x. Abusing a little of the notation, we assume that $\mathbf{h}$ is translated to pass through x. Denoting the (closed) half-spaces regarding $\mathbf{h}$ by $\mathbf{h}^+$ and $\mathbf{h}^-$, we may assume, without loss of generality, that $K \subseteq \mathbf{h}^-$. Hence we have that $K + \delta B \subseteq (\mathbf{h} + \delta u_K(z))^-$. Indeed, if $y \in \mathrm{int}(\mathbf{h} + \delta u_K(z))^+$, then $||y - w|| > \delta$ for any $w \in K$, and this, together with the compactness of K, leads to $\mathrm{dist}(y, K) > \delta$, from where we get that $y \notin K + \delta B$. It follows that $\mathbf{h} + \delta u_K(z)$ supports $K + \delta B$ at x, yielding that $u_K(z)$ is a Birkhoff normal vector of $K + \delta B$ at x. □

We know that the distance function to a convex body is convex, and hence it is continuous. However, we said nothing about the continuity of the metric projection (in case it is unique) thus far. This will be clarified next. In the last section of the paper, we will prove that the continuity of the metric projection gives that the derivative of the distance function is continuous (that is, d_K is of class C^1).

Proposition 3.10. *Let $K \subseteq X$ be a convex body, and assume that the metric projection $p_K \colon X \to K$ is well-defined. Then p_K is continuous.*

Proof. It is clear that p_K is continuous in the interior of K. Hence we consider the case where $x \in X$ is not an interior point of K. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence converging to x, and assume that $p_K(x_n)$ does not converge to $p_K(x)$. Since K is compact, it follows that there exists a subsequence $p_K(x_{n_j})$ converging to a point $w \in K$ with $w \neq p_K(x)$. From the continuity of the norm and of the distance function we get

$$||x - w|| = \lim_{j\to\infty} ||x_{n_j} - p_K(x_{n_j})|| = \lim_{j\to\infty} \mathrm{dist}(x_{n_j}, K) = \mathrm{dist}(x, K),$$

from where w is a metric projection of x onto K. This contradicts the uniqueness of the metric projection. Hence $p_K(x_n) \to p_K(x)$, and this shows that p_K is continuous at x. □

4. Norm gradients, and norm subgradients

For simplicity, we work from now on in the vector space $\mathbb{R}^n$; so we can use standard notions of differentiability. Also, we **always** suppose that the norm is smooth and strictly convex; so the Birkhoff orthogonality is "well-behaving". Let $f \colon U \to \mathbb{R}$ be a (Fréchet) differentiable function, where $U \subseteq \mathbb{R}^n$ is an open set. A number $c \in \mathrm{im}(f)$ is said to be a *regular value* of f if df_x is surjective for each $x \in f^{-1}(c)$. It is well known that the pre-image of any regular value is a hypersurface of $\mathbb{R}^n$. The *gradient vector* $\mathrm{grad} f(x)$ of f at $x \in U$ is the (unique) vector such that $df_x(v) = \langle \mathrm{grad} f(x), v\rangle$ for each $v \in \mathbb{R}^n$. We say that f is a *function of class C^1* if $\mathrm{grad} f(x)$ is continuous.

This definition relies on an inner product fixed in $\mathbb{R}^n$, and we would like to have a definition consistent with the geometry given by a (smooth and strictly convex) arbitrary norm.

Definition 4.1. Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Let c be a regular value of f. The *norm gradient* of f at a point $x \in f^{-1}(c)$ is the unique vector $\nabla f(x) \in \mathbb{R}^n$ with the following properties:

(i) $\nabla f(x) \dashv_B T_x(f^{-1}(c))$, and

(ii) $df_x(\nabla f(x)) = ||\nabla f(x)||^2$.

If $c \in \mathrm{im}(f)$ is not a regular value and $x \in f^{-1}(c)$, then we simply put $\nabla f(x) = 0$.

Remark 4.2. From now on we always use the symbol ∇ as a notation for the norm gradient, despite the fact that this notation is often used for the standard Euclidean gradient. This choice is justified because we will not work with the Euclidean gradient throughout the text.

It is not difficult to see that the norm gradient is well-defined. If c is a regular value, then $f^{-1}(c)$ is a hypersurface, and hence it has an $(n-1)$-dimensional tangent hyperplane $T_x(f^{-1}(c))$ at any point $x \in f^{-1}(c)$. This hyperplane supports the unit ball of the norm at a point which determines the Birkhoff left-orthogonal direction to $T_x(f^{-1}(c))$. The condition **(ii)** gives both, the choice of an orientation and a normalization.

This definition makes sense for C^1 functions, since in this case we can define the tangent space of the pre-image of a regular value. However, we can relax this regularity hypothesis when dealing with convex functions, and we are mostly concerned with this case. Based on Proposition 2.3, we define the *norm gradient* of a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ where f is differentiable. If $x \in \mathbb{R}^n$ is such that $c = f(x)$ is not a global minimum, then we take the unique non-zero vector $\nabla f(x)$ such that

(a) $\nabla f(x)$ is Birkhoff left-orthogonal to the (unique) supporting hyperplane of $\{f \leq c\}$ at x,

(b) $df_x(\nabla f(x)) = ||\nabla f(x)||^2$.

Otherwise (that is, if c is a global minimum), we put $\nabla f(x) = 0$.

Denote by $(\mathbb{R}^n)^*$ the dual space of $\mathbb{R}^n$, and assume that $||\cdot||$ is a smooth and strictly convex norm on $\mathbb{R}^n$. The *Legendre transform* on $(\mathbb{R}^n, ||\cdot||)$ is the map $\mathcal{L}\colon \mathbb{R}^n \to (\mathbb{R}^n)^*$ which associates to each non-zero vector $x \in \mathbb{R}^n$ the unique linear functional $\mathcal{L}(x) \in (\mathbb{R}^n)^*$ such that

(a) $\ker\mathcal{L}(x)$ is the (unique) hyperplane which is Birkhoff right-orthogonal to x,

(b) $\mathcal{L}(x) \cdot x = ||x||^2$.

We also define $\mathcal{L}(0) = 0$, that is, the null functional in $(\mathbb{R}^n)^*$. It is immediate to observe that $\mathcal{L}$ is a bijection.

Remark 4.3. Notice very carefully that, for defining the Legendre transform, we do not need the norm to be strictly convex. However, we always assume that this hypothesis is true, because we are more interested in this case.

Remark 4.4. If $|\cdot|$ is the Euclidean norm in $\mathbb{R}^n$ derived from the standard inner product $\langle\cdot,\cdot\rangle$, then it is easy to see that the Legendre transform of $|\cdot|$ is given as $\mathcal{L}(x) = \langle x, \cdot\rangle$. That is, in this case the Legendre transform is the natural isomorphism between $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$ given by the standard inner product.

Lemma 4.5. *The Legendre transform is homogeneous of degree one, that is,* $\mathcal{L}(\alpha x) = \alpha\mathcal{L}(x)$ *for any* $\alpha \in \mathbb{R}$ *and* $x \in \mathbb{R}^n$.

Proof. Of course, we may assume that $\alpha \neq 0$ and $x \neq 0$, since otherwise the proof is trivial. Let $\mathbf{h}$ be the hyperplane such that $x \dashv_B \mathbf{h}$. Then we also have that $\alpha x \dashv_B \mathbf{h}$, meaning that the linear functionals $\mathcal{L}(\alpha x)$ and $\alpha\mathcal{L}(x)$ have the same kernel $\mathbf{h}$. Now we observe that

$$\mathcal{L}(\alpha x)\cdot(\alpha x) = ||\alpha x||^2 = \alpha^2||x||^2 = \alpha\mathcal{L}(x)\cdot(\alpha x),$$

that is, the linear functionals $\mathcal{L}(\alpha x)$ and $\alpha\mathcal{L}(x)$ take the same value at αx. This concludes the proof, since now we have that both linear functionals agree on a basis of $\mathbb{R}^n$. □

We also have that the Legendre transform is a norm-preserving map when one considers the *dual norm* on $(\mathbb{R}^n)^*$. We prove this next.

Proposition 4.6. *Let* $||\cdot||_*$ *be the* dual norm *on* $(\mathbb{R}^n)^*$*, defined as*

$$||\phi||_* := \sup\{\phi(x) : x \in B\}.$$

The Legendre transform is a norm-preserving map of $(\mathbb{R}^n, ||\cdot||)$ *onto* $((\mathbb{R}^n)^*, ||\cdot||_*)$.

