On the Differentiability Properties of Convex Functions and Convex Bodies

by

Anthony R. Cappello

BA, Houghton College, 2013

MS, Emporia State University, 2018

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Anthony R. Cappello

It was defended on

May 30th 2024

and approved by

Dissertation Director: Piotr Hajłasz, Department of Mathematics, University of Pittsburgh

Jason DeBlois, Department of Mathematics, University of Pittsburgh

Giovanni Leoni, Department of Mathematics, Carnegie Mellon University

Juan Manfredi, Department of Mathematics, University of Pittsburgh

On the Differentiability Properties of Convex Functions and Convex Bodies

Anthony R. Cappello, PhD

University of Pittsburgh, 2024

There are three main results in this thesis. We first present a new proof of Theorem 40 that says a convex body K has boundary of class $C^{1,1}$ if and only if there is $R > 0$ such that K is the union of closed balls with radius R contained in K. The first main result, Theorem 51 extends the above result to a similar characterization of $C^{1,\alpha}$ convex bodies. Using this characterization, we find new proofs of the Kirchheim-Kristensen theorem (Theorem 70) about the differentiability of the convex envelope and the Krantz-Parks theorem (Theorem 77) about the regularity of the Minkowski sum of convex bodies. Namely we show that if a convex function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $f \in C^{1,\alpha}_{\text{loc}}(\mathbb{R}^n)$ and $f(x) \to \infty$ as $|x| \to \infty$, then the convex envelope of f, denoted $\text{conv}(f)$, satisfies $\text{conv}(f) \in C^{1,\alpha}_{\text{loc}}(\mathbb{R}^n)$. We also prove that the Minkowski sum of a convex body and a convex body of class $C^{1,\alpha}$ is a convex body of class $C^{1,\alpha}$. The tools from the characterization of $C^{1,1}$ convex bodies are used to prove the second main result, which is a new geometrically inspired proof of the Alexandrov theorem, Theorem 84, about the second order differentiability of convex functions. Moreover, we give a new proof of a result by Azagra-Hajlasz (Theorem 90) concerning the Lusin Approximation by $C^{1,1}$ convex functions. In the third main result, Theorem 108, we prove the set of normal directions to the k-dimensional faces on the boundary of an n-dimensional convex body is countable $(n-k-1)$-rectifiable. Finally we conclude by presenting characterizations of $C^{1,1}$ and $C^{1,\alpha}$ functions.

keywords convex body, convex function, Lipschitz gradient, Hölder gradient, convex envelope, Minkowski sum, Alexandrov's theorem, support function.

Table of Contents

Preface . . . viii

1.0 Introduction . . . 1

1.1 Summary of Main Chapters . . . 2

2.0 Preliminaries . . . 7

2.1 Notation . . . 7

2.2 Rademacher's Theorem . . . 9

2.3 Convex Sets . . . 13

2.4 Convex Functions . . . 18

2.5 Differentiability Properties of Convex Functions . . . 22

3.0 Differentiability Properties of the Boundaries of Convex Bodies . . . 35

3.1 Convex Bodies and Convex Domains . . . 35

3.2 Convex Bodies Satisfying the Uniform Inner Ball Condition . . . 40

3.3 Convex Bodies Satisfying the Approximate Inner Ball Condition . . . 46

3.4 Regularity of the Convex Envelope . . . 59

3.5 Sum of Convex Bodies . . . 66

4.0 Second Differentiability of Convex Functions . . . 69

4.1 Approximating Convex Bodies with Lipschitz Outer Normal Vectors . . . 69

4.2 Alexandrov's Theorem . . . 71

4.3 The Differentiability of the Subdifferential of a Convex Function . . . 76

4.4 Approximating Convex Functions Globally with Continuously Differentiable Convex Functions with Lipschitz Gradient . . . 77

5.0 Support Functions and Boundary Properties of Convex Bodies . . . 80

5.1 The Support Function . . . 80

5.2 Properties of the Support Function . . . 83

5.3 Rectifiability of the Normal Directions . . . 85

6.0 Equivalence of Continuously Differentiable Convex Functions with Hölder Continuous Gradient 93
6.1 Convex Functions with Lipschitz Continuous Gradient 93
6.2 Convex Functions with Hölder Continuous Gradient 99
Bibliography 106

List of Figures

Figure 1: Illustration of Notation for Theorem 51 51
Figure 2: Proof of Lemma 68 64

Preface

First, I would like to thank my advisor Dr. Hajłasz, for helping from the very beginning of my time at Pitt to achieve my dream. Thank you for pushing me to be better, providing me with research questions that were perfect for my interests, and for teaching me how to be a mathematician.

I would like to thank my undergraduate professors Dr. Rebekah Yates, Dr. Kristen Camenga, and Dr. Jill Jordan, for inspiring me and creating a love of mathematics that has never diminished.

I would like to thank my coworkers and fellow math teachers at Asheboro High School, especially Beth Beckwith and Adam Reeder; the main members of our math parties. Not only did we have a lot of fun talking math but I also learned a lot about how to best teach and inspire others.

I would like to thank the members of my committe, Dr. DeBlois, Dr. Hajłasz, Dr. Leoni, and Dr. Manfredi for giving up time in their busy schedules to help me complete my degree. Also thanks to Dr. DeBlois, Dr. Hajłasz and Dr. Neilan, for writing letters of recommendations and being references while I searched for a job.

Thanks to all the family and friends who have supported me and a special thanks to my friend Dr. Scott Donahue Martens for supporting me and commiserating with me as we both traversed the minefield of completing a PhD with two young children.

Thanks to my dog Molly for providing tons of emotional support.

Finally, a huge massive thanks to my wife, Emily. You have been there for me through the many ups and downs, supported me, encouraged me, and sacrificed much for me to help me achieve my dream. Together we have co-authored two of my greatest theorems, Benjamin and Jack and I am extremely grateful to you for all that you have done for me and our family through all the craziness over the last few years. My success in completing this thesis is directly correlated to the love and support you have given to me and I am forever grateful to you.

1.0 Introduction

The study of convex sets began with the ancient Greeks, with Archimedes considered the first to provide a rigorous definition of convexity. Moreover, in the Elements of Euclid, polygons and polytopes are extensively studied; the building blocks of convex geometry. Up until the early 20th century, the study of convexity was centered around geometrical ideas such as Kepler's study of polytopes and packing of balls, Euler's famous relation between the vertices, edges, and faces of convex polytopes in three dimensions, and Cauchy's surface area formula connecting the surface area of a three dimensional convex body with the area of its projections onto two dimensional subspaces.

By the end of the 19th century convex functions entered the scene. As stated by Pachpatte in [25], mathematicians such as Hölder, Hadamard, and Stolz are attributed with being the first to work with convex functions. It wasn't until Jensen published his famous inequality in 1905 and 1906 that convex functions became a relevant field of study in mathematics. Throughout the 20th century the study of convex functions found its way into numerous areas of mathematics including functional analysis, complex analysis, and PDE's. Moreover, the application of convex functions to the field of convex optimization has led to many useful results in many fields including statistics, data analysis, and risk analysis (see [10]).

In this thesis the goal is to find connections between convex sets, functions, and their differentiability properties. Namely, if we consider compact convex sets with non empty interior, called convex bodies, we have that locally the boundary of a convex body is the graph of a convex function. Thus we are able to combine the geometry of a convex body with the differentiablity of convex functions to prove some truly wonderful results. There are many surprising and beautiful results that appear in convex analysis, especially given that at its core convex analysis is founded upon purely geometric definitions. With nothing more than the notion of line segments, balls, and hyperplanes we are able to extract numerous differentiability properties of convex functions and convex bodies.

1.1 Summary of Main Chapters

In Chapter 2 we collect known results about convex sets and convex functions that will be used throughout the thesis. We include proofs of most of the results so that the thesis will be self-contained. We also include the standard proof of the Rademacher theorem about the a.e. differentiability of Lipschitz functions, as the Rademacher theorem plays a central role in Chapter 4.

Chapter 3 is concerned with the differentiability properties of the boundary of a convex body, defined as a compact convex set with nonempty interior. Since the boundary of a convex body is locally the graph of a convex function, we can investigate the differentiability properties of convex functions through the study of the regularity of boundaries of convex bodies. This approach will be used later in Chapter 4.

We denote the class of functions with α-Hölder continuous gradients as $C^{1,\alpha}$ and in particular when $\alpha = 1$, a $C^{1,1}$ function has Lipschitz gradient. Moreover we say that a convex body is of class $C^{1,\alpha}$ if locally its boundary is the graph of $C^{1,\alpha}$ convex function.

When K is the union of closed balls of fixed radius, we say that K satisfies the uniform inner ball condition. Lucas [21] proved that a convex body is of class $C^{1,1}$, if and only if K satisfies the uniform inner ball condition (Theorem 40). Despite Lucas proving this result in his thesis [21], the result was never published in a paper. As a result, the theorem is not widely known to be Lucas' and it is difficult to find this result in published literature. The characterization of $C^{1,1}$ convex bodies does appear in Hormander [17, Proposition 2.4.3], though no references are made to the origin of the result. Moreover the proof in [17] is different than that of Lucas. We provide two new proofs of Lucas' theorem. The first is similar to that of Hormander's but is more geometric, while the second proof is completely new and is based on an application of the implicit function theorem for $C^{1,1}$ functions. Both proofs have been published in [3].

The main focus of Chapter 3 is an extension of Lucas' theorem that provides a geometric characterization of $C^{1,\alpha}$ convex bodies. We say $K \subset \mathbb{R}^n$ satisfies the (R, ε)-approximate inner ball condition, if for each $x \in \partial K$, there exists $h(x) \in K$ such that $\overline{B}(h(x), R) \subset K$ and $\mathrm{dist}(x, \overline{B}(h(x), R)) \leq \varepsilon$. We prove that a convex body K is of class $C^{1,\alpha}$ if and only if there exist constants $\varepsilon_0 > 0$ and $C > 0$, such that for all $0 < \varepsilon < \varepsilon_0$, K satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition (Theorem 51). The proof of the necessary condition is an extension of the first proof given of Lucas' theorem in this thesis, but the proof of the sufficient condition relies on making geometric estimates with the inner unit normal vectors. When $\alpha = 1$, we see that for all $0 < \varepsilon < \varepsilon_0$, K satisfies the (C, ε)-approximate inner ball condition. But by compactness we have there exists some $h(x)$ such that $x \in \overline{B}(h(x), C) \subset K$ showing that K is the union of closed balls of radius C. Therefore the characterization of $C^{1,\alpha}$ convex bodies is an extension of Lucas' result.

Using the characterization of $C^{1,\alpha}$ convex bodies we are able to find new geometric proofs of theorems related to the convex envelope and the sum of two convex bodies.

We define the convex envelope of a function f as

$$\operatorname{conv}(f)(x) := \sup\{g(x) : g \leq f \text{ and } g \text{ is convex}\}.$$

In [18], Kirchheim and Kristensen proved that if $f \in C^{1,\alpha}_{\text{loc}}$ and $f \to \infty$ as $|x| \to \infty$, then $\operatorname{conv}(f) \in C^{1,\alpha}_{\text{loc}}$ (Theorem 70). The proof presented in [18] follows from inequalities derived from an analytic view of the problem. The new proof presented in this thesis is based on geometrical principles and is elementary. The proof presented in this thesis uses an equivalent definition of the convex envelope which follows from the Carathéodory theorem, namely,

$$\operatorname{conv}(f)(x) = \inf\left\{\sum_{i=1}^{n+1} \lambda_i f(x_i) : \lambda_i \geq 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1, \quad \sum_{i=1}^{n+1} \lambda_i x_i = x\right\}.$$

Using this we are able to first apply the characterization of $C^{1,\alpha}$ convex bodies locally to the epigraph of f, and then taking convex combinations show that $\operatorname{conv}(f)$ also satisfies an approximate inner ball condition locally.

The other main application in Chapter 3 concerns the Minkowski sum of convex bodies. We define the Minkowski sum of sets $A, B \subset \mathbb{R}^n$ as,

$$A + B := \{a + b : a \in A \text{ and } b \in B\}.$$

If A, B are convex bodies, then so too is $A + B$. Moreover applying the characterization of $C^{1,\alpha}$ convex bodies, we can show that for convex body A of class $C^{1,\alpha}$, and a general convex body B, that $A + B$ is a convex body of class $C^{1,\alpha}$ (Theorem 77). The result is originally proved by Krantz and Parks in [20] through the use of coordinate systems but it is difficult. An alternate proof is given in [19], making use of the infimal convolution, though again relying on coordinate systems. Thus the arguments given in this thesis greatly simplify the proofs of their results.

In Chapter 4 we focus on one of the most fascinating differentiability properties of convex functions, specifically with regards to its second differentiability. In 1939 Alexandrov in [1] proved that for U open and convex and $f : U \to \mathbb{R}$ convex, f is twice differentiable almost everywhere and the gradient is differentiable almost everywhere in U. Numerous proofs have been given since by Rešetnjak; Krylov; Bangert; Rockafellar; Bianchi, Colesanti, and Pucci and a complete history of the theorem can be found in [8]. We present in this chapter a new geometrically inspired proof of Alexandrov's theorem given in [3] that is considered to be the most elementary proof of the theorem.

The theorem of Alexandrov can be stated in two parts (Theorem 84, Theorem 89). The first part states, for $f : \mathbb{R}^n \to \mathbb{R}$, convex, at almost every point where f is differentiable, there is a symmetric matrix denoted by $D^2 f(x)$ such that

$$\lim_{y \to x} \frac{f(y) - f(x) - Df(x)(y-x) - \frac{1}{2}(y-x)^T D^2 f(x)(y-x)}{|y-x|^2} = 0. \tag{1}$$

The second part then states, if $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then for all $x \in \mathbb{R}^n$ where f is twice differentiable as in (1), we have

$$\lim_{y \to x} \sup_{\sigma_y \in \partial f(y)} \frac{|\sigma_y - Df(x) - D^2 f(x)(y-x)|}{|y-x|} = 0 \tag{2}$$

where ∂f denotes the subdifferential of f. When proving the Alexandrov theorem, it has been standard to prove the second part first and use that to prove the first part. In this thesis we provide a proof of the first part and then use that to prove the second part. The proof of the first part is surprisingly simple. First we present a new proof of a result by McMullen [22], that for a convex body K, at almost every $x \in \partial K$, we can find a closed ball containing x contained in K (Theorem 79). Thus choosing some radius $R > 0$, sufficiently small, we can find a convex body $K(R) \subset K$, where $K(R)$ is the union of closed balls of radius R contained in K, and the surface area of $K \setminus K(R)$ is arbitrarily small. Moreover Lucas' theorem tells us that $K(R)$ is a $C^{1,1}$ convex body. Thus given a convex function $f : \mathbb{R}^n \to \mathbb{R}$

we can approximate the graph of f, locally with a $C^{1,1}$ convex body, and parameterizing the bottom part of the convex body yields a convex function $g \in C^{1,1}_{\text{loc}}$ that agrees with f on a set of small measure. Given that ∇g is Lipschitz, we then apply the Rademacher theorem to show that ∇g is differentiable almost everywhere, and this is precisely the second derivative we were seeking for f.

We conclude Chapter 4 by using a corollary needed for the proof of the Alexandrov theorem to find a new proof for a recent result by Azagra and Hajłasz in [4] concerning the Lusin-type property of $C^{1,1}$ convex functions (Theorem 90). The result can be stated as: for a convex $f : \mathbb{R}^n \to \mathbb{R}$, for every $R > 0$, and for every $\varepsilon > 0$, there exists a convex function $g \in C^{1,1}(\mathbb{R}^n)$, $g \geq f$, such that

$$|\{x \in B^n(0, R) : \ f(x) \neq g(x)\}| < \varepsilon.$$

The original proof for this is technical but the proof is simplified by the use of Lucas' theorem.

The goal of Chapter 5 is the study of the support function and its differentiability properties. The support function is defined as,

$$\sigma_K(x) = \sup_{k \in K} \langle x, k \rangle$$

and is a useful tool to describe the structure of the boundary of a convex body. We define the supporting hyperplane of K in the direction of u as,

$$H(u, K) = \{x \in \mathbb{R}^n : \langle x, u \rangle = \sigma_K(u)\}.$$

It is easy to see that σ_K is convex but moreover we have that σ_K is differentiable at $u \in \mathbb{R}^n \setminus \{0\}$ if and only if $H(u, K)$ intersects the boundary of K at exactly one point. The main result of this chapter is using this geometric understanding of the differentiability of the support function to show that the set of normal vectors $u \in \mathbb{S}^{n-1}$, such that the intersection of $H(u, K)$ with K is a d-dimensional face, is countably $(n - d - 1)$ rectifiable (Theorem 110). For convex f, we define

$$\Sigma^d(\partial f) := \{x \in \mathbb{R}^n : \dim((\partial f)(x)) \geq d\},$$

where ∂f is the subdifferential of f. To prove Theorem 110, we rely on a theorem originally proven by Zajíček [29] which shows we can cover $\Sigma^d(\partial f)$ by the graphs of locally Lipschitz functions of dimension $n - d$.

Finally in Chapter 6 we establish a list of equivalent statements for $C^{1,1}$ and $C^{1,\alpha}$ convex functions (Theorem 111, Theorem 114). These lists appear to be nonexistent as a whole in the published literature and thus the hope is this will be a useful resource for those interested in the differentiability of convex functions.