Proof. From the homogeneity, it suffices to prove that $||\mathcal{L}(x)||_* = 1$ whenever $||x|| = 1$. Let x be a unit vector, and assume that $\mathbf{h}$ is the hyperplane such that $x \dashv_B \mathbf{h}$. It is easy to see from the definition that $\mathcal{L}(x)\cdot y$ is positive if and only if y lies in the open half-space determined by $\mathbf{h}$ which contains x ($\mathbf{h}^+$, say). Also, $\mathbf{h}$ supports the unit ball at x, and consequently any vector $y \in B \cap \mathbf{h}^+$ can be written as

$$y = \frac{x+z}{||x+z||}$$

for some $z \in \mathbf{h}$. Since $||x+z|| \geq ||x||$, we get

$$\mathcal{L}(x)\cdot y = \mathcal{L}(x)\cdot\frac{x+z}{||x+z||} = \frac{\mathcal{L}(x)\cdot x}{||x+z||} = \frac{||x||^2}{||x+z||} \leq ||x||^2 = 1,$$

and hence $||\mathcal{L}(x)||_* = \sup\{\mathcal{L}(x)\cdot v : v \in B\} \leq 1$ is obtained. On the other hand, we have that $x \in B$ and $\mathcal{L}(x)\cdot x = ||x||^2 = 1$, and this yields $||\mathcal{L}(x)||_* = 1 = ||x||$. □

Our definition of the Legendre transform may be a little bit intrincated to work with, but it has the advantage of not demanding any differentiability properties of the norm (only the geometric features of being smooth and strictly convex). Actually, later our approach will give an easy proof of the fact that norms with these geometric properties must be of type C^1. Even without assuming *a priori* differentiability properties, we can guarantee that the Legendre transform is continuous.

Proposition 4.7. *Let* $||\cdot||$ *be a smooth and strictly convex norm on* $\mathbb{R}^n$*. Then the associated Legendre transform* $\mathcal{L}\colon (\mathbb{R}^n, ||\cdot||) \to ((\mathbb{R}^n)^*, ||\cdot||_*)$ *is continuous.*

Proof. First, observe that continuity at the origin comes immediately from the fact that the Legendre transform is norm-preserving. Also, since the Legendre transform is homogenous of degree 1 it suffices to show that for any sequence $(x_n)_{n\in\mathbb{N}}$ of unit vectors converging to a (unit) vector x we have that $\mathcal{L}(x_n) \to \mathcal{L}(x)$ in the dual norm as $n \to \infty$. We will divide the proof in steps.

Step 1: We prove that for each fixed $v \in \mathbb{R}^n$ we have convergence $\mathcal{L}(x_n)\cdot v \to \mathcal{L}(x)\cdot v$ as $n \to \infty$ (pointwise convergence). Denote by $\mathbf{h}_n$ the supporting hyperplane of B at x_n, and by $\mathbf{h}$ the supporting hyperplane of B at x. For each $n \in \mathbb{N}$, we can write $v = \alpha_n x_n + z_n$ for some $\alpha_n \in \mathbb{R}$ and $z_n \in \mathbf{h}_n$. Similarly, we write $v = \alpha x + z$, with $z \in \mathbf{h}$. We claim that, as $n \to \infty$, we have $\alpha_n \to \alpha$ (and, as a consequence, $z_n \to z$). First, observe

$$||v|| = ||\alpha_n x_n + z_n|| = |\alpha_n| \cdot \left|\left| x_n + \frac{z_n}{\alpha_n} \right|\right| \geq |\alpha_n| \cdot ||x_n|| = |\alpha_n|$$

whenever $\alpha_n \neq 0$, where the last inequality comes from $x_n \dashv_B z_n$. This shows that (α_n) is bounded. Suppose that some subsequence $\alpha_{n_k} \to \beta \neq \alpha$. Then $z_{n_k} \to v - \beta x$, and since Birkhoff orthogonality is a continuous relation, it follows that $x \dashv_B (v - \beta x)$. Therefore, we have two distinct decompositions of v in the direct sum $\mathbf{h} \oplus \text{span}\{v\}$, which is in contradiction to the fact that the norm is smooth. Consequently, every converging subsequence of (α_n) goes to α. This, together with the fact that (α_n) is bounded, yields that $\alpha_n \to \alpha$. Thus, we estimate

$$|\mathcal{L}(x_n) \cdot v - \mathcal{L}(x) \cdot v| = |\mathcal{L}(x_n) \cdot (\alpha_n x_n + z_n) - \mathcal{L}(x) \cdot (\alpha x + z)| = |\alpha_n - \alpha|,$$

and the latter goes to 0 as $n \to \infty$. This concludes the first step.

Step 2: We show that $\mathcal{L}(x_n) \to \mathcal{L}(x)$ in the dual norm up to a subsequence. Consider the sequence of the (continuous) functions $\mathcal{F} = (\mathcal{L}(x_n)|_B)_{n\in\mathbb{N}}$, where $|_B$ means that we are restricting the domain to B. Since $||\mathcal{L}(x_n)||_* = ||x_n|| = 1$ for every $n \in \mathbb{N}$, we have that this family is uniformly bounded. We also have that $\mathcal{F}$ is equicontinuous, because

$$|\mathcal{L}(x_n) \cdot v - \mathcal{L}(x_n) \cdot w| \leq ||\mathcal{L}(x_n)||_* ||v - w|| = ||v - w||,$$

for any $n \in \mathbb{N}$ and every $v, w \in B$. Noticing that B is compact, the Arzela-Ascoli theorem gives that there exists a subsequence $(\mathcal{L}(x_{n_k})|_B)_{k\in\mathbb{N}}$ such that $\mathcal{L}(x_{n_k})|_B \to \phi$ in the dual norm for some continuous function $\phi\colon B \to \mathbb{R}$. Since the convergence in the dual norm implies pointwise convergence, it follows from the first step of the proof that $\phi = \mathcal{L}(x)|_B$. Hence $\mathcal{L}(x_{n_k}) \to \mathcal{L}(x)$ in the dual norm.

Step 3: We prove that $\mathcal{L}(x_n) \to \mathcal{L}(x)$ with respect to $||\cdot||_*$ (that is, we can guarantee the convergence of the original sequence, without needing to pass to a subsequence). We proceed by contradiction. If this convergence does not hold, then there exist a number $\varepsilon > 0$ and a subsequence $(x_{n_j})_{j\in\mathbb{N}}$ such that $||\mathcal{L}(x_{n_j}) - \mathcal{L}(x)||_* > 2\varepsilon$ for every $j \in \mathbb{N}$. Hence there exists a sequence (v_{n_j}) of vectors of B such that

$$|\mathcal{L}(x_{n_j}) \cdot v_{n_j} - \mathcal{L}(x) \cdot v_{n_j}| > \varepsilon.$$

Passing to a subsequence if necessary, we can assume that $v_{n_j} \to v$ for some vector $v \in B$ (recall that B is compact). However, the left-hand side of the inequality above can be made arbitrarily small as $j \to \infty$.

This comes from the estimates

$$\begin{aligned}
&|\mathcal{L}(x_{n_j})\cdot v_{n_j} - \mathcal{L}(x)\cdot v_{n_j}| \\
&\leq |\mathcal{L}(x_{n_j})\cdot v_{n_j} - \mathcal{L}(x_{n_j})\cdot v| + |\mathcal{L}(x_{n_j})\cdot v - \mathcal{L}(x)\cdot v| + |\mathcal{L}(x)\cdot v - \mathcal{L}(x)\cdot v_{n_j}| \\
&\leq ||\mathcal{L}(x_{n_j})||_*||v_{n_j} - v|| + |\mathcal{L}(x_{n_j})\cdot v - \mathcal{L}(x)\cdot v| + ||\mathcal{L}(x)||_*||v - v_{n_j}|| \\
&= 2||v - v_{n_j}|| + |\mathcal{L}(x_{n_j})\cdot v - \mathcal{L}(x)\cdot v|,
\end{aligned}$$

where the reader may observe that, as a consequence of the first step of the proof, the last term converges to 0 as $j \to \infty$. □

Let $((\mathbb{R}^n)^{**}, ||\cdot||_{**})$ be the bi-dual of $\mathbb{R}^n$, which is the dual space of $(\mathbb{R}^n)^*$, endowed with the norm $||\cdot||_{**} := (||\cdot||_*)_*$, that is, the dual norm of $||\cdot||_*$. Let $J\colon \mathbb{R}^n \to (\mathbb{R}^n)^{**}$ be the canonical identification given by $J(x)\cdot\phi = \phi(x)$, for any $x \in \mathbb{R}^n$ and every $\phi \in (\mathbb{R}^n)^*$. It is well-known that J is a norm-preserving isomorphism. Next we discuss the duality of the Legendre transform.

Proposition 4.8. *Denote by $\mathcal{L}^*$ the Legendre transform of $((\mathbb{R}^n)^*, ||\cdot||_*)$ onto $(\mathbb{R}^n)^{**}$, and let $\phi \in (\mathbb{R}^n)^*$. Then $\mathcal{L}^*(\phi) = J(x)$ if and only if $\phi = \mathcal{L}(x)$. In other words, $\mathcal{L}^* = J \circ \mathcal{L}^{-1}$.*

Proof. Assume first that $\mathcal{L}^*(\phi) = J(x)$. Since J is linear, and the Legendre transform is always homogeneous of degree one, we may assume that $||x|| = 1$ (observe also that the case $x = 0$ is trivial). Since $\mathcal{L}^*$ and J are norm-preserving, we have

$$\phi(x) = J(x)\cdot\phi = \mathcal{L}^*(\phi)\cdot\phi = ||\phi||_*^2 = ||\mathcal{L}^*(\phi)||_{**}^2 = ||J(x)||_{**}^2 = ||x||^2 = 1.$$

It follows that $1 = \phi(x) = ||\phi||_* = \sup\{\phi(y) : y \in B\}$, that is, the dual norm of ϕ is attained at x. In particular, we have that $\mathcal{L}(x)\cdot x = \phi(x)$. For any $z \in \ker(\phi)$ the inequality

$$1 \geq \phi\left(\frac{x+tz}{||x+tz||}\right) = \frac{1}{||x+tz||}$$

holds for any $t \in \mathbb{R}$. Thus, $||x+tz|| \geq 1 = ||x||$ for every $t \in \mathbb{R}$, meaning that $x \dashv_B z$. Denoting by $\mathbf{h}$ the hyperplane which supports B at x, it follows that $\ker(\phi) \subseteq \mathbf{h}$. Since both are $(n-1)$-dimensional vector subspaces, we get that $\ker(\phi) = \mathbf{h}$, and this proves that $\phi = \mathcal{L}(x)$.

Now assume that $\mathcal{L}(x) = \phi$, and still consider that $||x|| = 1$. Since the Legendre transform is norm-preserving, we have $||\phi||_* = ||\mathcal{L}(x)||_* = ||x|| = 1$. Also,

$$J(x)\cdot\phi = \phi(x) = \mathcal{L}(x)\cdot x = ||x||^2 = 1.$$

On the other hand we have $\mathcal{L}^*(\phi)\cdot\phi = ||\phi||_*^2 = 1$, implying $J(x)\cdot\phi = \mathcal{L}^*(\phi)\cdot\phi$. It remains to prove that $J(x)$ and $\mathcal{L}^*(\phi)$ have the same kernel. If $\psi \in \ker(J(x))$, then $0 = J(x)\cdot\psi = \psi(x)$, and this leads to

$$||\phi + t\psi||_* \geq \phi(x) + t\psi(x) = \phi(x) = 1 = ||\phi||_* \quad \text{for any } t \in \mathbb{R}.$$

Thus, ψ is a vector of the supporting hyperplane of the dual unit ball B^* at ϕ, and hence $\psi \in \ker(\mathcal{L}^*(\phi))$. It follows that $\ker(J(x)) \subseteq \ker(\mathcal{L}^*(\phi))$, and using again the fact that both are $(n-1)$-dimensional linear subspaces, we obtain that the equality holds. Therefore, $\mathcal{L}^*(\phi) = J(x)$. □

Remark 4.9. What we have proved is that, up to the canonical identification between $\mathbb{R}^n$ and $(\mathbb{R}^n)^{**}$, the Legendre transform is self-dual. Notice that the proof relies heavily on the fact that the kernel of a non-zero functional is precisely the hyperplane which supports the unit ball in the boundary point where its dual norm is attained.

Corollary 4.10. *The inverse of the Legendre transform of a smooth and strictly convex normed space is continuous.*

Proof. Simply observe that $\mathcal{L}^{-1} = J^{-1} \circ \mathcal{L}^*$, and the latter is a composition of continuous maps ($\mathcal{L}^*$ is continuous because it is a Legendre transform). □

Next we give a characterization of the norm gradient of a differentiable convex function.

Theorem 4.11. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function differentiable at a point $x \in \mathbb{R}^n$. Then*

$$f(y) - f(x) \geq \mathcal{L}(\nabla f(x)) \cdot (y - x) \tag{1}$$

for any $y \in \mathbb{R}^n$. The converse is also true: the norm gradient $\nabla f(x)$ is the unique vector for which this inequality holds for every $y \in \mathbb{R}^n$.

Proof. We have to use some machinery from the theory of convex functions. First of all, we state and prove an inequality which "captures" the convexity in terms of one-sided derivatives. We have

$$f(x + w) - f(x) \geq f'_+(x, w) \tag{2}$$

for any $x, w \in \mathbb{R}^n$. To prove this inequality, observe that for any $\varepsilon > 0$ the inequality

$$\begin{aligned} f(x + \varepsilon w) - f(x) &= f(\varepsilon(x + w) + (1 - \varepsilon)x) - f(x) \\ &\leq \varepsilon f(x + w) + (1 - \varepsilon)f(x) - f(x) = \varepsilon(f(x + w) - f(x)) \end{aligned}$$

holds, yielding $f(x + w) - f(x) \geq \dfrac{f(x + \varepsilon w) - f(x)}{\varepsilon}$.

Hence, letting $\varepsilon \to 0^+$ we get (2).

For simplicity of notation, we write $v = \nabla f(x)$ and $c = f(x)$. First assume that c is not a global minimum of f. Then $v \neq 0$, and there is a unique hyperplane $\mathbf{h}$ supporting $\{f \leq c\}$ at x. Consequently, we have the direct sum

$$\mathbb{R}^n = \mathbf{h} \oplus \operatorname{span}\{v\},$$

and since any translation is bijective, we get that any point $y \in \mathbb{R}^n$ can be written in the form $y = \lambda v + x + z$, for some $\lambda \in \mathbb{R}$ and some $z \in \mathbf{h}$. From inequality (2) and the linearity of $f'_+(x, \cdot)$ we get

$$f(y) - f(x) \geq f'_+(x, \lambda v + z) = \lambda f'_+(x, v) + f'_+(x, z). \tag{3}$$

Since $f'_+(x, \cdot)$ is the Fréchet derivative of f at x, we get from the definition of the norm gradient that $f'_+(x, v) = df_x(v) = ||v||^2$.

On the other hand, since $z \in \mathbf{h}$ we have that, for any $\varepsilon \in \mathbb{R}$, the point $x + \varepsilon z$ lies in the hyperplane which supports $\{f \leq c\}$ at x. It follows that $f(x + \varepsilon z) \geq c = f(x)$ for any $\varepsilon > 0$, and this leads to

$$f'_+(x, z) = \lim_{\varepsilon \to 0^+} \frac{f(x + \varepsilon z) - f(x)}{\varepsilon} \geq 0.$$

Plugging this inequality in (3), we get

$$f(y) - f(x) \geq \lambda f'_+(x, v) = \lambda ||v||^2. \tag{4}$$

Now we observe that $\mathcal{L}(v) \cdot (y - x) = \mathcal{L}(v) \cdot (\lambda v + z) = \lambda \mathcal{L}(v) \cdot v + \mathcal{L}(v) \cdot z$.

From the definition of the Legendre transform we have that $\mathcal{L}(v) \cdot z = 0$, since z is a vector of the hyperplane which is Birkhoff right-orthogonal, and $\mathcal{L}(v) \cdot v = ||v||^2$. Consequently, $\mathcal{L}(v) \cdot (y - x) = \lambda ||v||^2$. Together with (4) this gives the desired inequality.

Now we assume that $c = f(x)$ is a global minimum of f, and we have $\nabla f(x) = 0$ by the definition. As a consequence, we get that $f(y) - f(x) \geq 0 = \mathcal{L}(\nabla f(x)) \cdot (y - x)$ for any $y \in \mathbb{R}^n$.

For the converse, suppose first that $c = f(x)$ is not the global minimum of f, and assume that $v \in \mathbb{R}^n$ is such that $f(y) - f(x) \geq \mathcal{L}(v) \cdot (y - x)$. Notice that we must have $v \neq 0$ (otherwise c is a global minimum), and let $\mathbf{h}$ be the hyperplane which is Birkhoff right-orthogonal to v, translated to pass through x. We claim that $\mathbf{h}$ supports the sub-level set $\{f \leq c\}$ at x. Indeed, if this is not true, then we can take a point $y \in \mathbf{h} \cap \operatorname{int}\{f \leq c\} = \mathbf{h} \cap \{f < c\}$. It follows that $v \dashv_B y - x$, which means that $\mathcal{L}(v) \cdot (y - x) = 0$. Thus,

$$f(y) - f(x) < c - f(x) = c - c = 0 = \mathcal{L}(v) \cdot (y - x),$$

and this contradiction shows that $v \dashv_B \mathbf{h}$. It still remains to prove that $df_x(v) = ||v||^2$. First we notice that for any $\lambda > 0$ we have

$$f(x + \lambda x) - f(x) \geq \mathcal{L}(v) \cdot (\lambda v) = \lambda ||v||^2,$$

from which we get

$$df_x(v) = f'_+(x, v) = \lim_{\lambda \to 0^+} \frac{f(x + \lambda v) - f(x)}{\lambda} \geq ||v||^2.$$

For the reverse inequality we observe that with $\lambda > 0$ also

$$f(x - \lambda v) - f(x) \geq \mathcal{L}(v) \cdot (-\lambda v) = -\lambda ||v||^2$$

holds, and hence $||v||^2 \geq \lim_{\lambda \to 0^+} \dfrac{f(x - \lambda v) - f(x)}{-\lambda} = df_x(v)$.