2.0 Preliminaries

2.1 Notation

An open ball in $\mathbb{R}^n$ will be denoted as $B(x,r)$ (or $B^n(x,r)$ if the dimension is not clear) and the closed ball will be denoted by $\overline{B}(x,r)$ ($\overline{B}^n(x,r)$). Often we will consider balls that are associated with a point, usually not the center. Thus we will use $h(x)$ or similarly $h_\varepsilon(x)$ to denote the center of the ball that has some relation to x e.g. for every $x \in K$ there exists some closed ball $\overline{B}(h(x),R) \subset K$ such that $x \in \overline{B}(h(x),R) \subset K$.

We denote the line segment from $x \in \mathbb{R}^n$ to $y \in \mathbb{R}^n$ as $[x,y]$, that is,

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

We will only be working within the Euclidean space $\mathbb{R}^n$, where $\langle \cdot, \cdot \rangle$ is the standard inner product given by $\langle x,y \rangle = \sum_{i=1}^n x_i y_i$. Moreover $|x| = \sqrt{\langle x,x \rangle}$ represents the Euclidean norm on $\mathbb{R}^n$. We define the unit-sphere in $\mathbb{R}^n$ as, $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$. We denote the hyperplane as

$$H_b(u) = \{x : \langle x,u \rangle = b\} \tag{3}$$

and the closed half space with outer normal u, as

$$H_b^-(u) = \{x : \langle x,u \rangle \leq b\} \tag{4}$$

For a function $f : U \to \mathbb{R}$, we define $\Gamma_f : U \to U \times \mathbb{R}$ to be the graph of f defined by $\Gamma_f(x) := (x, f(x))$. For any set $E \subset \mathbb{R}^n$, we denote the distance to E as,

$$\operatorname{dist}(x,E) = \inf_{y \in E} |x-y|$$

and it is easy to prove that for every closed set $E \subset \mathbb{R}^n$, and any $x \notin E$, there exists $y \in E$, not necessarily unique, such that $\operatorname{dist}(x,E) = |x-y|$. The diameter of E is given by,

$$\operatorname{diam}(E) = \sup_{x,y \in E} |x-y|.$$

The Lipschitz constant of $f : E \to \mathbb{R}$, when it exists, is denoted by

$$\operatorname{Lip}(f, E) = \sup_{\substack{x,y \in E \\ x \neq y}} \frac{|f(x) - f(y)|}{|x - y|},$$

and if f is Lipschitz on all of $\mathbb{R}^n$, we then denote the Lipschitz constant by $\operatorname{Lip}(f)$.

For a measurable set $A \subset \mathbb{R}^n$ we denote the n-dimensional Lebesgue measure of A as $\mathcal{L}^n(A)$. We define the constant ω_s, for $s \geq 0$, as

$$\omega_s = \frac{\pi^{s/2}}{\Gamma\left(1 + \frac{s}{2}\right)},$$

where $\Gamma(s) := \int_0^\infty e^{-x} x^{s-1}\, dx$ is the gamma function and we note that when $s = n \in \mathbb{N}$, ω_n is the volume of the unit ball in $\mathbb{R}^n$. Let X be a metric space. For $\varepsilon > 0$ and $E \subset X$, we define,

$$\mathcal{H}^s_\varepsilon(E) = \inf \left\{ \frac{\omega_s}{2^s} \sum_{i=1}^\infty (\operatorname{diam} A_i)^s : E \subset \bigcup_{i=1}^\infty A_i \text{ with } \operatorname{diam} A_i < \varepsilon \right\}.$$

We call $\mathcal{H}^s(E)$ the s *dimensional Hausdorff measure* of E which is given by,

$$\mathcal{H}^s(E) = \lim_{\varepsilon \to 0^+} \mathcal{H}^s_\varepsilon(E)$$

and as the function $\varepsilon \mapsto \mathcal{H}^s_\varepsilon(E)$ is nonincreasing, this limit will always exist in the extended reals.

It is well known that for any measurable $A \subset \mathbb{R}^n$, we have $\mathcal{L}^n(A) = \mathcal{H}^n(A)$ (see [13, Theorem 2.5]), and we generally refer to $\mathcal{L}^n(A)$ as the volume of A. Similarly, we will refer to $\mathcal{H}^{n-1}(\partial A)$ as the surface area of A. By the definition of $\mathcal{H}^n$, if $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with $\operatorname{Lip}(f) = L$, then $\mathcal{H}^n(f(A)) \leq L^n \mathcal{H}^n(A)$ [13, Theorem 2.8].

For $A \subset \mathbb{R}^n$, x is a *density point of* A if,

$$\lim_{r \to 0} \frac{\mathcal{L}^n(A \cap B^n(x, r))}{\mathcal{L}^n(B^n(x, r))} = 1$$

and by the Lebesgue differentiation theorem, for a measurable set A, a.e. point in A is a density point.

We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is positively *1-homogeneous* if for all $t > 0$ and for all $x \in \mathbb{R}^n$ we have, $f(tx) = tf(x)$, and *subadditive* if for all $x, y \in \mathbb{R}^n$,

$$f(x + y) \leq f(x) + f(y).$$

2.2 Rademacher's Theorem

Before discussing the differentiability properties of convex functions we first will provide details for the well known Rademacher's theorem. We include the details here for completeness sake but this section stands alone and the material contained in it is not needed elsewhere, aside from the statement of Rademacher's theorem. The proof of the Rademacher theorem is included, as well as theorem statements for well known results in measure theory and the details for many of the theorems can easily be found in standard textbooks on real analysis. The results below may work with more general measures but have been stated using only the Lebesgue measure as that is the only measure needed for this section.

Let $U \subset \mathbb{R}^n$ be open. The support of a function $f : U \to \mathbb{R}$ is denoted $\operatorname{supp}(f)$ and is defined as

$$\operatorname{supp}(f) = \overline{\{x \in U : f(x) \neq 0\}}.$$

We define $C_c^\infty(U)$ to be the space of smooth functions defined on U with compact support.

Theorem 1. *$C_c^\infty(U)$ is dense in $L^1(U)$.*

Lemma 2. *(Fundamental lemma of Calculus of Variations) If $g \in L^1_{\text{loc}}(U)$ and for every $\phi \in C_c^\infty(U)$ we have*

$$\int_U g(x)\phi(x)\, dx = 0$$

then $g = 0$ a.e.

The proof presented here for Rademacher's theorem can be found in [14, Theorem 113].

Theorem 3 (Rademacher's theorem)**.** *If $f : U \to \mathbb{R}$ is Lipschitz continuous, where $U \subset \mathbb{R}^n$ is open, then*

$$\nabla f(x) = \left\langle \frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right\rangle$$

exists almost everywhere. Moreover for all $x \in U$ where $\nabla f(x)$ exists we have,

$$\lim_{y \to x} \frac{f(y) - f(x) - \nabla f(x) \cdot (y - x)}{|y - x|} = 0.$$

Proof. Let $\nu \in \mathbb{S}^{n-1}$ and let

$$D_\nu f(x) = \frac{d}{dt} f(x + tv)\big|_{t=0}$$

be the directional derivative of f in the direction ν wherever it exists. Let A_ν denote the set of $x \in U$ such that $D_\nu f(x)$ does not exist. As both

$$\liminf_{t\to 0} \frac{f(x+t\nu)-f(x)}{t} \quad \text{and} \quad \limsup_{t\to 0} \frac{f(x+t\nu)-f(x)}{t}$$

are Borel measurable functions and $D_\nu f(x)$ exists when

$$\liminf_{t\to 0} \frac{f(x+t\nu)-f(x)}{t} = \limsup_{t\to 0} \frac{f(x+t\nu)-f(x)}{t}$$

we have that A_ν is Borel measurable. We also know that as f is Lipschitz, the function $t \mapsto f(x+t\nu)$ is absolutely continuous and hence differentiable almost everywhere. Hence the intersection of the set A_ν with any line parallel to ν has one dimensional measure zero and we can apply Fubini's theorem to see that A_ν has measure zero. Therefore for every $\nu \in \mathbb{S}^{n-1}$, $D_\nu f(x)$ exists for almost every $x \in U$. Fix $\nu \in \mathbb{S}^{n-1}$ and $\phi \in C_c^\infty(U)$ and choose $h > 0$ small enough so that $x + h\nu \in U$ for all $x \in \operatorname{supp}(\phi)$. By the invariance of the integral with respect to translations, we have

$$\int_U f(x+h\nu)\phi(x)\,dx = \int_U f(x)\phi(x-h\nu)\,dx$$

so that for h sufficiently small

$$\int_U \frac{f(x+h\nu)-f(x)}{h}\phi(x)\,dx = -\int_U \frac{\phi(x-h\nu)-\phi(x)}{-h} f(x)\,dx.$$

By the Dominated convergence theorem we have, letting $h \to 0$,

$$\int_U D_\nu f(x)\phi(x)\,dx = -\int_U f(x) D_\nu \phi(x)\,dx.$$

As this is true for any $\nu \in \mathbb{S}^{n-1}$, in particular it is true for $v = e_i$ so that

$$\int_U \frac{\partial f}{\partial x_i}(x)\phi(x)\,dx = -\int_U f(x)\frac{\partial \phi}{\partial x_i}(x)\,dx \quad \text{for } i = 1, 2, \dots, n.$$

Hence,

$$\begin{aligned}\int_U D_\nu f(x)\phi(x)\,dx &= -\int_U f(x) D_\nu \phi(x)\,dx = -\int_U f(x)(\nabla\phi(x)\cdot\nu)\,dx \\ &= -\sum_{i=1}^n \int_U f(x)\frac{\partial\phi}{\partial x_i}(x)\nu_i\,dx = \sum_{i=1}^n \int_U \frac{\partial f}{\partial x_i}(x)\phi(x)\nu_i\,dx \\ &= \int_U \phi(x)(\nabla f(x)\cdot\nu)\,dx. \end{aligned} \tag{5}$$

Applying Lemma 2 we have that $D_\nu f(x) = \nabla f(x)\cdot\nu$ a.e. Let $\nu_1, \nu_2, \ldots$ be a countable dense subset of $\mathbb{S}^{n-1}$ and define the sets

$$A_k = \{x \in U : \nabla f(x),\, D_{\nu_k} f(x) \text{ exist and } D_{\nu_k} f(x) = \nabla f(x)\cdot\nu_k\},$$

and $A = \bigcap_{k=1}^\infty A_k$. Given that $|U \setminus A_k| = 0$ for each k this implies that $|U \setminus A| = 0$ and by the definition of A,

$$D_{\nu_k} f(x) = \nabla f(x)\cdot\nu_k \quad \text{for al } x \in A \text{ and all } k = 1, 2, \ldots$$

The claim is that f is differentiable on A. Consider the function

$$Q(x,\nu,h) = \frac{f(x+h\nu) - f(x)}{h} - \nabla f(x)\cdot\nu,$$

where $x \in A$, $\nu \in \mathbb{S}^{n-1}$, and $h > 0$. To show f is differentiable on A we need only show that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $0 < h < \delta$ implies $|Q(x,\nu,h)| < \varepsilon$. Given that f is L-Lipschitz and $\frac{\partial f}{\partial x_i}(x)$ exists a.e. we have that $|\partial f/\partial x_i(x)| \le L$ a.e. and by Cauchy-Schwarz, $|\nabla f(x)| \le \sqrt{n}L$ a.e. Thus for any $\nu, \nu' \in \mathbb{S}^{n-1}$ we have,

$$\begin{aligned}&|Q(x,\nu,h) - Q(x,\nu',h)| \\ &\qquad = \left|\frac{f(x+h\nu)-f(x)}{h} - \nabla f(x)\cdot\nu - \left(\frac{f(x+h\nu')-f(x)}{h} - \nabla f(x)\cdot\nu'\right)\right| \\ &\qquad = \left|\frac{f(x+h\nu)-f(x+h\nu')}{h} + \nabla f(x)\cdot\nu' - \nu\right| \\ &\qquad \le L|\nu-\nu'| + |\nabla f(x)||\nu-\nu'| = (\sqrt{n}+1)L|\nu-\nu'|.\end{aligned}$$

By the density of $\{\nu_k\}$ and the compactness of $\mathbb{S}^{n-1}$ for every $\varepsilon > 0$ there exists N large enough so that

$$\mathbb{S}^{n-1} \subset \bigcup_{i=1}^{N} B\left(\nu_i, \frac{\varepsilon}{2(\sqrt{n}+1)L}\right),$$

i.e. for each $\nu \in \mathbb{S}^{n-1}$ there exists some $k = 1, 2, \dots, N$ such that

$$|\nu - \nu_k| \leq \frac{\varepsilon}{2(\sqrt{n}+1)L}.$$

By the construction of A we have for all $x \in A$ and $\{\nu_i\}_{i=1}^{\infty}$,

$$\lim_{h \to 0^+} Q(x, \nu_i, h) = 0$$

and thus for each $x \in A$ and ν_i there exists $\delta_i > 0$ such that $0 < h < \delta_i$ implies $|Q(x, \nu_i, h)| < \varepsilon/2$. Choosing $\delta := \min\{\delta_1, \dots, \delta_N\}$ yields for all $0 < h < \delta$ and for all $i = 1, \dots, N$,

$$|Q(x, \nu_i, h)| < \varepsilon/2.$$

Therefore, combining the aforementioned inequalities we have, for $x \in A$, $\nu \in \mathbb{S}^{n-1}$, and $0 < h < \delta$,

$$|Q(x, \nu, h)| \leq |Q(x, \nu_k, h)| + |Q(x, \nu_k, h) - Q(x, \nu, h)| < \frac{\varepsilon}{2} + (\sqrt{n}+1)L|\nu - \nu_K| < \varepsilon$$

completing the proof. **QED**

In this chapter we will cover much of the preliminary facts of convex geometry and convex analysis needed throughout the thesis. There are many texts used in the compilation of these topics in convex analysis including, [11], [16], [23], [26], and [27]. While most of the stated definitions and theorems are elementary and standard in convex analysis we provide all necessary details to ensure all readers are able to understand the entirety of the thesis without needing to look up results.

2.3 Convex Sets

We say a set $C \subset \mathbb{R}^n$ is *convex* if for every $x, y \in C$, $[x, y] \subset C$. Moreover given a set $A \subset \mathbb{R}^n$ we define its *convex hull* as

$$\operatorname{co}(A) := \left\{ \sum_{i=1}^{m} \lambda_i x_i : \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1, \text{ and } x_i \in A, m \in \mathbb{N} \right\},$$

where we call $\sum_{i=1}^{m} \lambda_i x_i$ a *convex combination* of $x_1, \ldots, x_m \in A$ when the λ_i's satisfy $\lambda_i \geq 0$ and $\sum_{i=1}^{m} \lambda_i = 1$. Thus, in words, we can say the convex hull of A is the set of convex combinations of A. Equivalently we can define the convex hull of A to be the intersection of all convex sets containing A. Similarly we can define the *affine hull* of A to be the set of affine combinations of A where we define an affine combination of A to be of the form

$$\sum_{i=1}^{m} \lambda_i x_i \text{ with } \sum_{i=1}^{m} \lambda = 1 \text{ and } x_i \in A,$$

or equivalently the intersection of all affine sets containing A. The main difference is that a convex combination has only non-negative coefficients and affine combinations can have negative coefficients. From these definitions we can see that every convex set is contained in an affine set and using this we define the *dimension of a convex set* C to be the dimension of its affine hull.

When discussing affine sets, we say that the vectors $x_0, \ldots, x_k$ are affinely independent if $x_1 - x_0, \ldots, x_k - x_0$ are linearly independent. It is equivalent to say that the vectors are affinely independent if

$$\sum_{i=0}^{k} \lambda_i x_i = 0 \quad \text{and} \quad \sum_{i=0}^{k} \lambda_i = 0 \tag{6}$$

implies $\lambda_0 = \cdots = \lambda_k = 0$.

In the above characterization of the convex hull of a set we are considering the convex combinations of m points in A, where m is an arbitrary natural number. In 1911 Carathéodory strengthened this representation by showing that for a compact set $A \subset \mathbb{R}^n$, its convex hull can be represented by the convex combination of only $n + 1$ elements of A and in 1914 Steinitz extended the result for general sets. A proof of Carathéodory's theorem can be found in [27, Theorem 17.1] but we present here the proof in [28, Theorem 1.1.4].