Finally, assume that $c = f(x)$ is a global minimum of f, and let $v \in \mathbb{R}^n$ be a vector such that $f(y) - f(x) \geq \mathcal{L}(v) \cdot (y - x)$ for every $y \in \mathbb{R}^n$. Since c is a global minimum, we have that $df_x = 0$, and then, in particular, $df_x(v) = 0$. For any $\lambda > 0$ we have

$$f(x + \lambda v) - f(x) \geq \mathcal{L}(v) \cdot (\lambda v) = \lambda ||v||^2,$$

and, similarly to what we have done before, we obtain

$$||v||^2 \leq \lim_{\lambda \to 0^+} \frac{f(x+\lambda v) - f(x)}{\lambda} = f'_+(x,v) = df_x(v) = 0,$$

implying also that $v = 0$. The proof is complete. □

Notice carefully that inequality (1) provides an equivalent definition for the norm gradient of a convex function f in a point x where f is differentiable. Inspired by that, we extend this notion for a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ which is not necessarily differentiable. We say that $v \in \mathbb{R}^n$ is a *norm subgradient* of f at x if

$$f(y) - f(x) \geq \mathcal{L}(v) \cdot (y - x) \tag{5}$$

for any $y \in \mathbb{R}^n$. For each $x \in \mathbb{R}^n$, the set $\partial f(x)$ of the norm subgradients of f at x is called the *norm subdifferential* of f at x. It is immediate to check that $\partial f(x)$ is always closed. It is also easy to see that $f(x) = c$ is a global minimum of f if and only if $0 \in \partial f(x)$. At this point, we warn the reader that other generalizations of the concept of subgradient were studied, also in view of differentiability properties (see, e.g., [5] and [10]). As in the Euclidean case, we have a characterization of the norm subgradients in terms of one-sided derivatives (see [6], for example).

Lemma 4.12. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function, and let $x \in \mathbb{R}^n$. We have that $v \in \partial f(x)$ if and only if $f'_+(x,u) \geq \mathcal{L}(v) \cdot u$ for every $u \in \mathbb{R}^n$.*

Proof. Assume first that $v \in \partial f(x)$. Then for each $u \in \mathbb{R}^n$ and any $\lambda > 0$ we have

$$f(x+\lambda u) - f(x) \geq \mathcal{L}(v) \cdot \lambda u = \lambda \mathcal{L}(v) \cdot u.$$

It follows that $\dfrac{f(x+\lambda u) - f(x)}{\lambda} \geq \mathcal{L}(v) \cdot u$, for any $\lambda > 0$. Letting $\lambda \to 0^+$, we have the desired inequality.

Now assume that $v \in \mathbb{R}^n$ is a vector such that $f'_+(x,u) \geq \mathcal{L}(v) \cdot u$. If $y \in \mathbb{R}^n$, then from inequality (2) we get

$$f(y) - f(x) \geq f'_+(x, y-x) \geq \mathcal{L}(v) \cdot (y - x),$$

and this concludes the proof. □

Theorem 4.13. *If $f\colon \mathbb{R}^n \to \mathbb{R}$ is a convex function, then for any $x \in \mathbb{R}^n$ we have that $\partial f(x) \neq \emptyset$ and $f'_+(x,u) = \max\{\mathcal{L}(w) \cdot u : w \in \partial f(x)\}$, for each $u \in \mathbb{R}^n$.*

Proof. By the previous lemma, we already have that the inequality

$$\max\{\mathcal{L}(w) \cdot u : w \in \partial f(x)\} \leq f'_+(x,u)$$

holds for any $u \in \mathbb{R}^n$. Hence we only have to show that for an arbitrarily given vector $u \in \mathbb{R}^n$ there exists a vector $w \in \partial f(x)$ such that $\mathcal{L}(w) \cdot u = f'_+(x,u)$.

To prove that, we first assume that $u \neq 0$. Choose a basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$ with the property that $e_1 = u$. Let $g_0(\cdot) := f'_+(x, \cdot)$, and define recursively functions

$g_1, \ldots, g_n$ by setting $g_m(\cdot) := (g_{m-1})'_+(e_m, \cdot)$. Each of these functions (including g_0) is sub-linear, and from property **(d)** of Lemma 2.2 we have that

$$\operatorname{lin}(g_m) \supseteq \operatorname{lin}(g_{m-1}) + \operatorname{span}(e_m)$$

for each $m = 1, \ldots, n$. From property **(e)** of Lemma 2.2 we have that g_n is a linear functional. Since the Legendre transform is a bijection, there exists $w \in \mathbb{R}^n$ such that $g_n = \mathcal{L}(w)$. We claim that w is a norm subgradient such that $\mathcal{L}(w) \cdot u = f'_+(x, u)$. To check that, we first observe that by property **(b)** of Lemma 2.2 the inequalities $g_0 \geq g_1 \geq \ldots \geq g_n$ hold. This yields

$$\mathcal{L}(w) \cdot (y - x) = g_n(y - x) \leq g_0(y - x) = f'_+(x, y - x) \leq f(y) - f(x),$$

for any $x, y \in \mathbb{R}^n$, where the last inequality comes from (2). This shows that $w \in \partial f(x)$ and, in particular, we also get that $\partial f(x) \neq \emptyset$ (notice that this construction holds true for any $u \neq 0$). Finally, from properties **(a)** and **(b)** of Lemma 2.2 it follows that

$$g_n(u) \leq g_0(u) = -(g_0)'_+(u, -u) = -(g_0)'_+(e_1, -u) = -g_1(-u) \leq -g_n(-u) = g_n(u),$$

from where we get that $\mathcal{L}(w) \cdot u = g_n(u) = g_0(u) = f'_+(x, u)$. This concludes the proof of the case $u \neq 0$. If $u = 0$, then any $w \in \partial f(x)$ satisfies $\mathcal{L}(w) \cdot u = 0 = f'_+(x, u)$. Since we already have that $\partial f(x) \neq \emptyset$, the proof is finished. □

Remark 4.14. This proof appears for the Euclidean sub-case in [6, Theorem 3.1.8]. The reader may notice that the proof constructs a linear functional rather than a subgradient. Only after that, we identify this linear functional with a vector. In the classical theory, this is made via an inner product, and in our case we use the Legendre transform.

The next and important corollary is also an immediate analogue of the Euclidean subcase. It states that we can "detect" differentiability of a convex point at a given interior point of its domain looking at the norm subdifferential at this point.

Corollary 4.15. *A convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable at $x \in \mathbb{R}^n$ if and only if $\partial f(x)$ is a singleton. In this case, the unique norm subgradient is the norm gradient, and we have*

$$df_x(u) = \mathcal{L}(\nabla f(x)) \cdot u \quad \textit{for every } u \in \mathbb{R}^n.$$

Proof. The uniqueness part of Theorem 4.11 already gives that if f is differentiable at x, then $\nabla f(x)$ is the unique norm subgradient of f at x. Hence it remains to prove the converse.

Assume that $\partial f(x) = \{v\}$. Then for any $u \in \mathbb{R}^n$ we have

$$f'_+(x, -u) = \mathcal{L}(v) \cdot (-u) = -\mathcal{L}(w) \cdot u = -f'_+(x, u),$$

meaning that every $u \in \mathbb{R}^n$ is a linearity direction of $f'_+(x, \cdot)$, which is, therefore, a linear map. It follows from Proposition 2.1 that f is differentiable at x.

To finish the proof, just notice that if f is differentiable at $x \in \mathbb{R}^n$, then $\partial f(x) = \{\nabla f(x)\}$, and consequently

$$df_x(u) = f'_+(x, u) = \max\{\mathcal{L}(w) \cdot u : w \in \partial f(x)\} = \mathcal{L}(\nabla f(x)) \cdot u$$

for each $u \in \mathbb{R}^n$. □

Remark 4.16. We have defined functions of class C^1 as differentiable functions whose respective Euclidean gradients are continuous. This is clearly the same as demanding that their respective differential maps are continuous as maps of $\mathbb{R}^n$ onto $(\mathbb{R}^n)^*$. Namely, this follows easily from the fact that the identification between $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$ given by the standard inner product is linear (and hence continuous). Since the Legendre transform and its inverse are also continuous, it follows from the last corollary that a convex differentiable function is of class C^1 if and only if its norm gradient is continuous for any norm.

For the sake of completeness, we state and prove two other properties of the norm subdifferential which are completely analogous to the usual subdifferential. One of them is a way to "detect" convexity via the norm subdifferential, and the other one is a version of Rockafellar's theorem.

Proposition 4.17. *A function $f\colon (\mathbb{R}^n, ||\cdot||) \to \mathbb{R}$ is convex if and only if $\partial f(x) \neq \emptyset$ for every $x \in \mathbb{R}^n$.*

Proof. We already know that if f is convex, then $\partial f(x)$ is non-empty for every $x \in \mathbb{R}^n$ (from Theorem 4.13). Hence we prove the converse. Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$, and observe that we may take a vector $w \in \partial f((1-\lambda)x+\lambda y)$. Therefore,

$$f(x) - f((1-\lambda)x+\lambda y) \geq \mathcal{L}(w)\cdot(x-(1-\lambda)x-\lambda y) = \lambda\mathcal{L}(w)\cdot(x-y), \quad \text{and}$$

$$f(y) - f((1-\lambda x)+\lambda y) \geq \mathcal{L}(w)\cdot(y-(1-\lambda)x-\lambda y) = (1-\lambda)\mathcal{L}(w)\cdot(y-x).$$

Multiplying the first inequality by $(1-\lambda)$, the second by λ, and adding both, we get

$$f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y),$$

and this proves that f is convex. □

In what follows, we define the *norm subdifferential* of a convex function f as the set $\partial f \subseteq \mathbb{R}^n \times \mathbb{R}^n$ given by $\partial f := \{(x,w) \in \mathbb{R}^n \times \mathbb{R}^n : w \in \partial f(x)\}$.

A subset $S \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is said to be *norm cyclically monotonic* if for every $m \in \mathbb{N}$ and any subset $\{(x_0,w_0),\ldots,(x_m,w_m)\} \in S$ the number

$$\mathcal{L}(w_0)\cdot(x_1-x_0)+\mathcal{L}(w_1)\cdot(x_2-x_1)+\ldots+\mathcal{L}(w_{m-1})\cdot(x_m-x_{m-1})+\mathcal{L}(w_m)\cdot(x_0-x_m)$$

is non-positive. If f is a convex function, then it is clear that any finite ordered subset $\{(x_0,w_0)\ldots,(x_m,w_m)\}$ of ∂f is norm cyclically monotonic, because

$$\mathcal{L}(w_0)\cdot(x_1-x_0)+\mathcal{L}(w_1)\cdot(x_2-x_1)+\ldots+\mathcal{L}(w_{m-1})\cdot(x_m-x_{m-1})+\mathcal{L}(w_m)\cdot(x_0-x_m)$$
$$\leq f(x_1)-f(x_0)+f(x_2)-f(x_1)+\ldots+f(x_m)-f(x_{m-1})+f(x_0)-f(x_m)=0.$$

The version of Rockafellar's theorem that we will prove next states, roughly speaking, that any norm cyclically monotonic set is contained in the norm subdifferential of some convex function. But for that sake we need to extend (in the natural way) the definition of convex functions to functions which take values on the extended real line $(-\infty,+\infty]$. A convex function $f\colon \mathbb{R}^n \to (-\infty,+\infty]$ is said to be *proper* if $\{f=\infty\} \neq \mathbb{R}^n$.

Theorem 4.18. *A non-empty set $S \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is norm cyclically monotonic if and only if there exists a proper convex function $f\colon \mathbb{R}^n \to (-\infty, +\infty]$ such that $S \subseteq \partial f$.*

Proof. We already know that if $S \subseteq \partial f$ for some proper convex function f, then S is norm cyclically monotonic. Thus, we prove the converse. Let $S \subseteq \mathbb{R}^n \times \mathbb{R}^n$ be norm cyclically monotonic. We fix $(x_0, w_0) \in S$ and define $f\colon \mathbb{R}^n \to (-\infty, +\infty]$ by

$$f(x) = \sup\{\mathcal{L}(w_m) \cdot (x - x_m) + \mathcal{L}(w_{m-1}) \cdot (x_m - x_{m-1}) + \ldots + \mathcal{L}(w_0) \cdot (x_1 - x_0)\},$$

where the supremum is taken over all values of $m \in \mathbb{N}$ and every possible choice of points $(x_j, w_j) \in S$ for $j = 1, \ldots, m$. The function f defined this way is the supremum of affine functions, and hence it is a convex function (see [29, Chapter 1]). Moreover, since S is norm cyclically montonic, it follows that $f(x_0) = 0$, and then f is proper.

Now let $(x, w) \in S$. We have to prove that w is a norm subgradient of f at x. For this purpose, choose a number $\alpha < f(x)$. Hence there exist pairs $(x_1, w_1), \ldots, (x_m, w_m)$ such that

$$\alpha < \mathcal{L}(w_m) \cdot (x - x_m) + \mathcal{L}(w_{m-1}) \cdot (x_m - x_{m-1}) + \ldots + \mathcal{L}(w_0) \cdot (x_1 - x_0).$$

Putting $(x, w) = (x_{m+1}, w_{m+1})$, it also comes from the definition of f that

$$f(y) \geq \mathcal{L}(w_{m+1}) \cdot (y - x_{m+1}) + \mathcal{L}(w_m) \cdot (x_{m+1} - x_m) + \ldots + \mathcal{L}(w_0) \cdot (x_1 - x_0),$$

for every $y \in \mathbb{R}^n$. These two inequalities yield immediately that

$$f(y) > \alpha + \mathcal{L}(w_{m+1}) \cdot (y - x_{m+1}) = \alpha + \mathcal{L}(w) \cdot (y - x).$$

Since this holds for any $\alpha < f(x)$ and every $y \in \mathbb{R}^n$, we have that

$$f(y) \geq f(x) + \mathcal{L}(w) \cdot (y - x)$$

for each $y \in \mathbb{R}^n$. It follows that $w \in \partial f(x)$, and the proof is done. □

Remark 4.19. The proof of the Euclidean case is given in [29, Theorem 1.5.16].

Here a disclaimer is due, namely on reasons why all of that "works so well" when we only have a (smooth and strictly convex) norm to work with. The idea behind subgradients is to "control" the one-sided derivatives using linear functionals. As it is pointed out in Remark 4.14, in the classical theory these functionals are simply identified with vectors via the standard inner product of $\mathbb{R}^n$. When we have a norm, we still can identify vectors with linear functionals in a way which is coherent with the norm of the domain by using the Legendre transform.

Things get more interesting when we adopt the geometric point of view. In the classical theory, subdifferentials are related to normal cones of sub-level sets. The natural question that arises is whether this is also true in our case, when one replaces the standard normality notion (combined with the standard inner product) by Birkhoff orthogonality. The next theorems are devoted to answer this question.

Theorem 4.20. *Let $x \in \mathbb{R}^n$ and $c = f(x)$. Assume that c is not a global minimum of f, and let $v \in \mathbb{R}^n$ be a norm subgradient of f at x. Then the hyperplane $\mathbf{h}$ which is Birkhoff right-orthogonal to v supports $\{f \leq c\}$ at x. Moreover, we have the estimates*

$$\sup\{f'_-(x, v + z) : z \in \mathbf{h}\} \leq ||v||^2 \leq \inf\{f'_+(x, v + z) : z \in \mathbf{h}\}. \tag{6}$$

Remark 4.21. If $v \neq 0$ and **(i)** holds, then one can easily see that $f'_-(x,v)$ and $f'_+(x,v)$ are positive numbers.

Proof. By abuse of notation, we assume that $\mathbf{h}$ is translated to pass through x. If $\mathbf{h}$ does not support $\{f \leq c\}$ at x, then one may take a point $y \in \mathbf{h} \cap \{f < c\}$. Since $y - x \in \mathbf{h}$ and $v \dashv_B \mathbf{h}$, we have that $\mathcal{L}(v) \cdot (y-x) = 0$. Once $f(y) < c = f(x)$, we get $f(y) - f(x) < 0 = \mathcal{L}(v) \cdot (y - x)$, which contradicts the fact that v is a subgradient of f at x. This proves that $\mathbf{h}$ supports $\{f \leq c\}$ at x.

The estimates come from Lemma 4.12. For any $z \in \mathbf{h}$ we have that

$$f'_+(x, v+z) \geq \mathcal{L}(v) \cdot (v+z) = \mathcal{L}(v) \cdot v = ||v||^2 \quad \text{and}$$

$$f'_-(x, v+z) = -f'_+(x, -v-z) \leq -\mathcal{L}(v) \cdot (-v-z) = ||v||^2,$$

and this concludes the proof. □

The natural question that arises here is whether the converse holds true, namely the following implication: if a vector v is Birkhoff left-orthogonal to some hyperplane which supports $\{f \leq c\}$ at x, and if (6) also holds, is v then a norm subgradient of f? To answer that question (positively), we first need to check whether the estimates (6) really "make sense".

Proposition 4.22. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function, and assume that $f(x) = c$ is not a global minimum. If $\mathbf{h}$ is a supporting hyperplane of $\{f \leq c\}$ at x, and if $v \dashv_B \mathbf{h}$, then $f'_-(x, v+z_1) \leq f'_+(x, v+z_2)$ for any $z_1, z_2 \in \mathbf{h}$. As a consequence, we have that*

$$\sup\{f'_-(x, v+z) : z \in \mathbf{h}\} \leq \inf\{f'_+(x, v+z) : z \in \mathbf{h}\}.$$

Proof. Suppose that there exist $z_1, z_2 \in \mathbf{h}$ such that $f'_-(x, v+z_1) > f'_+(x, v+z_2)$. This yields

$$f'_+(x, z_2 - z_1) \leq f'_+(x, v+z_2) + f'_+(x, -v-z_1) = f'_+(x, v+z_2) - f'_-(x, v+z_1) < 0,$$

and this is a contradiction because $z_2 - z_1 \in \mathbf{h}$. Indeed, since $\mathbf{h}$ supports $\{f \leq c\}$ at x, we have that, at x, f is non-decreasing in the direction of $z_2 - z_1$. □

It follows from Proposition 4.22 that if $\mathbf{h}$ supports $\{f \leq c\}$ at x, then we always can choose a vector $v \dashv_B \mathbf{h}$ for which (6) holds. Next we will prove that such a vector is a norm subgradient. This guarantees the existence of norm subgradients pointing in all Birkhoff outer normal directions of $\{f \leq c\}$ at x.