Theorem 4 (Carathéodory's Theorem)**.** *For any subset $A \subset \mathbb{R}^n$, its convex hull admits the representation*

$$co(A) = \left\{ \sum_{i=1}^{n+1} \lambda_i x_i : \sum_{i=1}^{n+1} \lambda_i = 1,\ \lambda_i \geq 0,\ x_i \in A \right\}.$$

Proof. By definition, for $x \in \mathrm{co}(A)$, we have

$$x = \sum_{i=1}^{m} \lambda_i x_i, \quad \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0,\ x_i \in A \text{ for } i = 1, \dots, m, \tag{7}$$

for some $m \in \mathbb{N}$, where we assume that m is the smallest possible value. Thus x cannot be written as the convex combination of k elements of A if $k < m$. This means we can assume $\lambda_i > 0$ for each $i = 1, \dots, m$. If $m > n + 1$, then the points $x_1, \dots, x_m$ must be affinely dependent. Hence, by (6), there exists constants $\mu_1, \dots, \mu_m$, at least one non-zero, such that

$$\sum_{i=1}^{m} \mu_i x_i = 0 \quad \text{and} \quad \sum_{i=1}^{m} \mu_i = 0. \tag{8}$$

Moreover we know there exists $i = 1, \dots, m$ such that $\mu_i > 0$. Thus we can reorder the λ_i and μ_i such that $\mu_m > 0$ and

$$\frac{\lambda_m}{\mu_m} = \min \left\{ \frac{\lambda_i}{\mu_j} : i = 1, \dots, m \text{ and } \mu_j > 0 \right\} > 0. \tag{9}$$

Hence we have, by (8)

$$0 = \frac{\lambda_m}{\mu_m} \sum_{i=1}^{m} \mu_i x_i = \sum_{i=1}^{m-1} \frac{\lambda_m \mu_i}{\mu_m} x_i + \lambda_m x_m$$

and subtracting this from (7) yields

$$x = \sum_{i=1}^{m-1} \left(\lambda_i - \frac{\lambda_m \mu_i}{\mu_m} \right) x_i \tag{10}$$

where obviously the coefficients are positive if $\mu_i \leq 0$ and by (9), $\lambda_i - \frac{\lambda_m \mu_i}{\mu_m} \geq 0$ if $\mu_i > 0$. Finally, we see by (8)

$$\sum_{i=1}^{m-1} \left(\lambda_i - \frac{\lambda_m \mu_i}{\mu_m} \right) = \sum_{i=1}^{m-1} \lambda_i + \lambda_m \sum_{i=1}^{m-1} \frac{-\mu_i}{\mu_m} = \sum_{i=1}^{m-1} \lambda_i + \lambda_m = 1$$

so that by (10), x is the convex combination of $m - 1$ elements of A, contradicting the minimality of m. **QED**

If $C \subset \mathbb{R}^n$ is a closed convex set, then for every $x \in \mathbb{R}^n$, there is a unique point denoted by $\pi_C(x)$ such that

$$\pi_C(x) \in C \qquad \text{and} \qquad |x - \pi_C(x)| = \operatorname{dist}(x, C). \tag{11}$$

Indeed, if there exist $c_1, c_2 \in C$ such that $c_1 \neq c_2$ and

$$|x - c_1| = |x - c_2| = \operatorname{dist}(x, C),$$

then by convexity $\frac{c_1+c_2}{2} \in B(x, \operatorname{dist}(x, C))$, which implies

$$\left| x - \left(\frac{c_1}{2} + \frac{c_2}{2} \right) \right| < \operatorname{dist}(x, C).$$

But by convexity $\frac{c_1}{2} + \frac{c_2}{2} \in C$, a contradiction. We call the mapping $\pi_C : \mathbb{R}^n \to C$ the *metric projection* onto C. Clearly, if $x \notin C$, then $\pi_C(x) \in \partial C$.

Lemma 5. *If $C \subset \mathbb{R}^n$ is closed and convex, then $\pi_C : \mathbb{R}^n \to C$ is 1-Lipschitz.*

Proof. Let $x, y \in \mathbb{R}^n$. By convexity of C, $t\pi_C(x) + (1-t)\pi_C(y) \in C$ for all $t \in (0, 1)$ and hence

$$\begin{aligned}
&|y - \pi_C(y)|^2 = \operatorname{dist}(y, C)^2 \leq |y - (t\pi_C(x) + (1-t)\pi_C(y))|^2 \\
&= |(y - \pi_C(y)) - t(\pi_C(x) - \pi_C(y))|^2 \\
&= |y - \pi_C(y)|^2 - 2t\langle y - \pi_C(y), \pi_C(x) - \pi_C(y)\rangle + t^2|\pi_C(x) - \pi_C(y)|^2
\end{aligned}$$

which can be simplified to

$$2\langle y - \pi_C(y), \pi_C(x) - \pi_C(y)\rangle \leq t|\pi_C(x) - \pi_C(y)|^2.$$

Letting $t \to 0^+$ yields

$$\langle y - \pi_C(y), \pi_C(x) - \pi_C(y)\rangle \leq 0. \tag{12}$$

By switching the role of x and y we also have

$$\langle x - \pi_C(x), \pi_C(y) - \pi_C(x)\rangle \leq 0. \tag{13}$$

Adding inequalities (12) and (13) yields

$$|\pi_C(x) - \pi_C(y)|^2 \leq \langle x - y, \pi_C(x) - \pi_C(y)\rangle \leq |x - y|\, |\pi_C(x) - \pi_C(y)|$$

and hence $|\pi_C(x) - \pi_C(y)| \leq |x - y|$. **QED**

Given $x \in \mathbb{R}^n \setminus C$ the vector $x - \pi_C(x)$ is of particular import, so we define the unit vector

$$u_C(x) = \frac{x - \pi_C(x)}{|x - \pi_C(x)|}. \tag{14}$$

Using the proof of Lemma 5 we can obtain the following result:

Corollary 6. *Let $C \subset \mathbb{R}^n$ be a closed convex set and $x \in \mathbb{R}^n \setminus C$. Then there exists $b \in \mathbb{R}$ such that $C \subset H_b^-(u_C(x))$ and $\pi_C(x) \in H_b(u_C(x))$, where H_b and H_b^- are defined in* (3) *and* (4)

Proof. Note that the hyperplane through $\pi_C(x)$ with outer unit normal $u_C(x)$, is given by,

$$\{z \in \mathbb{R}^n : \langle u_C(x), z - \pi_C(x)\rangle = 0\}.$$

Hence for $y \in C$, i.e. $y = \pi_C(y)$, the proof in Lemma 5 showed that

$$\langle u_C(x), y - \pi_C(x)\rangle \leq 0,$$

implying C is contained in the half space $H_b^-(u_C(x))$ where $b = \langle u_C(x), \pi_C(x)\rangle$ and obviously $\pi_C(x) \in H_b(u_C(x))$. **QED**

This corollary leads to a foundational aspect of convex geometry: given a closed convex set C, for every $x \in \partial C$, there exists a hyperplane $H_b(u)$ containing x, such that $C \subset H_b^-(u)$. We call such a hyperplane, $H_b(u)$, a *supporting hyperplane of C at x*. Thus Corollary 6 can be restated as, for every $x \in \pi_C(\mathbb{R}^n \setminus C) \subset \partial C$ there exists a supporting hyperplane of C at x. Therefore, to show this is true for every point in ∂C we need only show the following lemma:

Lemma 7. *For a closed convex set C, $\pi_C(\mathbb{R}^n \setminus C) = \partial C$.*

Proof. It is clear that $\pi_C(\mathbb{R}^n \setminus C) \subset \partial C$. Thus consider $x \in \partial C$. Define a sequence $x_k \in \mathbb{R}^n \setminus C$ such that $|x - x_k| < \frac{1}{k}$. Then considering the unit vector, $u_C(x_k)$, and by the compactness of $\mathbb{S}^{n-1}$, there will exist a subsequence, $u_C(x_{k_i})$ converging to some $y \in \mathbb{S}^{n-1}$. Moreover by Corollary 6 it is clear that $\pi_C\big(\pi_C(x_{k_i}) + u_C(x_{k_i})\big) = \pi_C(x_{k_i})$. Hence as $x \in \partial C$ implies $\pi_C(x) = x$ and the 1-Lipschitzness of π_C we have,

$$|x - \pi_C\big(\pi_C(x_{k_i}) + u_C(x_{k_i})\big)| = |\pi_C(x) - \pi_C(x_{k_i})| \leq |x - x_{k_i}| < \frac{1}{k_i}.$$

Therefore by the continuity of the metric projection, letting $i \to \infty$ yields $x = \pi_C(x + y)$. Given that

$$\operatorname{dist}\big(\pi_C(x_k) + u_C(x_k), C\big) = |\pi_C(x_k) + u_C(x_k) - \pi_C(x_k)| = |u_C(x_k)| = 1$$

this implies $\operatorname{dist}(x + y, C) = 1$ and hence $x + y \in \mathbb{R}^n \setminus C$ concluding the proof. **QED**

Combining Corollary 6 and Lemma 7 yields the fundamental result,

Theorem 8. *Given a closed convex set $C \subset \mathbb{R}^n$, for every $x \in \partial C$ there exists a supporting hyperplane.*

A simple consequence of this theorem is that a closed convex set is uniquely defined by its supporting hyperplanes.

Corollary 9. *Given a closed convex set $C \subset \mathbb{R}^n$, we have*

$$C = \bigcap \{H_b^-(u) \subset \mathbb{R}^n : H_b(u) \textit{ is a supporting hyperplane of } C\}.$$

In general it is not guaranteed that a supporting hyperplane at x is unique. In fact uniqueness of a supporting hyperplane at a point of a convex set will play a crucial role later on as we discuss differentiability of convex functions.

Consider a closed convex set $C \subset \mathbb{R}^n$. Then it is a well known result (see for example [16, Theorem 4.1.1]) that for any $x_0 \notin C$, there exists $u \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that the hyperplane $H_b(u)$ separates x_0 and C, where we say the hyperplane $H_b(u)$ separates C and x_0 if

$$C \subset \{x \in \mathbb{R}^n : \langle x, u \rangle < b\} \quad \text{and} \quad \langle x_0, u \rangle > b.$$

Corollary 10. *Let $C \subset \mathbb{R}^n$ be a closed convex set and $x_0 \notin C$. Then there exists a hyperplane, $H_b(u)$, separating C and x_0.*

Proof. By Theorem 8, as $\pi_C(x_0) \in C$, there exists $u \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ such that $H_b(u)$ is a supporting hyperplane of C at $\pi_C(x_0)$. Moreover as $x_0 \neq \pi_C(x_0)$,

$$\operatorname{dist}(x_0, H_b(u)) \geq |x_0 - \pi_C(x_0)| > 0$$

so that the hyperplane with outer normal u through the midpoint of the line segment $[\pi_C(x_0), x_0]$ will separate C and x_0. **QED**

2.4 Convex Functions

Now that we have established the fundamental properties of convex sets we turn to the study of convex functions. Let $U \subset \mathbb{R}^n$ be a convex set. We say $f : U \to \mathbb{R}$ is a *convex function* if for every $x, y \in U$ and $t \in [0, 1]$

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

In other words, a function is convex if every line segment connecting two points on the graph of f lies above the graph of f. It is then clear that there is a connection between convex functions and convex sets, given both are founded upon the use of line segments. To accomplish this connection, for $f : U \to \mathbb{R}$, with $U \subset \mathbb{R}^n$ convex, we define the *epigraph of a function*, denoted by $\text{epi}(f)$, as

$$\text{epi}(f) := \{(x, t) \in U \times \mathbb{R} : f(x) \leq t\}.$$

Note, the graph of f is contained in the boundary of its epigraph and if U is closed then $\text{epi}(f)$ is closed. Therefore an equivalent definition for a convex function can be found by converting the function to a set and showing the set is convex. So we can say for $U \subset \mathbb{R}^n$ convex, a function $f : U \to \mathbb{R}$ is convex if and only if its epigraph, $\text{epi}(f)$, is a convex subset of $U \times \mathbb{R}$.

We observe that the definition of a convex function requires only the convex combination of two points on the graph of f, but viewing a convex function as a function with a convex epigraph we see that there is no reason to restrict ourselves to two points. The next result says that in fact a function is convex if any convex combination of points on the graph of f is in the epigraph of f.

Theorem 11 (Jensen's Inequality)**.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for any $\lambda_i \geq 0$, $i = 1, ..., m$ which satisfy $\sum_{i=1}^m \lambda_i = 1$ and for any elements $x_i \in \mathbb{R}^n$, $i = 1, ..., m$, it holds that*

$$f\left(\sum_{i=1}^m \lambda_i x_i\right) \leq \sum_{i=1}^m \lambda_i f(x_i)$$

Proof. This follows by noting that

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) = f\left(\lambda_1 x_1 + (\lambda_2 + \cdots + \lambda_m)\left(\frac{\lambda_2 x_2}{\lambda_2 + \cdots + \lambda_m} + \cdots + \frac{\lambda_m x_m}{\lambda_2 + \cdots + \lambda_m}\right)\right),$$

applying the definition of convexity, and then an induction argument. **QED**

A similar characterization of convex functions can be given if we know that a function is continuous and midpoint convex.

Theorem 12. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be continuous. Then g is convex if and only if for all $x, y \in \mathbb{R}^n$,*

$$g\left(\frac{x+y}{2}\right) \leq \frac{g(x)}{2} + \frac{g(y)}{2}. \tag{15}$$

Proof. If g is convex, then (15) follows from the definition. Suppose that g satisfies (15) and fix $x, y \in \mathbb{R}^n$. Then for any $t \in (0,1)$ we can approximate $(1-t)x + ty$ with elements of the form

$$\frac{k_n}{2^n}x + \frac{m_n}{2^n}y \quad k_n, m_n \in \mathbb{N}$$

such that $\frac{k_n}{2^n} + \frac{m_n}{2^n} = 1$, $\frac{k_n}{2^n} \to (1-t)$ and $\frac{m_n}{2^n} \to t$. We then apply an induction argument using (15) to show that

$$g\left(\frac{k_n}{2^n}x + \frac{m_n}{2^n}y\right) \leq \frac{k_n}{2^n}g(x) + \frac{m_n}{2^n}g(y)$$

and by the continuity of g, letting $n \to \infty$ proves g is convex. **QED**

We now establish that the slopes of secant lines for any convex function $f : \mathbb{R}^n \to \mathbb{R}$ are non-decreasing.

Proposition 13. *Let $U \subset \mathbb{R}^n$ be convex and $f : U \to \mathbb{R}$, a convex function. Then for any $x, z \in U$ and any $y \in [x, z]$,*

$$\frac{f(y) - f(x)}{|y - x|} \leq \frac{f(z) - f(x)}{|z - x|}. \tag{16}$$

Proof. Let $t \in (0,1)$ such that $y = (1-t)x + tz$. Then by definition of convexity of f,

$$f(y) \leq tf(z) + (1-t)f(x) \iff f(y) - f(x) \leq tf(z) - tf(x).$$

Then dividing both sides of the inequality by $t|z-x|$ yields,

$$\frac{f(y)-f(x)}{t|z-x|} \leq \frac{f(z)-f(x)}{|z-x|}.$$

The result follows by noting that $y - x = t(z-x)$. **QED**

Using this fact we can show that a convex function $f : U \to \mathbb{R}$ is continuous and in fact locally Lipschitz continuous.

Theorem 14. *Let $U \subset \mathbb{R}^n$ be convex and $f : U \to \mathbb{R}$ convex. Then f is locally Lipschitz continuous on* $\operatorname{int} U$ *with,*

$$\operatorname{Lip}(f, \bar{B}(x,r)) \leq \frac{\operatorname{osc}(f, \bar{B}(x,2r))}{r} \quad \text{for every } \overline{B}(x,2r) \subset U$$

where we define the oscillation of f on the set E by,

$$\operatorname{osc}(f,E) = \sup_{x,y \in E} |f(x) - f(y)|.$$

Proof. First note that a convex function is locally bounded in $\operatorname{int} U$. Indeed, for any $x \in \operatorname{int} U$ there exists a neighborhood V_x contained in the interior of a simplex with vertices in U and hence the function is bounded on V_x by values on the vertices since any point in the simplices is their convex combination. Let $y, z \in \bar{B}(x,r)$. Without loss of generality, we may assume that $f(z) \geq f(y)$. Let u be the intersection of $\partial B(x,2r)$ with the ray from y to z. Then $|u-y| \geq r$ and as the difference quotients of a convex function of one variable are increasing, by Proposition 13, we have the following inequality,

$$\frac{f(z)-f(y)}{|z-y|} \leq \frac{f(u)-f(y)}{|u-y|} \leq \frac{\operatorname{osc}(f, \overline{B}(x,2r))}{r}.$$

Taking the supremum over $y, z \in \overline{B}(x,r)$ yields the result. **QED**

Now that we can see a convex function through the lens of a convex set we return to our previous study of hyperplanes and convex sets. Consider $f : \mathbb{R}^n \to \mathbb{R}$ convex, and recall that $\text{epi}(f)$ is a closed convex set. Then by Theorem 8 we know that for every $(x, y) \in \partial(\text{epi}(f))$ there will exist a supporting hyperplane. Moreover, as f is defined on all of $\mathbb{R}^n$, we see that $\partial(\text{epi}(f)) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x) = y\}$, the graph of f. Therefore for each $x \in \mathbb{R}^n$ there will exist a hyperplane passing through $\Gamma_f(x)$ such that $\text{epi}(f)$ will be contained above this hyperplane. This leads to the following characterization of convex functions.

Theorem 15. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $x \in \mathbb{R}^n$, then there is $v \in \mathbb{R}^n$ such that*

$$f(y) \geq f(x) + \langle v, y - x \rangle \quad \textit{for all } y \in \mathbb{R}^n. \tag{17}$$

We define the *subdifferential of f* to be the set,

$$\partial f(x) := \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x \rangle \text{ for all } y \in \mathbb{R}^n\}$$

so we see that Theorem 15 shows, for a convex function $f : \mathbb{R}^n \to \mathbb{R}$, $\partial f(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$. Later we will see that the subdifferential of a convex function is intimately linked with differentiability properties of a convex function. It is easy to verify, by the definition that $\partial f(x)$ is convex, but moreover we have:

Theorem 16. *If $K \subset \mathbb{R}^n$ is compact and $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then*

$$\partial f(K) := \bigcup_{x \in K} \partial f(x)$$

is compact.

Proof. As $\partial f(K) \subset \mathbb{R}^n$ we need only show it is closed and bounded. To show $\partial f(x)$ is bounded suppose to the contrary that there is a sequence $x_k \in K$ and $\sigma_k \in \partial f(x_k)$ such that $|\sigma_k| \to \infty$ as $k \to \infty$. As K and $\mathbb{S}^{n-1}$ are both compact by taking subsequences we may assume that there exist $x \in K$ and $\sigma \in \mathbb{S}^{n-1}$ such that

$$x_k \to x \quad \text{and} \quad \frac{\sigma_k}{|\sigma_k|} \to \sigma \quad \text{as } k \to \infty.$$

Thus by the convexity of f and (17) we have,

$$f\left(x_k + \frac{\sigma_k}{|\sigma_k|}\right) \geq f(x_k) + \left\langle \sigma_k, \left(x_k + \frac{\sigma_k}{|\sigma_k|}\right) - x_k \right\rangle = f(x_k) + |\sigma_k|.$$

By the assumption we know that $f(x_k)+|\sigma_k| \to \infty$ as $k \to \infty$ but we also have by continuity $f(x_k + \frac{\sigma_k}{|\sigma_k|}) \to f(x+\sigma)$ implying that $f(x+\sigma) = \infty$ a clear contradiction. Therefore, $\partial f(K)$ is bounded. To show $\partial f(K)$ is closed let $\sigma_k \in \partial f(K)$ be such that for each k, $\sigma_k \in \partial f(x_k)$, with $x_k \in K$ and $\sigma_k \to \sigma$ as $k \to \infty$. Again, by compactness of K, taking a subsequence, we can assume that $x_k \to x \in K$. Thus for all $z \in \mathbb{R}^n$ we have,

$$f(z) \geq f(x_k) + \langle \sigma_k, z - x_k \rangle$$

and letting $k \to \infty$ implies

$$f(z) \geq f(x) + \langle \sigma, z - x \rangle$$

which shows that $\sigma \in \partial f(x) \subset \partial f(K)$ and hence $\partial f(K)$ is closed. **QED**

2.5 Differentiability Properties of Convex Functions

As we have discussed in the previous sections, convex sets and functions are defined using purely geometric tools. Despite their geometric nature, convex functions have many astounding properties related to differentiability. The next corollary follows from applying the Rademacher theorem to Theorem 14.

Corollary 17. *Convex functions are differentiable almost everywhere.*

This result can be strengthened even further as restricting a convex function to the set of points where it is differentiable shows that it is in fact continuously differentiable, in the sense of sequential continuity.

Theorem 18. *Let D be the set of points where a convex function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then $\nabla f\big|_D$ is continuous.*

Proof. Assume that f is differentiable at x. We will prove that if f is differentiable at y_k and $y_k \to x$ as $k \to \infty$, then $\nabla f(y_k) \to \nabla f(x)$ as $k \to \infty$. Let

$$h(y) = f(y) - f(x) - \nabla f(x) \cdot (y - x)$$

and note that

$$\nabla h(y_k) = \nabla f(y_k) - \nabla f(x).$$

Thus we need only show that $\nabla h(y_k) \to 0$ as $k \to \infty$. As h is the sum of a convex function and a linear function, it is convex. If $y_k \in B(x, r_k)$ where $r_k \to 0$ as $k \to \infty$, then Theorem 14 and the convexity of h yields,

$$|\nabla h(y_k)| \leq \mathrm{Lip}(h, \overline{B}(x, r_k)) \leq \frac{\mathrm{osc}(h, \overline{B}(x, 2r_k))}{r_k} \leq \frac{2\sup(|h|, \overline{B}(x, 2r_k))}{r_k}$$

As the supremum is taken over a compact set we can find $z_k \in \overline{B}(x, 2r_k)$ such that,

$$|h(z_k)| = \sup(|h|, \overline{B}(x, 2r_k))$$

where $r_k \to 0$ implies $z_k \to x$ as $k \to \infty$. Hence we have,

$$\begin{aligned} |\nabla h(y_k)| &\leq \frac{2|h(z_k)|}{r_k} = \frac{4|f(z_k) - f(x) - \nabla f(x) \cdot (z_k - x)|}{2r_k} \\ &\leq \frac{4|f(z_k) - f(x) - \nabla f(x) \cdot (z_k - x)|}{|z_k - x|} \to 0 \quad \text{as } k \to \infty \end{aligned}$$

where the limit follows from the differentiablity of f at x. **QED**

Corollary 19. *If a convex function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable everywhere, then f is of class C^1.*

A standard result in advanced calculus says that the continuity of the partial derivatives f at x implies the differentiability of f at x. For a convex function the existence of partials is all that is needed.

Lemma 20. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. If partial derivatives $\frac{\partial f}{\partial x_i}(x)$ exist for all $i = 1, ..., n$, then f is differentiable at x.*

Proof. Let,

$$A = \left[\frac{\partial f}{\partial x_1}(x), ..., \frac{\partial f}{\partial x_n}(x)\right]$$

and define $\phi : \mathbb{R}^n \to \mathbb{R}$ by,

$$\phi(h) = f(x+h) - f(x) - Ah.$$

We need to show that

$$\frac{\phi(h)}{|h|} \to 0 \quad \text{ as } h \to 0$$

proving that f is differentiable at x with $\nabla f(x) = A$. Note that ϕ is convex being the sum of a convex function and linear function. Also, the following inequality follows from the Cauchy-Schwarz inequality,

$$\sum_{i=1}^{n} u_i v_i = u \cdot v \leq |u||v| \leq |u| \sum_{i=1} |v_i|. \tag{18}$$

Let $\{e_1, \ldots, e_n\}$ be the standard basis in $\mathbb{R}^n$ so that,

$$h = (h_1, ..., h_n) = h_1 e_1 + \cdots + h_n e_n$$

and thus by Jensen's inequality and (18),

$$\phi(h) = \phi\left(\frac{1}{n}\sum_{i=1}^{n} n h_i e_i\right) \leq \frac{1}{n}\sum_{i=1}^{n} \phi(n h_i e_i) = \sum_{i:h_i \neq 0} h_i \left(\frac{\phi(n h_i e_i)}{n h_i}\right) \leq |h| \sum_{i:h_i \neq 0} \left|\frac{\phi(n h_i e_i)}{n h_i}\right|.$$

In a similar fashion we have,

$$\begin{aligned}\phi(-h) = \phi\left(\frac{1}{n}\sum_{i=1}^{n} -n h_i e_i\right) &\leq \frac{1}{n}\sum_{i=1}^{n} \phi(-n h_i e_i) = \sum_{i:h_i \neq 0} -h_i \left(\frac{\phi(-n h_i e_i)}{-n h_i}\right) \\ &\leq |h| \sum_{i:h_i \neq 0} \left|\frac{\phi(-n h_i e_i)}{-n h_i}\right|.\end{aligned}$$

The convexity of ϕ and the fact that $\phi(0) = 0$ yields,

$$0 = \phi\left(\frac{h + (-h)}{2}\right) \leq \frac{1}{2}\left(\phi(h) + \phi(-h)\right)$$

and thus,

$$-|h| \sum_{i:h_i \neq 0} \left|\frac{\phi(-n h_i e_i)}{-n h_i}\right| \leq -\phi(-h) \leq \phi(h) \leq |h| \sum_{i:h_i \neq 0} \left|\frac{\phi(n h_i e_i)}{n h_i}\right|.$$

This in turn implies that,

$$-\sum_{i:h_i\neq 0}\left|\frac{\phi(-nh_ie_i)}{-nh_i}\right| \leq \frac{\phi(h)}{|h|} \leq \sum_{i:h_i\neq 0}\left|\frac{\phi(nh_ie_i)}{nh_i}\right|. \tag{19}$$

Now as,

$$\begin{aligned}\lim_{t\to 0}\frac{\phi(te_i)}{t} &= \lim_{t\to 0}\frac{f(x+te_i)-f(x)-A(te_i)}{t} = \lim_{t\to 0}\frac{f(x+te_i)-f(x)}{t} - \frac{\partial f}{\partial x_i}(x)\\ &= \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(x) = 0\end{aligned}$$

we have, by letting $|h|\to 0$ (and thus $h_i \to 0$) that,

$$-\sum_{i:h_i\neq 0}\left|\frac{\phi(-nh_ie_i)}{-nh_i}\right| \to 0 \quad \text{and} \quad \sum_{i:h_i\neq 0}\left|\frac{\phi(nh_ie_i)}{nh_i}\right| \to 0$$

and thus, by (19) we have shown that,

$$\lim_{|h|\to 0}\frac{\phi(h)}{|h|} = 0$$

as desired. **QED**

Theorem 21. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then one-sided partial derivatives exist at every point $x \in \mathbb{R}^n$ and*

$$\frac{\partial^- f}{\partial x_i}(x) \leq \frac{\partial^+ f}{\partial x_i}(x).$$

Moreover, for any $x \in \mathbb{R}^n$ and $1 \leq i \leq n$, if $s \in \mathbb{R}$ satisfies,

$$\frac{\partial^- f}{\partial x_i}(x) \leq s \leq \frac{\partial^+ f}{\partial x_i}(x)$$

then $f(x+te_i) \geq f(x) + st$ for all $t \in \mathbb{R}$.

Proof. The existence of the one sided derivatives at every $x \in \mathbb{R}^n$ follows from the fact that the secant slopes are non-decreasing for a convex function (Proposition 13) and are clearly bounded, by Theorem 16. If $s \in \mathbb{R}$ satisfies $\frac{\partial^- f}{\partial x_i}(x) \leq s \leq \frac{\partial^+ f}{\partial x_i}(x)$, then for all $t < 0$, we have

$$\frac{f(x + te_i) - f(x)}{t} \leq \frac{\partial^- f}{\partial x_i}(x) \leq s$$

and similarly for all $t > 0$,

$$\frac{f(x + te_i) - f(x)}{t} \geq \frac{\partial^+ f}{\partial x_i}(x) \geq s$$

completing the proof. **QED**

Theorem 22. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then the following conditions are equivalent:*

(a) f *is convex*

(b) $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ *for all* $x, y \in \mathbb{R}^n$

(c) $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$ *for all* $x, y \in \mathbb{R}^n$

Remark 23. Condition (b) means that f is bounded below by its tangent spaces and if f is differentiable at x, then $\nabla f(x) \in \partial f(x)$. Condition (c) is called monotonicity of ∇f.

Proof. We first assume (a). Convexity of f implies that

$$\frac{f(x + t(y - x)) - f(x)}{t} \leq f(y) - f(x)$$

for any $t \in (0, 1)$. Passing to the limit as $t \to 0^+$ proves (b).

Now assume (b). It follows that

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \implies f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$$

and similarly,

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \implies f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle.$$

Adding these two inequalities yields (c).

Finally assume (c). To prove f is convex we will show for any $x, y \in \mathbb{R}^n$ the function

$$\phi(t) = f(x + t(y - x))$$

is convex. To this end it suffices to prove that $\phi'(t)$ is increasing. We have,

$$\phi'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle$$

If $t_1 < t_2$, then denoting $x_1 = x + t_1(y - x)$ and $x_2 = x + t_2(y - x)$ yields

$$\phi'(t_2) - \phi'(t_1) = \langle \nabla f(x_2) - \nabla f(x_1), y - x \rangle = \frac{1}{t_2 - t_1} \langle \nabla f(x_2) - \nabla f(x_1), x_2 - x_1 \rangle \geq 0$$

as desired. **QED**

There is a strong connection between the subdifferential of a convex function, $\partial f(x)$, and its differentiability. We will explore this more in Theorem 27 but before stating and proving the theorem we will need a convex version of the Hahn Banach theorem. The proof of the Hahn Banach theorem requires Zorn's lemma, so we first establish the necessary definitions for stating Zorn's lemma.

Definition 24. We define a *partially ordered set* X as a set equipped with a binary relation $\leq$ that satisfies the properties:

(1) $x \leq x$ for all $x \in X$;

(2) $x \leq y$ and $y \leq x$ implies $x = y$;

(3) $x \leq y$ and $y \leq z$ implies $x \leq z$.

We say a subset $C \subset X$ is a *totally ordered set* if for every $x, y \in C$, either $x \leq y$ or $y \leq x$. An *upper bound* of $Y \subset X$, where X is a partially ordered set, is an element $m \in X$ such that $y \leq m$ for all $y \in Y$. We say $m \in X$ is a *maximal element* of X if $m \leq x$ for some $x \in X$ implies $m = x$.

Theorem 25 (Zorn's Lemma)**.** *Let X be a non empty partially ordered set. If every totally ordered subset of X has an upper bound, then X has at least one maximal element.*

The classic Hahn-Banach theorem states that a linear functional can be extended if bounded above by a 1-homogeneous and subadditive function, but we can weaken this to a convex function. Proofs of Hahn Banach using a convex majorant can be hard to find but we present here one given in [26, Theorem A, pg. 105].

Theorem 26 (Hahn-Banach Theorem)**.** *Let X be a real linear space and $p : X \to \mathbb{R}$ a convex function. Let $\lambda : Y \to \mathbb{R}$ be a linear functional defined on a linear subspace $Y \subset X$ such that $\lambda(v) \le p(v)$ for all $v \in Y$. Then there is a linear functional $\Lambda : X \to \mathbb{R}$ such that*

$$\Lambda(v) \le p(v) \textit{ for all } v \in X \quad \textit{and} \quad \Lambda(v) = \lambda(v) \textit{ for all } v \in Y.$$

Proof. Let $z \notin Y$ and define $\overline{Y} := \text{span}\{Y, z\}$. We will first show we can extend λ to $\overline{Y}$. Let $\bar{\lambda}$ be the linear extension of λ to $\overline{Y}$. To do this we need only define $\bar{\lambda}(z)$ as by the linearity of $\bar{\lambda}$ we have for $a \in \mathbb{R}$ and $y \in Y$ (and thus $az + y \in \overline{Y}$),

$$\bar{\lambda}(az + y) = a\bar{\lambda}(z) + \bar{\lambda}(y) = a\bar{\lambda}(z) + \lambda(y).$$

Suppose that $y_1, y_2 \in Y$ and $\alpha, \beta > 0$. Then,

$$\begin{aligned}
\beta\lambda(y_1) + \alpha\lambda(y_2) = \lambda(\beta y_1 + \alpha y_2) &= (\alpha + \beta)\lambda\left(\frac{\beta}{\alpha + \beta}y_1 + \frac{\alpha}{\alpha + \beta}y_2\right) \\
&\le (\alpha + \beta)\rho\left(\frac{\beta}{\alpha + \beta}y_1 + \frac{\alpha}{\alpha + \beta}y_2\right) \\
&= (\alpha + \beta)\rho\left(\frac{\beta}{\alpha + \beta}(y_1 - \alpha z) + \frac{\alpha}{\alpha + \beta}(y_2 + \beta z)\right) \\
&\le \beta\rho(y_1 - \alpha z) + \alpha\rho(y_2 + \beta z)
\end{aligned}$$

which in turn implies that,

$$\beta(\lambda(y_1) - \rho(y_1 - \alpha z)) \le \alpha(\rho(y_2 + \beta z) - \lambda(y_2)).$$

As $\alpha, \beta > 0$, we have

$$\frac{1}{\alpha}(\lambda(y_1) - \rho(y_1 - \alpha z)) \le \frac{1}{\beta}(\rho(y_2 + \beta z) - \lambda(y_2)).$$

Then we can find $a \in \mathbb{R}$ such that,

$$\sup_{\substack{y \in Y \\ \alpha > 0}} \frac{1}{\alpha}(\lambda(y) - \rho(y - \alpha z)) \le a \le \inf_{\substack{y \in Y \\ \alpha > 0}} \frac{1}{\alpha}(\rho(y + \alpha z) - \lambda(y)). \tag{20}$$

We now define $\bar{\lambda}(z) = a$ and we need to show that the given extension satisfies the property $\bar{\lambda}(x) \le \rho(x)$ for all $x \in \overline{Y}$. By (20) we know for all $\alpha > 0$ and $y \in Y$ that,

$$\bar{\lambda}(z) \le \frac{1}{\alpha}(\rho(y + \alpha z) - \lambda(y))$$

which, by the linearity of $\bar{\lambda}$ and the fact that $\lambda = \bar{\lambda}$ on Y, implies that

$$\bar{\lambda}(y + \alpha z) \leq \rho(y + \alpha z).$$

Similarly using (20) again, for all $\alpha > 0$ and $y \in Y$, we have

$$\bar{\lambda}(z) \geq \frac{1}{\alpha}(\lambda(y) - \rho(y - \alpha z))$$

which again implies that,

$$\bar{\lambda}(y - \alpha z) \leq \rho(y - \alpha z)$$

and as every element of $\overline{Y}$ is of the form $y \pm \alpha z$ we have shown the extension $\bar{\lambda}$ satisfies the required property $\bar{\lambda}(x) \leq \rho(x)$ for all $x \in \overline{Y}$. This shows that we can extend λ by one dimension. The remainder of the proof follows from Zorn's lemma (Theorem 25). Let E be the set of extensions, e, of λ which satisfy $e(x) \leq \rho(x)$ for all x where e is defined. We assign a partial ordering of E by letting $e_1 \preceq e_2$ if e_2 is defined on a larger set than e_1 and $e_1(x) = e_2(x)$ for all x where they are both defined. Thus we can let $\{e_\alpha\}_{\alpha \in A}$ be a totally ordered subset of E where each e_α is defined on a set X_α. Hence we can define e on the set

$$X = \bigcup_{\alpha \in A} X_\alpha$$

by letting $e(x) = e_\alpha(x)$ whenever $x \in X_\alpha$. It is obvious that for all $\alpha \in A$, $e_\alpha \preceq e$ so every totally ordered subset of E has an upper bound. Therefore by Zorn's lemma, E has a maximal element Λ defined on some set X' with the property that $\Lambda(x) \leq \rho(x)$ for all $x \in X'$. If $X' \neq X$ then by the first part of the proof we can extend Λ to $\overline{\Lambda}$ defined on a space whose dimension is $\dim(X') + 1$, contradicting the maximality of Λ. Thus $X' = X$ and Λ is defined on all of X with the properties that $\Lambda(x) = \lambda(x)$ for $x \in Y$ and $\Lambda(x) \leq \rho(x)$ for all $x \in X$. **QED**

We recall by Theorem 15 that a convex function has a nonempty subdifferential at every point in its domain, and that by Corollary 17 it is differentiable almost everywhere. The following theorem shows us that the subdifferential of a convex function is an extension of the derivative of a convex function.