Theorem 4.23. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. As usual, assume that $f(x) = c$ is not a global minimum, and that the hyperplane $\mathbf{h}$ supports $\{f \leq c\}$ at x. If $v \in \mathbb{R}^n$ is a vector such that*

(i) $v \dashv_B \mathbf{h}$ *and*

(ii) $\sup\{f'_-(x, v+z) : z \in \mathbf{h}\} \leq ||v||^2 \leq \inf\{f'_+(x, v+z) : z \in \mathbf{h}\}$,

then v is a norm subgradient of f at x.

Proof. First observe that, by Remark 4.21, v is non-zero. We have to prove that $f(y) - f(x) \geq \mathcal{L}(v) \cdot (y - x)$. Any $y \in \mathbb{R}^n$ can be written as $y = x + z + \lambda v$ for some $\lambda \in \mathbb{R}$ and some $z \in \mathbf{h}$. Assuming first that $\lambda > 0$, we have

$$\begin{aligned} f(y) - f(x) = f(x + z + \lambda v) - f(x) \geq f'_+(x, z + \lambda v) = \lambda f'_+\Big(x, v + \frac{z}{\lambda}\Big) \geq \lambda ||v||^2 \\ = \mathcal{L}(v) \cdot (\lambda v) = \mathcal{L}(v) \cdot (z + \lambda v) = \mathcal{L}(v) \cdot (y - x), \end{aligned}$$

where the first inequality comes from (2). We also used the definition of the Legendre transform. If $\lambda < 0$, we have

$$\begin{aligned} f(y) - f(x) \geq f'_+(x, z + \lambda v) = -\lambda f'_+\Big(x, -v - \frac{z}{\lambda}\Big) = \lambda f'_-\Big(x, v + \frac{z}{\lambda}\Big) \\ \geq \lambda ||v||^2 = \mathcal{L}(v) \cdot (z + \lambda v) = \mathcal{L}(v) \cdot (y - x). \end{aligned}$$

It only remains to prove the case $\lambda = 0$. To do so, we recall that $f'_+(x, z) \geq 0$, because z is a vector of a supporting hyperplane of $\{f \leq c\}$ at x. Hence

$$f(y) - f(x) = f(x + z) - f(x) \geq f'_+(x, z) \geq 0 = \mathcal{L}(v) \cdot z = \mathcal{L}(v) \cdot (y - x),$$

and the proof is complete. □

Corollary 4.24. *Under the same conditions as in the previous theorem, let u be a unit outward pointing vector Birkhoff orthogonal to a supporting hyperplane $\mathbf{h}$ of $\{f \leq c\}$ at x. If*

$$\sup\{f'_-(x, u + z) : z \in \mathbf{h}\} \leq \lambda \leq \inf\{f'_+(x, u + z) : z \in \mathbf{h}\},$$

then $\lambda u \in \partial f(x)$.

Proof. We have to prove that conditions **(i)** and **(ii)** of Theorem 4.23 hold. The first is obvious, because Birkhoff orthogonality is homogeneous. For the second, recall that $\lambda > 0$ (see Remark 4.21), notice that

$$\lambda \cdot \sup\{f'_-(x, u + z) : z \in \mathbf{h}\} = \sup\{f'_-(x, \lambda u + z) : z \in \mathbf{h}\},$$

and that the same holds for the infimum. Then

$$\sup\{f'_-(x, \lambda v + z) : z \in \mathbf{h}\} \leq \lambda^2 = ||\lambda u||^2 \leq \inf\{f'_+(x, \lambda u + z) : z \in \mathbf{h}\},$$

which is condition **(ii)** of Theorem 4.23. □

As we shall see later, Theorem 4.20 is a key ingredient in proving that a distance function to a convex body in a (smooth) normed space is differentiable outside the body. But this can also be used to give an easy proof to a well-known result regarding the regularity of norms.

Theorem 4.25. *Let $\rho\colon \mathbb{R}^n \to \mathbb{R}$ be a norm. If its unit ball B is smooth, then ρ is C^1 on $\mathbb{R}^n \setminus \{0\}$. Moreover, its norm gradient is given by*

$$\nabla\rho(x) = \frac{x}{\rho(x)}, \quad \textit{for each } x \in \mathbb{R}^n \setminus \{0\}.$$

Proof. First notice that ρ is a sub-linear function, and that hence it is convex. Thus, to show that ρ is differentiable at a given point x, it suffices to prove that the norm subdifferential of ρ at x contains a unique element.

Let $x \in \mathbb{R}^n \setminus \{0\}$ be such that $\rho(x) = c$ (> 0). The sub-level set $\{\rho \leq c\}$ is the ball cB, which by the smoothness hypothesis is supported at x by a unique hyperplane ($\mathbf{h}$, say). Thus, given $v \in \partial\rho(x)$ we have $v \dashv_B \mathbf{h}$, meaning that $v = \beta x$ for some $\beta \in \mathbb{R}$. Since

$$\frac{d}{dt}\rho(x \pm tx)\bigg|_{t=0} = \pm\rho(x),$$

we have that $\rho'_+(x,x) = \rho(x) = -\rho'_+(x,-x)$. With Lemma 4.12 this yields

$$\rho(x) = \rho'_+(x,x) \geq \mathcal{L}(v)\cdot x = \mathcal{L}(\beta x)\cdot x = \beta\rho(x)^2$$

and
$$-\rho(x) = \rho'_+(x,-x) \geq \mathcal{L}(v)\cdot(-x) = -\mathcal{L}(\beta x)\cdot x = -\beta\rho(x)^2.$$

These inequalities give that $\beta = 1/\rho(x)$. It follows that $\partial\rho(x)$ is a singleton, and hence ρ is differentiable. The unique element of $\partial\rho(x)$ is the norm gradient

$$\nabla\rho(x) = \beta x = \frac{x}{\rho(x)},$$

which is clearly continuous. It follows that ρ is C^1 on $\mathbb{R}^n \setminus \{0\}$. □

Remark 4.26. Notice that by the continuity of the Legendre transform we have that if the norm gradient is continuous (as a map onto $\mathbb{R}^n$), then the Euclidean gradient is also continuous. Indeed, the Euclidean gradient is the image of the norm gradient under the composition of the Legendre transform with the isomorphism between $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$ given by the standard inner product.

In some fields (such as Finsler geometry, for example) it is more common to define the Legendre transform by means of the derivative of the norm. However, in the beginning we did not assume differentiability of the norm, but only strict convexity and smoothness. The last theorem states that, under these hypotheses, the norm is indeed differentiable, and hence we can characterize now the Legendre transform by means of the derivative of the norm.

Corollary 4.27. *Let ρ be a smooth norm on $\mathbb{R}^n$. The associated Legendre transform can be written as $\mathcal{L}(x) = \rho(x)\cdot d\rho_x(\cdot)$, for each $x \in \mathbb{R}^n$.*

Proof. If $x = 0$, then there is nothing to prove. Hence we assume that $x \neq 0$ and put $c = \rho(x)$. With B as unit ball of ρ, let $\mathbf{h}$ be the hyperplane such that $x \dashv_B \mathbf{h}$. If $z \in \mathbf{h}$, then we may take a differentiable curve $\gamma(t)\colon I \subseteq \mathbb{R} \to c\cdot\partial B$ such that $\gamma(0) = x$ and $\gamma'(0) = z$. Thus,

$$\rho(x)\cdot d\rho_x(z) = \rho(x)\cdot\frac{d}{dt}\rho\circ\gamma(t)\bigg|_{t=0} = 0 = \mathcal{L}(x)\cdot z,$$

because $\rho\circ\gamma(t) = c$ for every t. Finally, we calculate

$$\rho(x)\cdot d\rho_x(x) = \rho(x)\cdot\frac{d}{dt}\rho(x+tx)\bigg|_{t=0} = \rho(x)^2 = \mathcal{L}(x)\cdot x.$$

□

Remark 4.28. Having the differentiability of the norm (except at the origin, of course) *a priori*, the Legendre transform can be equivalently defined as

$$\mathcal{L}(x) \cdot v := \frac{1}{2} \frac{d}{dt} \rho(x + tv)^2 \Big|_{t=0}, \tag{7}$$

for any $x, v \in \mathbb{R}^n$. This approach was taken in [3], for example. Even more, in [17] the authors study the map $[v, x] := \mathcal{L}(x) \cdot v$, which they call a *semi-inner product*.