Theorem 27. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex.*

(1) If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$;

(2) f is differentiable at x if and only if $\partial f(x)$ is a singleton

Proof. By Theorem 22, we know that $\nabla f(x) \in \partial f(x)$. We now wish to show that this is the only point. Suppose to the contrary that there exists $v \in \partial f(x)$ such that $v \neq \nabla f(x)$. Then there exists some $v_i \neq \frac{\partial f}{\partial x_i}(x)$, and without loss of generality we may assume that $v_i < \frac{\partial f}{\partial x_i}(x)$. Then we consider the function,

$$g(t) := f(x + te_i)$$

and note that $\frac{\partial f}{\partial x_i}(x), v_i \in \partial g(0)$. Moreover, $g'(0) = \frac{\partial f}{\partial x_i}(x)$. As $v_i < g'(0)$ we have, by Proposition 13, that in fact,

$$g'_{-}(0) \leq v_i < g'(0)$$

where $g'_{-}(0)$ is the left sided derivative of g at 0. But this contradicts the differentiability of g as $g'_{-}(0) = g'_{+}(0) = g'(0)$.

To prove (2), we note that (1) shows if f is differentiable at x, then $\partial f(x)$ is a singleton. It suffices to show that if $\partial f(x) = \{v\}$, then

$$\frac{\partial f}{\partial x_i}(x) = v_i$$

proving that all the partials exist and thus by Lemma 20, f is differentiable at x. Let $g(t) := f(x + te_i)$ and by the previous comment we need to show that $g'(0) = v_i$. Clearly as $\partial f(x) = \{v\}$, we have $v_i \in \partial g(0)$ and thus,

$$g(t) \geq g(0) + v_i t \quad \text{for all } t \in \mathbb{R}.$$

By the convexity of g and Theorem 21 we know the one sided derivatives exist, i.e.

$$g'_{+}(0) = \lim_{t \to 0^+} \frac{g(t) - g(0)}{t} \quad \text{and} \quad g'_{-}(0) = \lim_{t \to 0^-} \frac{g(t) - g(0)}{t}.$$

Thus to show the differentiability of g at 0 it suffices to prove that $g'_{-}(0) = g'_{+}(0)$ which, by the first part, will imply $g'(0) = v_i$ as desired. Suppose to the contrary that,

$$g'_{-}(0) \leq a < b \leq g'_{+}(0).$$

Then by the second part of Theorem 21, we have that for all $t \in \mathbb{R}$,

$$g(t) \geq g(0) + at.$$

Let $Y = \operatorname{span}\{e_i\}$ and thus for $w = te_i \in Y$ we can rewrite the above inequality as,

$$at \leq f(x + te_i) - f(x).$$

Equivalently we have,

$$(ae_i) \cdot w \leq f(x + w) - f(x).$$

Setting $\lambda(w) = (ae_i) \cdot w$ and $p(w) = f(x + w) - f(x)$ we can see that $\lambda(w)$ is linear, $p(w)$ is convex, and $\lambda(w) \leq p(w)$. Thus by the Hahn Banach theorem there exists $A \in \mathbb{R}^n$ such that

$$A \cdot u \leq f(x + u) - f(x)$$

for all $u \in \mathbb{R}$, where $A \cdot e_i = a$. In a similar fashion we have,

$$g(t) \geq g(0) + bt$$

and thus by the Hahn Banach theorem there exists $B \in \mathbb{R}^n$ such that

$$B \cdot u \leq f(x + u) - f(x)$$

for $u \in \mathbb{R}^n$ and $B \cdot e_i = b$. As $A \cdot e_i \neq B \cdot e_i$ we know that $A \neq B$ but as

$$f(x + u) \geq f(x) + A \cdot u$$

and

$$f(x + u) \geq f(x) + B \cdot u$$

this implies $A, B \in \partial f(x)$ a contradiction of the assumption that $\partial f(x) = \{v\}$. **QED**

Theorem 28. *If $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex functions and g is differentiable at x, then*

$$\partial(f + g)(x) = \partial f(x) + \nabla g(x).$$

Proof. It suffices to show that $\partial(f+g)(x) \subset \partial f(x)+\nabla g(x)$ as the opposite inclusion follows immediately from the definition. Let $v \in \partial(f+g)(x)$, that is, for any $z \in \mathbb{R}^n$

$$f(z)+g(z) \geq f(x)+g(x)+\langle v, z-x\rangle. \tag{21}$$

As g is differentiable at x, we have

$$g(z) = g(x) + \langle \nabla g(x), z-x\rangle + o(|z-x|)$$

and combining this with (21) yields

$$f(z)+g(x)+\langle \nabla g(x), z-x\rangle + o(|z-x|) \geq f(x)+g(x)+\langle v, z-x\rangle.$$

This in turn implies

$$f(z) \geq f(x) + \langle v-\nabla g(x), z-x\rangle + o(|z-x|). \tag{22}$$

Let $h(z) = f(z) - \langle v - \nabla g(x), z-x\rangle$ and note that h is convex, being the sum of convex functions, and $h(x) = f(x)$. Thus rewriting (22) gives us,

$$h(z) \geq h(x) + o(|z-x|).$$

It then follows from the convexity of h that $h(z) \geq h(x)$ for all $z \in \mathbb{R}^n$. Thus,

$$f(z) \geq f(x) + \langle v-\nabla g(x), z-x\rangle$$

proving that $v - \nabla g(x) \in \partial f(x)$. Therefore

$$v = (v-\nabla g(x)) + \nabla g(x) \in \partial f(x) + \nabla g(x)$$

and the proof is complete. **QED**

We conclude the section concerning the differentiability properties of convex functions with a generalization of Theorem 18, namely that the subdifferential of a convex function is sequentially continuous at every point where ∇f exists.

Theorem 29. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable at x. If $y_k \to x$ and $\sigma_k \in \partial f(y_k)$, then $\sigma_k \to \nabla f(x)$.*

Remark 30. We can see this is a generalization of Theorem 21 as if f is differentiable, by Theorem 27, then $\sigma_k = \nabla f(y_k)$.

Proof. Define $h : \mathbb{R}^n \to \mathbb{R}$ as,

$$h(z) = f(z) - f(x) - \nabla f(x) \cdot (z - x).$$

Then, $f(z) \geq f(y_k) + \sigma_k \cdot (z - y_k)$ implies that

$$\begin{aligned} h(z) &= f(z) - f(x) - \nabla f(x)(z - x) \\ &\geq f(y_k) + \sigma_k \cdot (z - y_k) - f(x) - \nabla f(x) \cdot (z - x) \\ &= f(y_k) - f(x) - \nabla f(x) \cdot (y_k - x) + (\sigma_k - \nabla f(x)) \cdot (z - y_k) \\ &= h(y_k) + (\sigma_k - \nabla f(x)) \cdot (z - y_k) \end{aligned}$$

which shows that $\sigma_k - \nabla f(x) \in \partial h(y_k)$. Thus if we can show that $\xi_k \in \partial h(y_k)$ implies $\xi_k \to 0$ as $k \to \infty$, then we will have shown that $\sigma_k \to \nabla f(x)$. Note that $h(z)$ is convex being the sum of a convex and linear function. As $y_k \to x$, choose $r_k \to 0$ such that $y_k \in B(x, r_k)$. Then we know that

$$|\xi_k| \leq \text{Lip}(h, \overline{B}(x, r_k)). \tag{23}$$

Indeed, as

$$h(z) \geq h(y_k) + \xi_k \cdot (z - y_k)$$

and letting $z = y_k + \varepsilon u$ for $|u| = 1$ and $\varepsilon > 0$, we have

$$\varepsilon(\xi_k \cdot u) \leq h(y_k + \varepsilon u) - h(y_k) \leq \text{Lip}(h, \overline{B}(x, r_k))\varepsilon$$

provided ε is small enough so that $y_k + \varepsilon u \in B(x, r_k)$. Thus

$$\xi_k \cdot u \leq \text{Lip}(h, \overline{B}(x, r_k))$$

and taking the supremum over all $|u| = 1$ yields (23). The remainder of the proof follows the proof of Theorem 18 though we include the details here for completeness sake. By Theorem 14,

$$|\xi_k| \leq \text{Lip}(h, \overline{B}(x, r_k)) \leq \frac{\text{osc}(h, \overline{B}(x, 2r_k))}{r_k} \leq \frac{2\sup(|h|, \overline{B}(x, 2r_k))}{r_k}.$$

As the supremum is taken over a compact set we can find $z_k \in \mathbb{R}^n$ such that,

$$|h(z_k)| = \sup(|h|, \overline{B}(x, 2r_k))$$

where $r_k \to 0$ implies $z_k \to x$ as $k \to \infty$. Hence we have,

$$\begin{aligned} |\nabla h(y_k)| &\le \frac{2|h(z_k)|}{r_k} = \frac{4|f(z_k) - f(x) - \nabla f(x) \cdot (z_k - x)|}{2r_k} \\ &\le \frac{4|f(z_k) - f(x) - \nabla f(x) \cdot (z_k - x)|}{|z_k - x|} \to 0 \quad \text{as } k \to \infty \end{aligned}$$

where the limit follows from the differentiablity of f at x. **QED**

3.0 Differentiability Properties of the Boundaries of Convex Bodies

While convex functions naturally have many nice differentiability properties, convex sets on their own are generally too simple to conduct any sort of analysis with them. To overcome this we consider the subset of convex sets called convex bodies. Locally the boundaries of convex bodies are graphs of convex functions and hence we can identify differentiability properties of convex bodies by studying the differentiability of these locally defined functions. Specifically in this chapter we will explore how the differentiability of convex bodies is related to the relationship of closed balls in the interior of the convex body and the boundary.

3.1 Convex Bodies and Convex Domains

We say a convex domain is a convex set with non-empty interior and a *convex body* is a compact convex set with non-empty interior. Throughout this thesis a convex domain or body in $\mathbb{R}^n$ will generally be denoted by K. As previously stated, locally the boundary of a convex domain is the graph of a convex function. To see this we consider the *lower-bound function* of K, which is defined by,

$$x' \in \mathbb{R}^{n-1} \mapsto \ell_K(x') := \inf\{t \in \mathbb{R} : (x', t) \in K\} \tag{24}$$

and we can show that if K is convex, then ℓ_K is convex. The lower bound function, as defined, can be found in [16, Theorem 1.3.1].

Proposition 31. *If K is a convex domain, then for any $x \in \partial K$ there exists a convex, open U such that, by rotating, $\partial K \cap U$ is the graph of a convex function.*

Proof. By the definition of ℓ_K, for fixed $\varepsilon > 0$, let $(x_1', t_1), (x_2', t_2) \in K$, where $x_1', x_2' \in \mathbb{R}^{n-1}$ and $t_1, t_2 \in \mathbb{R}$, such that

$$t_1 \leq \ell_K(x_1') + \varepsilon \quad \text{and} \quad t_2 \leq \ell_K(x_2') + \varepsilon. \tag{25}$$

By the convexity of K,

$$(\lambda x_1' + (1-\lambda)x_2', \lambda t_1 + (1-\lambda)t_2) = \lambda(x_1', t_1) + (1-\lambda)(x_2', t_2) \in K$$

and thus, by (25)

$$\ell_K(\lambda x_1' + (1-\lambda)x_2') \leq \lambda t_1 + (1-\lambda)t_2 \leq \lambda \ell_K(x_1') + (1-\lambda)\ell_K(x_2') + 2\varepsilon.$$

Letting $\varepsilon \to 0^+$ shows that ℓ_K is convex. We define $\pi : \mathbb{R}^n \to \mathbb{R}^{n-1}$ as the orthogonal projection given by $\pi(x_1, \ldots, x_n) = (x_1, \ldots, x_{n-1})$. If we fix $x \in \partial K$, by rotating K, we can ensure that $\pi(x) \in \operatorname{int}(\pi(K))$ and $x = (\pi(x), \ell_K(\pi(x)))$. We then note $\pi(x) \in \operatorname{int}(\pi(K))$ implies there exists some $\delta > 0$ such that $B^{n-1}(\pi(x), \delta) \subset \pi(K)$. By decreasing δ, there exists $M > 0$ such that $B^{n-1}(\pi(x), \delta) \times M \subset K$. Thus we can let $U = B^{n-1}(\pi(x), \delta) \times (-\infty, M)$ and $\partial K \cap U$ is the graph of ℓ_K defined on $B^{n-1}(\pi(x), \delta)$. **QED**

Definition 32. For a convex domain K, we say that the boundary of K is of class C^1 if by rotating the set K, the boundary of K can be locally represented as the graph of a C^1 convex function.

Proposition 33. *A convex domain $K \subset \mathbb{R}^n$ has a unique supporting hyperplane at each $x \in \partial K$ if and only if K is a C^1 hypersurface.*

Proof. By Proposition 31 we know locally ∂K is the graph of a convex function. We then apply Theorem 27 and Corollary 19 to each of these locally defined functions to show that they are C^1 convex functions. **QED**

Given a convex body K, there are useful convex bodies we can identify in the interior of K. An important one is the *inner parallel body* of K, defined to be,

$$K_r := \{x \in K : \operatorname{dist}(x, \partial K) \geq r\}. \tag{26}$$

Lemma 34. *Let $K \subset \mathbb{R}^n$ be a convex body. Then K_r is convex for any $r > 0$ and if K contains a ball of radius r_0 then for all $r \in (0, r_0)$, K_r is a convex body.*

Proof. Let $x, y \in K_r$. We need to show that $[x, y] \subset K_r$. Clearly, $\overline{B}(x, r), \overline{B}(y, r) \subset K$ and for any $z \in [x, y]$, $\overline{B}(z, r) \subset \operatorname{co}(\overline{B}(x, r) \cup \overline{B}(y, r)) \subset K$, so $\operatorname{dist}(z, \partial K) \geq r$. Thus, $z \in K_r$, showing $[x, y] \subset K_r$ and proving the convexity of K_r. Moreover if $B(x, r_0) \subset K$, then for any $r \in (0, r_0)$ there exists $\delta > 0$ such that $B(x, r + \delta) \subset K$. Fix $y \in B(x, \delta)$ and let $z \in \partial K$ be such that $\operatorname{dist}(y, \partial K) = |y - z|$. Then there exist $y_\delta \in \partial B(x, \delta) \cap [y, z]$ and $y_r \in \partial B(x, r) \cap [y, z]$, where y_δ and y_r clearly satisfy

$$|y_\delta - y_r| = r - \delta \quad \text{and} \quad |y_r - z| > \delta.$$

Then we can see, as $y_\delta, y_r \in [y, z]$, that,

$$\operatorname{dist}(y, \partial K) = |y - z| \geq |y_\delta - z| = |y_\delta - y_r| + |y_r - z| > r - \delta + \delta = r$$

showing that $y \in K_r$. Therefore $B(x, \delta) \subset K_r$, showing that K_r has non empty interior. **QED**

The other important set inside of a convex body is the union of closed balls of fixed radius defined by,

$$K(R) := \bigcup \{\overline{B}(x, R) : \overline{B}(x, R) \subset K\}. \tag{27}$$

It is clear that if there exists $B(a, R) \subset K(R)$, then $K(R)$ has non empty interior and $K(R)$ is compact. Moreover $K(R)$ is convex, as for any $\overline{B}(x_0, R), \overline{B}(y_), R) \subset K$, we have $\operatorname{co}(\overline{B}(x_0, R) \cup \overline{B}(y_0, R)) \subset K$, so that if $x \in \overline{B}(x_0, R)$ and $y \in \overline{B}(y_0, R)$, i.e. $x, y \in K(R)$, then $[x, y] \subset \operatorname{co}(\overline{B}(x_0, R) \cup \overline{B}(y_0, R) \subset K$. Therefore $K(R)$, for R small enough is a convex body. In fact there is a special relationship between the sets K_r and $K(R)$.

Proposition 35. *Let $K \subset \mathbb{R}^n$ be a convex body. Then,*

$$\partial K_r = \{x \in K : \overline{B}(x, r) \subset K \text{ and } \overline{B}(x, r) \cap \partial K(r) \neq \emptyset\}.$$

Moreover, if $x \in \partial K(r)$, then $x \in \overline{B}(\pi_{K_r}(x), r) \subset K$.

Remark 36. This statement shows that the centers of the balls of radius r tangent to the boundary of K precisely define the boundary of K_r.

Proof. If $z \in \partial K_r = \{x \in K : \text{dist}(x, \partial K) = r\}$, then $\overline{B}(z,r) \subset K$ and moreover, by compactness, there exists some $y \in \partial K$ such that $|z - y| = r$. Thus $\overline{B}(z,r) \subset K$ and $\partial K \cap \overline{B}(z,r) \neq \emptyset$, so that $\partial K_r \subset \{x \in K : \overline{B}(x,r) \subset K \text{ and } \overline{B}(x,r) \cap \partial K \neq \emptyset\}$. If $z\{x \in K : \overline{B}(x,r) \subset K \text{ and } \overline{B}(x,r) \cap \partial K \neq \emptyset\}$, then $B(z,r) \subset K$ implies $z \in K_r$. Also as $\partial K \cap \overline{B}(z,r) \neq \emptyset$, $\text{dist}(z, \partial K) = r$, so that $z \in \partial K_r$.

Finally if $x \in \partial K(r)$, then we know there exists $h(x)$ such that $x \in \overline{B}(h(x), r) \subset K$. Thus $h(x) \in K_r$. As $|x - h(x)| = r$, this shows that in fact $\text{dist}(x, \partial K_r) = r$ and thus $h(x) = \pi_{K_r}(x)$. **QED**

Definition 37. Let $U \subset \mathbb{R}^n$ be open. We say a function $f : U \to \mathbb{R}$ is of class $C^{1,\alpha}$, for $\alpha \in (0,1]$, if $f \in C^1(U)$ and ∇f is α-Hölder continuous, i.e. there exists $L > 0$ such that for every $x, y \in U$,

$$|\nabla f(x) - \nabla f(y)| \leq L|x - y|^\alpha.$$

Note that if $\alpha = 1$, we have $f \in C^{1,1}$ when ∇f is Lipschitz continuous. We now extend the definition of a C^1 convex body to the more general $C^{1,\alpha}$ convex body.

Definition 38. For a convex body K, we say that the boundary of K is of class $C^{1,\alpha}$ for $\alpha \in (0,1]$ if by rotating the set K, the boundary of K can be locally represented as the graph of a C^1 convex function whose gradient is α-Hölder continuous (Lipschitz continuous when $\alpha = 1$).

An equivalent definition for a $C^{1,\alpha}$ convex body can be stated by the following proposition:

Proposition 39. *If the outer unit normal vector of a C^1 convex body K is α-Hölder continuous, then K is of class $C^{1,\alpha}$*

Proof. The boundary of a C^1 convex body is locally the graph of a C^1 convex function. Thus by a rotation and translation we can assume that $0 \in \partial K \subset \{x \in \mathbb{R}^n : x_n \geq 0\}$, and there exist some $L, R > 0$ such that $\partial K \cap (B^{n-1}(0, 2R) \times [0, L])$ is the graph of a C^1 convex function $f : B^{n-1}(0, 2R) \to \mathbb{R}$. We denote $x' = (x_1, \ldots, x_{n-1})$. Note on $B^{n-1}(0, R)$ there exists $M > 0$ such that $|\nabla f| \leq M$ as ∇f is assumed to be continuous on $B^n(0, 2R)$. Thus the inner unit normal vector at $\Gamma_f(x')$, in terms of ∇f, is given by

$$\nu(\Gamma_f(x')) = \frac{(-\nabla f(x'), 1)}{\sqrt{1 + |\nabla f(x')|^2}}, \quad \text{so,} \quad \pi(\nu(\Gamma_f(x')) = \frac{-\nabla f(x')}{\sqrt{1 + |\nabla f(x')|^2}},$$

where $\pi : \mathbb{R}^n \to \mathbb{R}^{n-1}$, $\pi(x', x_n) = x'$ is the orthogonal projection. Defining

$$\Psi(z) := \frac{-z}{\sqrt{1 - |z|^2}} \quad \text{and} \quad \Phi(z) := \frac{-z}{\sqrt{1 + |z|^2}},$$

we see that $\Psi(\Phi(z)) = z$ for all $z \in \mathbb{R}^{n-1}$, and it follows that

$$\nabla f(x') = \Psi\left(\frac{-\nabla f(x')}{\sqrt{1 + |\nabla f(x')|^2}}\right) = \Psi(\pi(\nu(\Gamma_f(x')))) \quad \text{for } x' \in U.$$

This proves that ∇f is α-Hölder continuous on $B^{n-1}(0, R)$, as it is the composition of Lipschitz functions, Ψ and π, with an α-Hölder continuous function, ν. The only issue could be the Lipschitz continuity of Ψ: it is a smooth function defined for $|z| < 1$, but it is unbounded. However, this does not cause any problems here, because

$$\left|\frac{-\nabla f(x')}{\sqrt{1 + |\nabla f(x')|^2}}\right| \leq \frac{M}{\sqrt{1 + M^2}} < 1.$$

QED

3.2 Convex Bodies Satisfying the Uniform Inner Ball Condition

We say that a convex body K satisfies the *r-uniform inner ball condition* if for every $x \in K$ there exists $\overline{B}(h(x), r) \subset K$ such that $x \in \overline{B}(h(x), r)$. In other words K satisfies the $r-$uniform inner ball condition if $K = K(r)$ as defined in 27. Note the uniformity here follows from the radius being independent of any of the individual points in K, and depends only on the convex body itself. The important aspect of the r-uniform inner ball condition is how these closed balls interact with the boundary of K. In this case we have that at every point on the boundary there is a closed ball tangent to the boundary completely contained in K. Thus the curvature of the convex body is controlled by the uniform radius of these closed balls.

The following result is a beautiful characterization of $C^{1,1}$ convex bodies originally proved by Lucas in [21, Theorem 1, pg. 32], in his unpublished thesis. Another proof can be found in [17, Proposition 2.4.3] and both proofs presented in this thesis are the ones given in [3].

Theorem 40. *A convex body K is of class $C^{1,1}$ if and only if there exists $r > 0$ such that K satisfies the r-uniform inner ball condition.*

Later we will show, in Theorem 51, that we can find a similar geometric characterization of $C^{1,\alpha}$ convex bodies. For both Theorem 40 and Theorem 51 we will need the following two lemmas.

Lemma 41. *Let $f \in C^{1,\alpha}(B(z,r))$ with $|\nabla f(x) - \nabla f(y)| \leq L|x-y|^\alpha$ for all $x, y \in B(z,r)$. Then for all $x, y \in B(z,r)$, we have, $f(x) \leq \frac{L}{1+\alpha}|x-y|^{1+\alpha} + \nabla f(y) \cdot (x-y) + f(y)$.*

Proof. The mean value theorem implies

$$f(x) - f(y) = \int_0^1 \nabla f((1-t)y + tx) \cdot (x-y)\, dt$$

and using the Hölder continuity of the derivative we have for all $x, y \in B(z,r)$,

$$\begin{aligned} f(x) - f(y) - \nabla f(y) \cdot (x-y) &= \int_0^1 (\nabla f((1-t)y+tx) - \nabla f(y)) \cdot (x-y)\, dt \\ &\leq \int_0^1 |\nabla f((1-t)y+tx) - \nabla f(y)||x-y|\, dt \\ &\leq L|x-y| \int_0^1 |(1-t)y + tx - y|^\alpha\, dt = \frac{L}{1+\alpha}|x-y|^{1+\alpha}. \end{aligned}$$

QED

Remark 42. Lemma 41 implies that if $f \in C^{1,\alpha}_{\text{loc}}(U)$, where $U \subset \mathbb{R}^n$ is open, then

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + O(|y - x|^{1+\alpha}) \quad \text{for all } x, y \in U. \tag{28}$$

The next lemma shows that the epigraph of a paraboloid satisfies the uniform inner ball condition, and gives an explicit constant for $R > 0$.

Lemma 43. *Let $a \neq 0$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$, and define the paraboloid $p : \mathbb{R}^n \to \mathbb{R}$ by $p(x) = a|x|^2 + b \cdot x + c$. Then* $\text{epi}(p)$ *satisfies the uniform inner ball condition with $R = \frac{1}{2a}$.*

Proof. We first assume that p is of the form $p(x) = a|x|^2$. By examining the behavior of the normal line at every point on the graph of p we can see that the set of points which do not have a unique metric projection onto the graph of p, defined as the medial axis, is equal to the set $\left\{x \in \mathbb{R}^n : x_1 = \cdots = x_{n-1} = 0 \text{ and } x_n \geq \frac{1}{2a}\right\}$. Moreover the distance from any point on the medial axis to the graph of p has distance at least $\frac{1}{2a}$. The result follows by noting that $p(x) = a|x|^2 + b \cdot x + c$ is a translated paraboloid of the form $a|x|^2$. **QED**

Recall, by Theorem 27, that a differentiable convex function has a unique supporting hyperplane at each point on its graph. Next we want to consider the properties of a convex function that is squeezed by a $C^{1,1}$ convex function and its unique supporting hyperplane.

Lemma 44. *Let $f, g : B^n(0, R) \to \mathbb{R}$ be convex functions. If $g \in C^{1,1}$, $f \leq g$ and $f(x) = g(x)$ for some $x \in B^n(0, R)$, then f is differentiable at x, $\nabla f(x) = \nabla g(x)$ and*

$$f(y) = f(x)\langle \nabla f(x), y - x \rangle + O(|y - x|^2). \tag{29}$$

Proof. If $v \in \partial f(x)$, then clearly, $v \in \partial g(x)$ and hence $v = \nabla g(x)$. Therefore, the result follows from the estimate

$$f(x) + \langle \nabla g(x), y - x \rangle \leq f(y) \leq g(y) = f(x) + \langle \nabla g(x), y - x \rangle + O(|y - x|^2),$$

where in the last equality we used (28) and the fact that $g(x) = f(x)$. **QED**

The proof of Theorem 40 relies on covering the boundary of K with a finite number of graphs of $C^{1,1}$ functions. Thus the following lemma shows that each of these functions satisfies a sort of inner ball condition.

Lemma 45. *Let $f \in C^{1,1}(B^{n-1}(0,2N))$. Then there exists $R > 0$ such that for every $x \in B^{n-1}(0,N)$ there exists $\overline{B}^n(h(x),R) \subset \text{epi}(f)$ satisfying $\Gamma_f(x) \in \overline{B}^n(h(x),R)$.*

Proof. As $f \in C^{1,1}(B^{n-1}(0,2N))$, there exists $L > 0$ such that for all $x, y \in B^{n-1}(0,2N)$,

$$|\nabla f(x) - \nabla f(y)| \leq L|x-y|.$$

Applying Lemma 41 we then have,

$$f(x) \leq \frac{L}{2}|x-y|^2 + \nabla f(y) \cdot (x-y) + f(y). \tag{30}$$

Fix $y \in B^{n-1}(0,N)$ and let $g_y : \mathbb{R}^{n-1} \to \mathbb{R}$ be defined by,

$$g_y(x) = \frac{L}{2}|x-y|^2 + \nabla f(y) \cdot (x-y) + f(y).$$

Thus $g_y(y) = f(y)$ and by (30), for all $x \in B^{n-1}(0,N)$, $f(x) \leq g_y(x)$. As $y \in B^{n-1}(0,N)$ varies we see that g_y is a translated paraboloid and thus, by Lemma 43, there exists $R > 0$, depending only on L, such that for every $y \in B^{n-1}(0,N)$ there exists some $h(y) \in \text{epi}(g_y)$ with the property,

$$\Gamma_f(y) = \Gamma_{g_y}(y) \in \overline{B}(h(y),R) \subset \text{epi}(g_y). \tag{31}$$

Note $\text{epi}(f) \subset B^{n-1}(0,2N) \times (-\infty,\infty)$. As, for all $x \in B^{n-1}(0,2N)$, $f(x) \leq g_y(x)$, we have

$$\text{epi}(g_y) \cap (B^{n-1}(0,2N) \times (-\infty,\infty)) \subset \text{epi}(f).$$

If we restrict $2R < N$, then for all $y \in B^{n-1}(0,N)$, we can see that

$$\overline{B}(h(y),R) \subset B^{n-1}(0,2N) \times (-\infty,\infty). \tag{32}$$

Thus by (31) and (32), for all $y \in B^{n-1}(0,N)$,

$$\Gamma_f(y) \in \overline{B}(h(y),R) \subset \text{epi}(f)$$

as desired. **QED**

Proof of Theorem 40. Let K be of class $C^{1,1}$. As K is compact, by Lemma 45, ∂K can be covered by open sets $U_1, \ldots, U_N$ such that for each U_i, there exists $R_i > 0$, where closed balls of radius R_i cover $U_i \cap \partial K$ and are contained in K. Setting $R = \min\{R_1, \ldots, R_N\}$ we have for every $x \in \partial K$ there exists $h(x) \in K$ such that $x \in \overline{B}(h(x), R) \subset K$.

Finally consider the set

$$B_R := \bigcup \{\overline{B}(x, R) : y \in \overline{B}(x, R) \subset K \text{ and } y \in \partial K\}.$$

The above argument shows B_R is covered by closed balls of radius in R contained in K. For any $x \in K \setminus B_R$ we have that $\operatorname{dist}(x, \partial K) \geq 2R$. Hence $B(x, R) \subset K$. Therefore K is the union of balls of radius R.

To prove the converse, by translating and rotating, we can assume $0 \in K \subset \{x \in \mathbb{R}^n : x_n \geq 0\}$ and that there exists some open ball $B^{n-1}(0, N)$ and $M > 0$, such that by Proposition 31, $\partial K \cap (B^{n-1}(0, N) \times (-\infty, M])$ is the graph of $f : B^{n-1}(0, N) \to \mathbb{R}$, where f is convex. By the hypothesis, there exists a radius $R > 0$ such that for all $y \in B^{n-1}(0, N)$, there exists $h(y) \in K$ satisfying $\Gamma_f(y) \in \overline{B}(h(y), R) \subset K$. Let g_y be the function, such that the graph of g_y is the boundary of the bottom hemisphere of $\overline{B}(h(y), R)$. Then g_y and f are convex, $g_y \geq f$, as the ball is contained in K, $g_y(y) = f(y)$, and $g_y \in C^{1,1}$. Thus by Lemma 44, f is differentiable at y. As this is true for any $y \in B^{n-1}(0, N)$, we have that f is differentiable on $B^{n-1}(0, N)$. Moreover as f is convex and differentiable, then f is $C^1(B^{n-1}(0, N))$ showing that K is a hypersurface of class C^1.

To prove K is of class $C^{1,1}$ we need only show that its outer normal vector satisfies a Lipschitz property and the result will follow from Proposition 39. Given the uniform inner ball condition, for each $x \in \partial K$ there exists $h(x) \in K$ such that $x \in \overline{B}(h(x), R) \subset K$ and Proposition 35 implies that $h(x) = \pi_{K_R}(x)$. It is clear the outer unit normal $\nu(x)$ of ∂K will be the same as the outer unit normal for $\partial B(h(x), R)$ at $x \in \partial K$. Thus the outer unit normal vector at $x \in \partial K$ is given by, $\nu(x) = \frac{1}{R}(x - \pi_{K_R}(x))$. Applying Lemma 5 we have, $|\pi_{K_R}(x) - \pi_{K_R}(y)| \leq |x - y|$ so that for any $x, y \in \partial K$ we have,

$$|\nu(x) - \nu(y)| = \left| \frac{x - \pi_{K_R}(x)}{R} - \frac{y - \pi_{K_R}(y)}{R} \right| \leq \frac{|x - y|}{R} + \frac{|\pi_{K_R}(x) - \pi_{K_R}(y)|}{R} \leq \frac{2|x - y|}{R}$$

proving that K is of class $C^{1,1}(\mathbb{R}^n)$ as desired. **QED**

The above proof is geometrically motivated and follows from elementary arguments. Another proof of the sufficient condition of Theorem 40 can be found using the Implicit function theorem. In the following section, when extending the result on $C^{1,1}$ convex bodies to $C^{1,\alpha}$ convex bodies, the argument will be similar to the first proof of Theorem 40 as it is not clear how to apply the Implicit Function theorem to the $C^{1,\alpha}$ case. First let's state the version of the Implicit Function theorem we will need.

Theorem 46 (Implicit Function Theorem). *Let $F : \mathbb{R}^{n-1} \times \mathbb{R} \to \mathbb{R}$ be of class $C^{1,1}(\mathbb{R}^n)$. If $F(x_1, \ldots, x_n) = 0$ and $\frac{\partial F}{\partial x_n}(x_1, \ldots, x_n) \neq 0$, then there are a neighborhood $U \subset \mathbb{R}^{n-1}$, of $(x_1, \ldots, x_{n-1})$, a neighborhood $V \subset \mathbb{R}$, of x_n, and a $C^{1,1}$ function $f : U \to V$ such that,*

$$F(x_1, \ldots, x_{n-1}, f(x_1, \ldots, x_{n-1})) = 0 \quad \text{for all } (x_1, \ldots, x_{n-1}) \in U.$$

Moreover if $F(x) = t$ and $\nabla F(x) \neq 0$, then $F^{-1}(t)$ defines a surface in $\mathbb{R}^n$ that is locally the graph of a $C^{1,1}$ function.