The reader experienced in convex analysis will notice that the results of this section, and also those of the next one, are extensions of the Euclidean case to finite-dimensional normed spaces. Using linear functionals to "describe" the notion of orthogonality, and regarding the definition of the Legendre transform as in (7), one can develop this theory in even more general settings. In this direction, we refer the reader to the books [7], [25] and [35]. For our goal here it is important to notice that if we use the widely-known form (7) of the Legendre transform of a norm to identify the "functional" subgradients as vectors, then we obtain the orthogonality in the sense of Birkhoff. Going even further, we see that our first definition of the Legendre transform (which, as common in convexity, only relies on the notion of support) is sufficient to construct a theory of convex functions over normed spaces which "dialogues" with usual concepts of Minkowski geometry (see again [30]).

To finish this section, we relate norm subdifferentials with normal cones of sub-level sets at boundary points. Let K be a convex body, and x be a boundary point. The (*Birkhoff*) *normal cone* of K at x, denoted by $\mathrm{NC}(K, x)$, is the set of all outward pointing vectors which are Birkhoff left-orthogonal to some hyperplane that supports K at x, together with the zero vector (for other normal cones appearing in the geometry of Banach spaces we refer the reader to [27]). Since an outward pointing unit Birkhoff left-orthogonal vector to some supporting hyperplane of K at x is called an outer normal, and since Birkhoff orthogonality is homogeneous, we have

$$\mathrm{NC}(K, x) = \mathbb{R}^+ \cdot \{\text{outer normals of } K \text{ at } x\},$$

where $\mathbb{R}^+ \cdot A := \{\lambda a : \lambda \geq 0 \text{ and } a \in A\}$. A consequence of Theorem 4.20 is that the inclusion

$$\mathbb{R}^+ \cdot \partial f(x) \subseteq \mathrm{NC}(\{f \leq c\}, x)$$

holds whenever $f \colon \mathbb{R}^n \to \mathbb{R}$ is a convex function and $f(x) = c$ is not a global minimum. Under the same hypotheses, Theorem 4.23 gives the reverse inclusion:

$$\mathrm{NC}(\{f \leq c\}, x) \subseteq \mathbb{R}^+ \cdot \partial f(x).$$

As it was mentioned before, the norm subdifferential is the pull-back of a (convex) set of linear functionals in $(\mathbb{R}^n)^*$ by the Legendre transform:

$$\partial f(x) = \mathcal{L}^{-1}\Big(\{\phi \in (\mathbb{R}^n)^* : f(y) - f(x) \geq \phi(y - x) \ \ \forall\, y \in \mathbb{R}^n\}\Big),$$

and the same holds for the Birkhoff normal cone, as the next lemma shows.

Lemma 4.29. *Let $K \subseteq \mathbb{R}^n$ be a convex body, and let $x \in \partial K$ be a boundary point.*

Then $$\mathrm{NC}(K, x) = \mathcal{L}^{-1}\Big(\{\phi \in (\mathbb{R}^n)^* : \phi(y - x) \leq 0 \ \ \forall\, y \in K\}\Big).$$

Proof. Let $v \in \mathrm{NC}(K,x)$. Of course, there is nothing to prove for the case $v = 0$, and hence we may assume that v is a non-zero vector. If $\mathbf{h}$ is the hyperplane such that $v \dashv_B \mathbf{h}$, then $\mathbf{h}$ supports K at x. Denote $\phi = \mathcal{L}(v)$. If $y \in K$, then y does not lie in the same half-space determined by $\mathbf{h}$ as v (recall that v is an outer normal), and hence we may write $y = \alpha v + z$, for some $z \in \mathbf{h}$ and some $\alpha \leq 0$. Thus,

$$\phi(y-x) = \mathcal{L}(v)\cdot(y-x) = \mathcal{L}(v)\cdot(\alpha v + z) = \alpha||v||^2 \leq 0,$$

and this shows the inclusion "$\subseteq$". Now assume that ϕ is a non-zero functional and observe that if $\phi(y-x) \leq 0$ for every $y \in K$, then $\mathbf{h} := \ker(\phi)$ supports K at x. Then, if v is a unit and outward pointing vector such that $v \dashv_B \mathbf{h}$, it follows that $\phi = \mathcal{L}(\lambda v)$ for some $\lambda > 0$. This gives the reverse inclusion. □

Remark 4.30. The "functional versions" of normal cones and subdifferentials can be used to investigate the case where the norms are not smooth or strictly convex. In this direction, we refer the reader to [23].

5. Differentiability of distance functions

Throughout this section we will **always** assume that the involved norms are smooth and strictly convex. Assume that $\mathbb{R}^n$ is endowed with such a norm, and let $K \subseteq \mathbb{R}^n$ be a convex body which is not necessarily smooth or strictly convex. For a given $x \in \mathbb{R}^n \setminus K$ the metric projection $p_K(x) \in \partial K$ is unique, and hence the outer normal $\eta_K(x)$ defined as

$$\eta_K(x) := \frac{x - p_K(x)}{||x - p_K(x)||}$$

is unique. Denote by d_K the distance function to K, defined as $d_K(x) := \mathrm{dist}(x,K)$.

In the next theorem we discuss the differentiability of d_K. In the same "spirit" as commented above, after Remark 4.28, the following results are counterparts of well-known results from convex analysis, obtained here in terms of the geometry of normed spaces. Thus, everything will be described in terms of Birkhoff orthogonality (cf. [30] and [21]).

Theorem 5.1. *The function d_K is differentiable in $\mathbb{R}^n \setminus K$. Moreover, we have*

$$\nabla d_K(x) = \eta_K(x) \quad \textit{for any } x \in \mathbb{R}^n \setminus K.$$

Proof. First observe that, for each $c > 0$, the sub-level set $\{d_K \leq c\}$ is precisely the convex body $K + cB$. Assume that $d_K(x) = c$, and let $\mathbf{h}$ be the (unique) supporting hyperplane of $K + cB$ at x. Assume that $v \in \partial d_K(x)$. Due to Theorem 4.20, the hyperplane which is right-orthogonal to v supports $K + cB$ at x, and hence $v \dashv_B \mathbf{h}$. From Propositions 3.5 and 3.9 we get that v is a multiple of $\eta_K(x)$. It follows that $\partial d_K(x) \subseteq \mathrm{span}\{\eta_K(x)\}$. From Proposition 3.6 we have that

$$d_K(x + t\eta_K(x)) = d_K(x) + t,$$

for $t \in \mathbb{R}$ small enough. From that equality we get immediately that

$$(d_K)'_+(x, \eta_K(x)) = 1 = -(d_K)'_+(x, -\eta_K(x)).$$

Now let $\alpha\eta_K(x) \in \partial d_K(x)$. Lemma 4.12 implies that

$$1 = (d_K)'_+(x, \eta_K(x)) \geq \mathcal{L}(\alpha\eta_K(x))\cdot\eta_K(x) = \alpha||\eta_K(x)||^2 = \alpha.$$

Finally, we have

$$-1 = (d_K)'_+(x, -\eta_K(x)) \geq \mathcal{L}(\alpha\eta_K(x)) \cdot (-\eta_K(x)) = -\alpha||\eta_K(x)||^2 = -\alpha.$$

It follows that $\alpha = 1$. Hence $\partial d_K(x) = \{\eta_K(x)\}$, that is, the norm subdifferential of d_K at x is a singleton. From Corollary 4.15 we get that d_K is differentiable at x, and $\nabla d_K(x) = \eta_K(x)$. □

Corollary 5.2. *The norm gradient ∇d_K of the distance function d_K is continuous in $\mathbb{R}^n \setminus K$. In particular, d_K is a function of class C^1.*

Proof. This follows immediately from the equality $x = p_K(x) + \operatorname{dist}(x, K) \cdot \eta_K(x)$, which holds for any $x \in \mathbb{R}^n \setminus K$. Since the distance function $\operatorname{dist}(\cdot, K)$ and the metric projection p_K are continuous functions in $\mathbb{R}^n \setminus K$ (see Proposition 3.10) we get that $\eta_K(x)$ is continuous in $\mathbb{R}^n \setminus K$. Consequently, ∇d_K is continuous in $\mathbb{R}^n \setminus K$, and from Remark 4.16 we get that d_K is a function of class C^1 in $\mathbb{R}^n \setminus K$. □

Observe that in the interior of K we clearly have that d_K is differentiable, and $\nabla d_K(x) = \{0\}$. In fact, one just has to recall that d_K is constant in K. Next we investigate what happens on the boundary of K. We show that d_K is not differentiable on ∂K, and characterize its norm subdifferential at these points.