For a proof of this version of the Implicit Function theorem, note that in the appendix of [24] there is an analogous version of the Inverse Function theorem for $C^{1,1}$ functions (see also [9, Theorem 2.1]). As the Inverse Function theorem implies the Implicit Function theorem, applying the result in [24] to the proof produces this $C^{1,1}$ version.

Definition 47. Let $K \subset \mathbb{R}^n$ be a convex body. Then $\delta_K : \mathbb{R}^n \to [0, \infty)$ is defined as

$$\delta_K(x) = \operatorname{dist}(x, K) = \inf\{|x - a| : a \in K\}.$$

Note that by the definition of the metric projection π_K, we have $\delta_K(x) = |x - \pi_K(x)|$.

The following is an easy consequence of the triangle inequality.

Proposition 48. *The function $\delta_K : \mathbb{R}^n \to [0, \infty)$ is 1-Lipschitz.*

In fact if K is a convex body we can apply the Lipschitz property of the metric projection π_K to show that the distance function squared is differentiable everywhere.

Theorem 49. *For a convex body $K \subset \mathbb{R}^n$, the function δ_K^2 is differentiable on $\mathbb{R}^n$ and*

$$\nabla \delta_K^2(x) = 2(x - \pi_K(x)).$$

Proof. For $x, y \in \mathbb{R}^n$, $\delta_K^2(x) \leq |x - \pi_K(y)|^2$ implies,

$$\delta_K^2(y) - \delta_K^2(x) \geq |y - \pi_K(y)|^2 - |x - \pi_K(y)|^2 = 2\langle y - x, x - \pi_K(y)\rangle + |y - x|^2$$

and by the 1-Lipshitz of π_K,

$$\begin{aligned}\langle y - x, x - \pi_K(y)\rangle &= \langle y - x, x - \pi_K(x)\rangle - \langle y - x, \pi_K(y) - \pi_K(x)\rangle \\ &\geq \langle y - x, x - \pi_K(x)\rangle - |y - x|^2.\end{aligned}$$

Thus,

$$\delta_K^2(y) - \delta_K^2(x) - 2\langle y - x, x - \pi_K(x)\rangle \geq -|y - x|^2.$$

Similarly,

$$\delta_K^2(y) - \delta_K^2(x) \leq |y - \pi_K(x)|^2 - |x - \pi_K(x)|^2 = |y - x|^2 + 2\langle y - x, x - \pi_K(x)\rangle$$

implying that

$$\delta_K^2(y) - \delta_K^2(x) - 2\langle y - x, x - \pi_K(x)\rangle \leq |y - x|^2.$$

Therefore,

$$\delta_K^2(y) - \delta_K^2(x) - 2\langle y - x, x - \pi_K(x)\rangle = o(|y - x|)$$

proving δ_K^2 is differentiable with $\nabla\delta_K^2(x) = 2(x - \pi_K(x))$. **QED**

Second Proof of Theorem 40. By Lemma 34, we may decrease $R > 0$ if necessary so that K_R is a convex body. By Theorem 49 the function $\delta_{K_R}^2(x) = \text{dist}(x, K_R)^2$ is differentiable and $\nabla\delta_{K_R}^2(x) = 2(x - \pi_{K_R}(x))$. Since the function π_{K_R} is Lipschitz by Lemma 5, we have that $\delta_{K_R}^2 \in C^{1,1}$ and for all $x \in \mathbb{R}^n \setminus K_R$, $\nabla\delta_{K_R}^2(x) \neq 0$. Therefore the set

$$\{x : \text{dist}(x, K_R) = R > 0\} = (\delta_{K_R}^2)^{-1}(R^2)$$

is locally the graph of a $C^{1,1}$ function by the Implicit function theorem. It remains to show that $\partial K = (\delta_{K_R}^2)^{-1}(R^2)$

Given the uniform inner ball condition, for each $x \in \partial K$ there exists $h(x) \in K$ such that $x \in \overline{B}(h(x), R) \subset K$. Moreover, $\text{dist}(h(x), \partial K) = R$ implies that $h(x) \in K_R$. As $\text{dist}(x, K_R) \geq R$ and $|x - h(x)| = R$ we have $\text{dist}(x, K_R) = R$. Hence $\partial K \subset \{y \in \mathbb{R}^n :$

$\text{dist}(y, K_R) = R\}$. Now if we let $x \in \{y \in \mathbb{R}^n : \text{dist}(y, K_R) = R\}$, then $\overline{B}(\pi_{K_R}(x), R) \subset K$, i.e. $x \in K$. If $x \in \text{int}\, K$, then extend the ray through $\pi_{K_R}(x)$ in the direction of $u_{K_R}(x)$, defined in (14), and denote its intersection with ∂K as y. By Proposition 35, we have $y \in \overline{B}(\pi_{K_R}(y), R)$. Obviously $\pi_{K_R}(y) = \pi_{K_R}(x)$ but

$$|y - \pi_{K_R}(y)| = |y - \pi_{K_R}(x)| > |x - \pi_{K_R}(x)| = R$$

a contradiction. Hence $x \in \partial K$. Therefore,

$$\partial K = \{x \in \mathbb{R}^n : \text{dist}(x, K_R) = R\} = (\delta^2_{K_R})^{-1}(R^2)$$

and K has boundary of class $C^{1,1}$. **QED**

3.3 Convex Bodies Satisfying the Approximate Inner Ball Condition

The goal of this section is to provide a generalization of Theorem 40 for convex bodies of class $C^{1,\alpha}$ by approximating the boundary of K with closed balls contained in K. As we characterized $C^{1,1}$ convex bodies with the uniform inner ball condition, we now establish a similar definition for an approximate inner ball condition.

Definition 50. We say a set K satisfies the (R, ε)-*approximate inner ball condition* if for each $x \in \partial K$ there exists $\overline{B}(h(x), R) \subset K$ such that $\text{dist}(x, \overline{B}(h(x), R)) \leq \varepsilon$.

Thus if K satisfies the (R, ε)-approximate inner ball condition, then we can approximate ∂K with closed balls of radius R at a distance of at most ε. It is important to note that in this definition R may depend upon ε, but is independent of the choice of $x \in \partial K$. Using the definition in (27), an equivalent definition of K satisfying the (R, ε)-approximate inner ball condition is that for all $x \in \partial K$,

$$\text{dist}(x, K(R)) \leq \varepsilon. \tag{33}$$

The main result of this chapter is:

Theorem 51. *Let $K \subset \mathbb{R}^n$ be a convex body. Then K is of class $C^{1,\alpha}$, for $\alpha \in (0,1]$, if and only if there exist $\varepsilon_0 > 0$ and $C > 0$ such that for each $0 < \varepsilon < \varepsilon_0$, K satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition.*

Remark 52. If we let $\alpha = 1$ in Theorem 51, then we have K satisfies the (C, ε)-approximate inner ball condition for all $\varepsilon < \varepsilon_0$, i.e. for every $x \in \partial K$ and for all $\varepsilon < \varepsilon_0$ there exists $\overline{B}(h_\varepsilon(x), C) \subset K$, such that $\operatorname{dist}(x, \overline{B}(h_\varepsilon(x), C)) \leq \varepsilon$. But it follows by compactness that there exists $h(x)$ such that $x \in \overline{B}(h(x), C) \subset K$, so that K is the union of closed balls of radius C and K satisfies the C-uniform inner ball condition. Therefore Theorem 51 is a generalization of Theorem 40.

Remark 53. It is important for the radius to be $O(\varepsilon^{\frac{1-\alpha}{1+\alpha}})$. If $\frac{1-\alpha}{1+\alpha} < 1$ the radius of the closed balls decrease at a rate slower than their distance from the boundary as $\varepsilon \to 0^+$. In fact if the radius of the closed balls was $O(\varepsilon)$, then it is possible for the boundary of K to not even be C^1. For example the cone $f(x) = |x|$ satisfies the $((\sqrt{2}-1)\varepsilon, \varepsilon)$-approximate inner ball condition.

To prove the forward direction of Theorem 51 we will need a lemma analogous to Lemma 45 for the $C^{1,1}$ case. In this case we need a definition for the (R, ε)-approximate inner ball condition for functions.

Definition 54. $f : U \to \mathbb{R}$, the $\operatorname{epi}(f)$ satisfies the (R, ε)-approximate inner ball condition on $V \subset U$, if for every $x \in V$ there exists $\overline{B}^{n+1}(h(x), R) \subset \operatorname{epi}(f)$ such that $\operatorname{dist}(\Gamma_f(x), \overline{B}^{n+1}(h(x), R)) \leq \varepsilon$.

The motivation for this definition follows from the desire to apply results of convex bodies to convex functions, but it is more common for functions to satisfy a local approximate inner ball condition.

Lemma 55. *Let $f \in C^{1,\alpha}(B^{n-1}(0, 2r))$. Then there exist constants $C, \varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,* epi$(f)$ *satisfies the (R, ε)-approximate inner ball condition on $B^{n-1}(0, r)$.*

Remark 56. Since $f \in C^{1,\alpha}(B^{n-1}(0, 2r))$, there exists $L > 0$ such that for all $x, y \in B^{n-1}(0, 2r)$, we have $|\nabla f(x) - \nabla f(y)| \leq L|x - y|^\alpha$. With this we can then find an explicit formula for the constant C; namely we will show $C = \frac{1+\alpha}{2} L^{\frac{-2}{1+\alpha}}$.

Proof. As $f \in C^{1,\alpha}(B^{n-1}(0,2r))$, let $L > 0$ be such that for all $x, y \in B^{n-1}(0,2r)$ we have

$$|\nabla f(x) - \nabla f(y)| \leq L|x-y|^\alpha.$$

By Lemma 41 we know that for all $x, y \in B^{n-1}(0,2r)$,

$$f(x) \leq \frac{L}{1+\alpha}|x-y|^{1+\alpha} + \nabla f(y) \cdot (x-y) + f(y). \tag{34}$$

Hence, for fixed $y \in B^{n-1}(0,2r)$ and any $\varepsilon > 0$, we can define two functions, $g_y : \mathbb{R}^n \to \mathbb{R}$ and $p_y^\varepsilon : \mathbb{R}^n \to \mathbb{R}$ by,

$$g_y(x) = \frac{L}{1+\alpha}|x-y|^{1+\alpha} + \nabla f(y) \cdot (x-y) + f(y)$$

$$p_y^\varepsilon(x) = \frac{L^{\frac{2}{1+\alpha}}}{(1+\alpha)\varepsilon^{\frac{1-\alpha}{1+\alpha}}}|x-y|^2 + \nabla f(y) \cdot (x-y) + f(y).$$

We note that $f(y) = g_y(y) = p_y^\varepsilon(y)$ and by (34) for all $x \in B^{n-1}(0,2r)$, we have $f(x) \leq g_y(x)$. We now claim that for all $x, y \in \mathbb{R}^{n-1}$ and for all $\varepsilon > 0$, we have $g_y(x) \leq p_y^\varepsilon(x) + \varepsilon$. To this end it suffices to show that

$$w(t) := \frac{L^{\frac{2}{1+\alpha}}}{(1+\alpha)\varepsilon^{\frac{1-\alpha}{1+\alpha}}}t^2 - \frac{L}{1+\alpha}t^{1+\alpha} + \varepsilon \geq 0 \quad \text{for all } t > 0$$

as $w(|x-y|) = \frac{L^{\frac{2}{1+\alpha}}}{(1+\alpha)\varepsilon^{\frac{1-\alpha}{1+\alpha}}}|x-y|^2 - \frac{L}{1+\alpha}|x-y|^{1+\alpha} + \varepsilon = p_y^\varepsilon(x) - g_y(x) + \varepsilon$. If $\alpha = 1$, then $w(t) = \varepsilon > 0$, so suppose $\alpha \in (0,1)$. Note as $t \to \infty$, $w(t) \to \infty$, thus by continuity, $w(t)$ will achieve its absolute minimum on $[0, \infty)$. For $t > 0$, we have,

$$w'(t) = \frac{2L^{\frac{2}{1+\alpha}}}{(1+\alpha)\varepsilon^{\frac{1-\alpha}{1+\alpha}}}t - Lt^\alpha$$

and solving $w'(t) = 0$, $t > 0$, we find the minimum will either be achieved at $t = 0$ or

$$t_0 = \left(\frac{\varepsilon}{L}\right)^{\frac{1}{1+\alpha}} \left(\frac{1+\alpha}{2}\right)^{\frac{1}{1-\alpha}}.$$

Obviously $w(0) = \varepsilon > 0$ and for $0 < \alpha < 1$ we have,

$$w(t_0) = \frac{\varepsilon(1+\alpha)^{\frac{2\alpha}{1-\alpha}}}{2^{\frac{1+\alpha}{1-\alpha}}} \left(\frac{1+\alpha}{2} - 1 + \frac{2^{\frac{1+\alpha}{1-\alpha}}}{(1+\alpha)^{\frac{2\alpha}{1-\alpha}}} \right).$$

Thus we must show, $\frac{1+\alpha}{2} - 1 + 2^{\frac{1+\alpha}{1-\alpha}}(1+\alpha)^{\frac{-2\alpha}{1-\alpha}} \geq 0$, which is equivalent to

$$2(1+\alpha)^{\frac{2\alpha}{1-\alpha}} \leq (1+\alpha)^{\frac{1+\alpha}{1-\alpha}} + 2^{\frac{2}{1-\alpha}}. \tag{35}$$

We then see that (35) follows from the fact that $0 < \alpha < 1$ implies,

$$2(1+\alpha)^{\frac{2\alpha}{1-\alpha}} < 2(2)^{\frac{2\alpha}{1-\alpha}} = 2^{\frac{1+\alpha}{1-\alpha}} \leq 1 + 2^{\frac{2}{1-\alpha}} < (1+\alpha)^{\frac{1+\alpha}{1-\alpha}} + 2^{\frac{2}{1-\alpha}}.$$

Thus we have proved that for all $x, y \in \mathbb{R}^{n-1}$, $g_y(x) \leq p_y^\varepsilon(x) + \varepsilon$ as desired.

We note that $p_y^\varepsilon + \varepsilon$ is a paraboloid and hence by Lemma 43, for every $y \in \mathbb{R}^{n-1}$, $p_y^\varepsilon + \varepsilon$ satisfies the R-uniform inner ball condition with $R = C\varepsilon^{\frac{1-\alpha}{1+\alpha}}$, where $C := \frac{(1+\alpha)}{2} L^{\frac{-2}{1+\alpha}}$. Thus for every $x, y \in \mathbb{R}^{n-1}$, there exists $h_y^\varepsilon(x) \in \operatorname{epi}(p_y^\varepsilon + \varepsilon)$ such that,

$$(x, p_y^\varepsilon(x) + \varepsilon) \in \overline{B}^n(h_y^\varepsilon(x), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset \operatorname{epi}(p_y^\varepsilon + \varepsilon). \tag{36}$$

For any $y \in B^{n-1}(0, 2r)$, as $f(y) = p_y^\varepsilon(y)$, we then have by (36),

$$\operatorname{dist}(\Gamma_f(y), \overline{B}^n(h^\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \leq \varepsilon, \tag{37}$$

where for simplicity the center of the ball denoted by $h_y^\varepsilon(y)$ is changed to $h^\varepsilon(y)$. It remains to show that there exists some $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ and for all $y \in B^{n-1}(0, r)$, we have $\overline{B}^n(h^\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset \operatorname{epi}(f)$. To accomplish this we fix $\varepsilon_0 > 0$ so that $\varepsilon_0 + 2C\varepsilon_0^{\frac{1-\alpha}{1+\alpha}} < r$. Let $0 < \varepsilon < \varepsilon_0$. Then, the fact that $2C\varepsilon^{\frac{1-\alpha}{1+\alpha}} + \varepsilon < r$, $y \in B^n(0, r)$, and (37) yield,

$$\overline{B}^n(h^\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset \overline{B}^n(\Gamma_f(y), r) \subset B^{n-1}(0, 2r) \times (-\infty, \infty).$$

As, $p_y^\varepsilon(x) + \varepsilon \geq f(x)$ for all $x \in B^{n-1}(0, 2r)$ we then have by (36)

$$\overline{B}^n(h^\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset \big(B(0, 2r) \times (-\infty, \infty)\big) \cap \operatorname{epi}(p_y^\varepsilon + \varepsilon) \subset \operatorname{epi}(f)$$

as desired. **QED**

Proof of Necessary Condition of Theorem 51. Let $K \subset \mathbb{R}^n$ be of class $C^{1,\alpha}$. By compactness and Proposition 31, we can cover ∂K by finitely many open sets $U_1, \ldots, U_N$ such that, for each $i = 1, \ldots, N$ we have $\partial K \cap U_i$ is the graph of $f_i \in C^{1,\alpha}$ and

$$|\nabla f_i(x) - \nabla f_i(y)| \leq L_i |x - y|^\alpha$$

for all $x, y \in \text{dom}(f_i)$. Let $L := \max\{L_1, \ldots, L_N\}$ and define $C_i := \frac{(1+\alpha)}{2} L_i^{\frac{-2}{1+\alpha}}$. Recall by Remark 56 that we can define $C = \frac{(1+\alpha)}{2} L^{\frac{-2}{1+\alpha}}$. We note that for all $i = 1, \ldots, N$, we have $C\varepsilon^{\frac{1-\alpha}{1+\alpha}} \leq C_i \varepsilon^{\frac{1-\alpha}{1+\alpha}}$. Thus for each $x \in \partial K$, there exists some $i \in \{1, \ldots, N\}$ such that $x \in \partial K \cap U_i$ and thus by Lemma 55, there exists $\varepsilon_0^i > 0$ such that for all $0 < \varepsilon < \varepsilon_0^i$ we have $\overline{B}(h_\varepsilon(x), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset \text{epi}(f_i)$ and

$$\text{dist}(x, \overline{B}(h_\varepsilon(x), C\varepsilon^{\frac{1-\alpha}{1+\alpha}})) \leq \varepsilon.$$

We can let $\varepsilon_0 = \min\{\varepsilon_0^1, \ldots, \varepsilon_0^N\}$, and by the compactness and convexity of K, decreasing ε_0 if necessary, we can guarantee that $\overline{B}(h_\varepsilon(x), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset K$ for all $\varepsilon < \varepsilon_0$. Therefore K satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition. **QED**

For the proof of the opposite direction of Theorem 51 the goal will be to use the approximate inner ball condition to first establish that at each point on the boundary of K, there is a unique supporting hyperplane. This ensures that K is of class C^1 by Proposition 33. Then applying geometric arguments we can choose closed balls in K that allow us to make estimates on the inner unit normal vectors of K, and helping us to show these inner unit normal vectors are α-Hölder continuous. Applying Proposition 39 will complete the result.