Theorem 5.3. *For each $x \in \partial K$ we have $\partial d_K(x) = \mathrm{NC}(K, x)$. In particular, d_K is not differentiable at a boundary point of K.*

Proof. The fact that $0 \in \partial d_K(x)$ follows from the property that $d_K(x) = 0$ is the global minimum of d_K. Now let v be a (unit) outer normal of K at x, and assume that $\mathbf{h}$ is the supporting hyperplane of K at x such that $v \dashv_B \mathbf{h}$. Since $d_K(x+tv) = t$ for $t \geq 0$, we have that $(d_K)'_+(x, v) = 1$. If $\theta \in (0, +\infty)$, then we will prove that $\theta v \in \partial d_K(x)$. Let $y \in \mathbb{R}^n$, and write $y = \lambda\theta v + z + x$ for some $z \in \mathbf{h}$ and some $\lambda \in \mathbb{R}$. First, assume that $\lambda > 0$. From (2) we have

$$d_K(\lambda\theta v + z + x) - d_K(x) \geq (d_K)'_+(x, \lambda\theta v + z).$$

Also, we claim that $(d_K)'_+(x, \lambda\theta v + z) \geq (d_K)'_+(x, \lambda\theta v)$. Indeed, observe that we have $d_K(x + tv) = t$ for any $t > 0$, and that Proposition 3.9 implies that z is a supporting direction of $\{d_K \leq t\}$ at $x + tv$. Hence $d_K(x + \alpha v + \beta z) \geq d_K(x + \alpha v)$ for any $\alpha > 0$ and $\beta \in \mathbb{R}$. Consequently,

$$\begin{aligned}(d_K)'_+(x, \lambda\theta v + z) &= \lim_{\varepsilon \to 0^+} \frac{d_K(x + \varepsilon(\lambda\theta v + z)) - d_K(x)}{\varepsilon} \\ &\geq \lim_{\varepsilon \to 0^+} \frac{d_K(x + \varepsilon\lambda\theta v) - d_K(x)}{\varepsilon} = (d_K)'_+(x, \lambda\theta v).\end{aligned}$$

Since $(d_K)'_+(x, \lambda\theta v) = \lambda\theta (d_K)'_+(x, v) = \lambda\theta$, we get

$$d_K(y) - d_K(x) \geq (d_K)'_+(x, \lambda\theta v + z) \geq \lambda\theta.$$

On the other hand, we have that

$$\mathcal{L}(\theta v) \cdot (y - x) = \mathcal{L}(\theta v) \cdot (\lambda v + z) = \lambda\theta||v||^2 = \lambda\theta,$$

and this concludes the case $\lambda > 0$. If $\lambda \leq 0$, then we recall again that $d_K(x) = 0$ is the global minimum of d_K, and write

$$\mathcal{L}(\theta v) \cdot (y - x) = \mathcal{L}(\theta v) \cdot (\lambda v + z) = \lambda\theta||v||^2 = \lambda\theta \leq 0 \leq d_K(y) - d_K(x).$$

This shows that $\theta v \in \partial d_K(x)$. Hence we have the inclusion $\partial d_K(x) \supseteq \mathrm{NC}(K,x)$. Now assume that $v \in \partial d_K(x)$. For any $y \in K$ we have

$$0 = d_K(y) - d_K(x) \geq \mathcal{L}(v) \cdot (y - x).$$

Consequently, we get from Lemma 4.29 that $v \in \mathrm{NC}(K,x)$. This gives the remaining inclusion $\partial d_K(x) \subseteq \mathrm{NC}(K,x)$. □

Acknowledgments. The authors wish to thank the anonymous referee for valuable hints to improve the manuscript.

References

[1] A. R. Alimov, I. G. Tsar'kov: *Connectedness and other geometric properties of suns and Chebyshev sets (Russian)*, Fundam. Prikl. Mat. 19/4 (2014) 21–91; English translation in J. Math. Sci. (N.Y.) 217/6 (2016) 683–730.

[2] J. Alonso, H. Martini, S. Wu: *On Birkhoff orthogonality and isosceles orthogonality in normed linear spaces*, Aequationes Math. 83/1-2 (2012) 153–189.

[3] J. C. Álvarez-Paiva, A. C. Thompson: *Volumes in normed and Finsler spaces*, in: *A Sampler of Riemann-Finsler Geometry*, D. Bao, R. Bryant, S. S. Chern, Z. Shen (eds.), Cambridge University Press, Cambridge (2004) 1–49.

[4] E. Asplund: *Differentiability of the metric projection in finite-dimensional Euclidean space*, Proc. Amer. Math. Soc. 38 (1973) 218–219.

[5] J. M. Borwein, S. P. Fitzpatrick, J. R. Giles: *The differentiability of real functions on normed linear space using generalized subgradients*, J. Math. Analysis Appl. 128/2 (1987) 512–534.

[6] J. M. Borwein, A. S. Lewis: *Convex Analysis and Nonlinear Optimization: Theory and Examples*, 2nd edition, CMS Books in Mathematics 3, Canadian Mathematical Society, Ottawa (2006).

[7] J. M. Borwein, J. D. Vanderwerff: *Convex Functions: Constructions, Characterizations and Counterexamples*, Encyclopedia of Mathematics and Its Applications 109, Cambridge University Press, Cambridge (2010).

[8] A. L. Brown: *Suns in normed linear spaces which are finite-dimensional*, Math. Annalen 279/1 (1987) 87–101.

[9] L. Bunt: *Bijdrage tot de theorie der konvekse puntverzamelingen*, Thesis, University of Groningen, Amsterdam (1934).

[10] F. H. Clarke: *Optimization and Nonsmooth Analysis*, Classics in Applied Mathematics 5, Society for Industrial and Applied Mathematics, Philadelphia (1990).

[11] R. Correa, D. Salas, L. Thibault: *Smoothness of the metric projection onto nonconvex bodies in Hilbert spaces*, J. Math. Analysis Appl. 457/2 (2018) 1307–1332.

[12] F. Deutsch: *The convexity of Chebyshev sets in Hilbert space*, in: *Topics in Polynomials of One and Several Variables and Applications*, World Scientific Publishing, New York (1993) 143–150.

[13] S. Fitzpatrick: *Metric projections and the differentiability of distance functions*, Bull. Aust. Math. Soc. 22 (1980) 291–312.

[14] J. R. Giles: *Differentiability of distance functions and a proximinal property inducing convexity*, Proc. Amer. Math. Soc. 104/2 (1988) 458–464.

[15] J. R. Giles: *A distance function property implying differentiability*, Bull. Aust. Math. Soc. 39/1 (1989) 59–70.

[16] L. Hetzelt: *On suns and cosuns in finite-dimensional normed real vector spaces*, Acta Math. Hungar. 45/1-2 (1985) 53–68.

[17] Á. G. Horváth, Z. Lángi, M. Spirova: *Semi-inner products and the concept of semi-polarity*, Results Math. 71/1-2 (2017) 127–144.

[18] L. A. Karlovitz: *The construction and application of contractive retractions in 2-dimensional normed linear spaces*, Indiana Univ. Math. J. 22/5 (1972) 473–481.

[19] S. G. Krantz, H. R. Parks: *On the vector sum of two convex sets in space*, Canad. J. Math. 43/2 (1991) 347–355.

[20] Y. Li, L. Nirenberg: *Regularity of the distance function to the boundary*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5) 29/1 (2005) 257–264.

[21] H. Martini, K. J. Swanepoel, G. Weiss,: *The geometry of Minkowski spaces – a survey. Part I*, Expositiones Math. 19 (2001) 97–142.

[22] D. Noll: *Directional differentiability of the metric projection in Hilbert space*, Pacific J. Math. 170/2 (1995) 567–592.

[23] J. P. Penot, R. Ratsimahalo: *Characterizations of metric projections in Banach spaces and applications*, Abstr. Appl. Analysis 3/1-2 (1998) 85–103.

[24] R. A. Poliquin, R. T. Rockafellar, L. Thibault: *Local differentiability of distance functions*, Trans. Amer. Math. Soc. 352/11 (2000) 5231–5249.

[25] R. T. Rockafellar: *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, Princeton University Press, Princeton (1970).

[26] M. Safdari: *The distance function from the boundary of a domain with corners*, Nonlinear Analysis 181 (2019) 294–310.

[27] D. Sain, K. Paul, A. Mal: *On approximate Birkhoff-James orthogonality and normal cones in a normed space*, J. Convex Analysis 26/1 (2019) 341–351.

[28] Sangeeta, T. D. Narang: *On suns in linear metric spaces and convex metric spaces*, East J. Approx. 17/2 (2011) 127–139.

[29] R. Schneider: *Convex Bodies: The Brunn-Minkowski Theory*, Encyclopedia of Mathematics and its Applications 151, Cambridge University Press, Cambridge (2014).

[30] A. C. Thompson: *Minkowski Geometry*, Encyclopedia of Mathematics and Its Applications 63, Cambridge University Press, Cambridge (1996).

[31] L. P. Vlasov: *On almost convex sets in Banach spaces (Russian)*, Dokl. Akad. Nauk SSSR 163 (1965) 18–21.

[32] L. P. Vlasov: *Approximatively convex sets in uniformly smooth spaces*, Mat. Zametki 1 (1967) 443–450.

[33] L. Zajíček: *Differentiability of the distance function and points of multivaluedness of the metric projection in Banach space*, Czechoslovak Math. J. 33/2 (1983) 292–308.

[34] L. Zajíček: *On the Fréchet differentiability of distance functions*, in: *Proceedings of the 12th Winter School on Abstract Analysis*, Srní 1984, Rend. Circ. Mat. Palermo 2, Suppl. No. 5 (1984) 161–165.

[35] C. Zălinescu: *Convex Analysis in General Vector Spaces*, World Scientific Publishing, River Edge (2002).

[36] Z. Wu: *A Chebyshev set and its distance function*, J. Approximation Theory 119/2 (2002) 181–192.