We first establish notation used throughout the proof of the sufficient condition of Theorem 51. Suppose that K satisfies the (R, ε)-approximate inner ball condition. For $x \in \partial K$, we define the collection of closed balls,

$$\mathfrak{B}^{(R,\varepsilon)}(x) := \{\overline{B}(y, R) \subset K : \text{dist}(x, \overline{B}(y, R)) \leq \varepsilon\}.$$

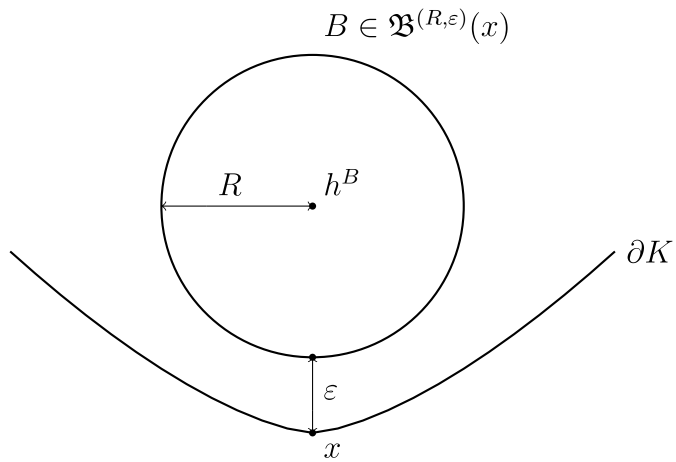
For each $B \in \mathfrak{B}^{(R,\varepsilon)}(x)$ we define h^B to be the center of B, so that $B := \overline{B}(h^B, R)$. Moreover, for each $x \in \partial K$ and each $B \in \mathfrak{B}^{(R,\varepsilon)}(x)$, the inner unit normal of B at $\pi_B(x) \in B$, is denoted ν^B and given by,

$$\nu^B = \frac{h^B - \pi_B(x)}{R}. \tag{38}$$

If K satisfies the uniform inner ball condition with closed balls of radius $R > 0$, then for each $x \in \partial K$ there is a unique closed ball of radius R such that $x \in \overline{B}(h(x), R) \subset K$. With the definition of the (R, ε)-approximate inner ball condition, for $x \in \partial K$, there may be an infinite number of closed balls satisfying $\overline{B}(y, R) \subset K$ and $\operatorname{dist}(x, B) \leq \varepsilon$. Moreover we know little about the location of these balls, so the challenge we must overcome in proving the opposite direction of Theorem 51 is to either ensure the results hold for any closed balls in the collection $\mathfrak{B}^{(R,\varepsilon)}(x)$ or show that we can always find closed balls in $\mathfrak{B}^{(R,\varepsilon)}(x)$ satisfying desired properties. Lemma 58 is an example of the former, while Lemma 60 is an example of the latter.

Remark 57. Under specific restrictions on R and ε, we may assume that if K satisfies the (R, ε)-approximate inner ball condition, then there exists $B \in \mathfrak{B}^{(R,\varepsilon)}(x)$, such that $\operatorname{dist}(x, B) = \varepsilon$. To see this, we restrict ε so that K_ε is a convex body, by Lemma 34, and restrict $R > 0$ so that there exists $B(y, R) \subset K_\varepsilon$. Then for any $x \in \partial K$, $\operatorname{dist}(x, B(y, R)) \geq \varepsilon$. Given that K satisfies the (R, ε)-approximate inner ball condition we also know for any $x \in \partial K$ there exists $\overline{B}(z, R) \in \mathfrak{B}^{(R,\varepsilon)}(x)$ such that $\operatorname{dist}(x, \overline{B}(z, R)) \leq \varepsilon$. Then we can find $\overline{B}(z', R) \subset \operatorname{co}(\overline{B}(z, R) \cup \overline{B}(y, R))$ such that $\operatorname{dist}(x, \overline{B}(z', R)) = \varepsilon$ and by convexity $\overline{B}(z', R) \subset K$, so that $\overline{B}(z', R) \in \mathfrak{B}^{(R,\varepsilon)}(x)$.

Figure 1: Illustration of Notation for Theorem 51



We first show that if K satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition for all $\varepsilon < \varepsilon_0$ then each point of the boundary has a unique supporting hyperplane. For the

$C^{1,1}$ case, as each point on the boundary is tangent to a closed ball, the uniqueness of the supporting hyperplane followed immediately by the convexity of K. For the $C^{1,\alpha}$ case, we show that the inner unit normal of any supporting hyperplane of $x \in \partial K$ must be sufficiently close to the inner unit normal ν^B for every $B \in \mathfrak{B}^{(C\varepsilon^{\frac{1-\alpha}{1+\alpha}},\varepsilon)}(x)$, with ε sufficiently small.

Lemma 58. *Let K satisfy the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}},\varepsilon)$-approximate inner ball condition for all $0 < \varepsilon < \varepsilon_0$. Let $H_b(\nu(x))$ be a supporting hyperplane of K at $x \in \partial K$ with inner unit normal vector $\nu(x)$ of K. Then for every $\delta > 0$ there exists $\eta(\delta) > 0$ such that for every $\varepsilon < \eta(\delta)$ and $B \in \mathfrak{B}^{(C\varepsilon^{\frac{1-\alpha}{1+\alpha}},\varepsilon)}(x)$, we have $|\nu(x) - \nu^B| \leq \delta$, where ν^B is defined in (38). Moreover $H_b(\nu(x))$ is the unique supporting hyperplane of K at x.*

Remark 59. If $\delta < \sqrt{2}$, we can find an explicit formula for η. Namely,

$$\eta(\delta) = \left(\frac{\delta^2}{2-\delta^2}\right)^{\frac{1+\alpha}{2\alpha}}$$

satisfies the claim.

Proof. Without loss of generality we can assume $C = 1$. Also we can assume $x = 0$, such that $H_0(e_n) := \{y \in \mathbb{R}^n : y_n = 0\}$ is a supporting hyperplane, and $K \subset \{y \in \mathbb{R}^n : y_n \geq 0\}$, as by a rotation and translation we get the same result for any $x \in \partial K$. Thus we may take the inner unit normal to be $\nu(0) = e_n$, where $e_n = (0, \dots, 0, 1)$. Then, for fixed $B \in \mathfrak{B}^{(\varepsilon^{\frac{1-\alpha}{1+\alpha}},\varepsilon)}(0)$, we have $\mathrm{dist}(x, B) = \beta \leq \varepsilon$. Thus we have $\pi_B(0) = \beta\nu^B$, as B is tangent to $\overline{B}(0,\beta)$, i.e. $\{\pi_B(0)\} = B \cap \overline{B}(0,\beta)$. Hence by (38),

$$\varepsilon^{\frac{1-\alpha}{1+\alpha}}\nu^B = h^B - \pi_B(0) = h^B - \beta\nu^B$$

and solving for h^B yields,

$$h^B = \beta\nu^B + \varepsilon^{\frac{1-\alpha}{1+\alpha}}\nu^B. \tag{39}$$

Given that $(h_1^B, \dots, h_{n-1}^B, h_n^B - \varepsilon^{\frac{1-\alpha}{1+\alpha}}) \in \partial B \subset K \subset \{x \in \mathbb{R}^n : x_n \geq 0\}$, we have

$$0 \leq h_n^B - \varepsilon^{\frac{1-\alpha}{1+\alpha}}. \tag{40}$$

Combining (39) and (40) gives us, $0 \leq \beta\nu^B + \varepsilon^{\frac{1-\alpha}{1+\alpha}}\nu_n^B - \varepsilon^{\frac{1-\alpha}{1+\alpha}}$ and as $\beta \leq \varepsilon$ this gives us,

$$0 \leq \beta\nu_n^B - (1-\nu_n^B)\varepsilon^{\frac{1-\alpha}{1+\alpha}} \leq \varepsilon\nu_n^B - (1-\nu_n^B)\varepsilon^{\frac{1-\alpha}{1+\alpha}}. \tag{41}$$

If $\nu_n^B = 1$, as $|\nu^B| = 1$, then $\nu^B = e_n$, and we are done. So consider $\nu_n^B < 1$. It is clear that $\pi_B(0) \in \{x \in B : h_n^B > x_n\}$ so that by (38), $\nu_n^B > 0$. Thus $0 < \nu_n^B < 1$ implies $0 < 1 - \nu_n^B < 1$ and solving (41) for ε yields,

$$\varepsilon \geq \left(\frac{(1 - \nu_n^B)}{\nu_n^B} \right)^{\frac{1+\alpha}{2\alpha}}. \tag{42}$$

Note that as, $\langle \vec{e_n}, \nu^B \rangle = \nu_n^B > 0$ we have,

$$|\vec{e_n} - \nu^B|^2 = \langle \vec{e_n}, \vec{e_n} \rangle - 2\langle \vec{e_n}, \nu^B \rangle + \langle \nu^B, \nu^B \rangle \leq |\vec{e_n}|^2 + |\nu^B|^2 = 2$$

and thus, $|e_n - \nu^B| \leq \sqrt{2}$. Therefore the result is trivial for $\delta \geq \sqrt{2}$. Let $0 < \delta < \sqrt{2}$ and fix $u \in \mathbb{R}^n$ such that $|e_n - u| = \delta$, with $|u| = 1$, $u_n \geq 0$. As $|e_n - u|^2 = \delta^2$ we have,

$$\delta^2 = u_1^2 + \cdots + u_{n-1}^2 + (1 - u_n)^2 \tag{43}$$

and as u is a unit vector, $1 - u_n^2 = u_1^2 + \cdots + u_{n-1}^2$. Combining this with (43) we can write,

$$\delta^2 = 1 - u_n^2 + (1 - u_n)^2 = 2(1 - u_n)$$

and solving this equation for u_n yields,

$$u_n = 1 - \frac{\delta^2}{2} = \frac{2 - \delta^2}{2}. \tag{44}$$

Using (44) we define η to be

$$\eta := \left(\frac{(1 - u_n)}{u_n} \right)^{\frac{1+\alpha}{2\alpha}} = \left(\frac{\delta^2}{2 - \delta^2} \right)^{\frac{1+\alpha}{2\alpha}}.$$

Thus for every $\varepsilon < \eta$ and $B \in \mathfrak{B}^\varepsilon(0)$, we must necessarily have $0 \leq u_n < \nu_n^B < 1$ as otherwise if $\nu_n^B \leq u_n$, then we have $1 - u_n \leq 1 - \nu_n^B$ and $\frac{1}{u_n} \leq \frac{1}{\nu_n^B}$ which implies,

$$\varepsilon < \eta = \left(\frac{(1 - u_n)}{u_n} \right)^{\frac{1+\alpha}{2\alpha}} \leq \left(\frac{(1 - \nu_n^B)}{\nu_n^B} \right)^{\frac{1+\alpha}{2\alpha}}$$

contradicting (42). As $|\nu^B| = |u| = 1$ and $0 \leq u_n < \nu_n^B < 1$,

$$|(\nu_1^B, \ldots, \nu_{n-1}^B)|^2 \leq |(u_1, \ldots, u_{n-1})|^2$$

and therefore,

$$|e_n - \nu^B|^2 = |(\nu_1^B, \dots, \nu_{n-1}^B)|^2 + (1-\nu_n^B)^2 \le |(u_1, \dots, u_{n-1})|^2 + (1-u_n)^2$$
$$= |e_n - u|^2 = \delta^2.$$

If there exists another inner normal vector of K at 0, $\nu'(0)$, such that $H_0(\nu'(0))$ is a supporting hyperplane at 0, and $\nu'(0) \neq \nu(0)$, then there exists $\delta_0 > 0$ such that $|\nu'(0) - \nu(0)| = \delta_0$. By the first part of the lemma there exists $\eta > 0$ such that for $0 < \varepsilon < \eta$ and $B \in \mathfrak{B}^{\varepsilon}(0)$ we have

$$|\nu(0) - \nu^B| \le \frac{\delta_0}{4}. \tag{45}$$

Likewise for $0 < \varepsilon < \eta$ and $B \in \mathfrak{B}^{\varepsilon}(0)$ we have

$$|\nu'(0) - \nu^B| \le \frac{\delta_0}{4}. \tag{46}$$

Hence for $0 < \varepsilon < \eta$ and $B \in \mathfrak{B}^{\varepsilon}(0)$ by (45) and (46)

$$\delta_0 = |\nu(0) - \nu'(0)| \le |\nu(0) - \nu^B| + |\nu'(0) - \nu^B| \le \frac{\delta_0}{2}$$

a contradiction, proving the uniqueness of $H_0(\nu(0))$. **QED**

Lemma 60. *Let K be a convex body such that K satisfies the (R,ε)-approximate inner ball condition. Then for every $x, y \in \partial K$ there exist $B_1 \in \mathfrak{B}^{(R,\varepsilon)}(x)$ and $B_2 \in \mathfrak{B}^{(R,\varepsilon)}(y)$ such that*

$$|\nu^{B_1} - \nu^{B_2}| \le \frac{2|x-y|}{R}$$

where ν^{B_1}, ν^{B_2} are defined in (38).

Proof. Fix $x, y \in \partial K$. Then we have, by the (R, ε)-approximate inner ball condition, that

$$|x - \pi_{K(R)}(x)| \leq \varepsilon \quad \text{and} \quad |y - \pi_{K(R)}(y)| \leq \varepsilon. \tag{47}$$

As $\pi_{K(R)}(x), \pi_{K(R)}(y) \in K(R)$ there exist closed balls $B_1, B_2 \subset K(R) \subset K$ such that, by (47),

$$\operatorname{dist}(x, B_1) \leq \varepsilon \quad \text{and} \quad \operatorname{dist}(y, B_2) \leq \varepsilon.$$

Therefore $B_1 \in \mathfrak{B}^{(R,\varepsilon)}(x)$ and $B_2 \in \mathfrak{B}^{(R,\varepsilon)}(y)$. Moreover the centers, h^{B_1} and h^{B_2}, of B_1, B_2 respectively, by Proposition 35, are in fact the metric projection of $\pi_{K(R)}(x)$ and $\pi_{K(R)}(y)$ onto the inner parallel body $(K(R))_R$, given by

$$(K(R))_R = \{z \in K(R) : \operatorname{dist}(z, \partial K(R)) \geq R\}.$$

Hence, by Lemma 5,

$$|h^{B_1} - h^{B_2}| \leq |x - y|. \tag{48}$$

Also note that the inner unit normals of B_1 and B_2, given in (38), can be written as,

$$\nu^{B_1} = \frac{h^{B_1} - \pi_{K(R)}(x)}{R} \quad \text{and} \quad \nu^{B_2} = \frac{h^{B_2} - \pi_{K(R)}(y)}{R}$$

so that again by Lemma 5 and (48), we have,

$$\begin{aligned} |\nu^{B_1} - \nu^{B_2}| = \left| \frac{h^{B_1} - \pi_{K(R)}(x)}{R} - \frac{h^{B_2} - \pi_{K(R)}(y)}{R} \right| &\leq \frac{|\pi_{K(R)}(x) - \pi_{K(R)}(y)|}{R} + \frac{|h^{B_1} - h^{B_2}|}{R} \\ &\leq \frac{2|x - y|}{R} \end{aligned}$$

completing the proof of the lemma. **QED**

We are now ready to complete the proof of Theorem 51.

Proof of Sufficient Condition of Theorem 51. Assume that there exists some $\varepsilon_0 > 0$ and without loss of generality let $C = 1$, such that for all $0 < \varepsilon < \varepsilon_0$, we have K satisfies the $(\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition. Lemma 58 and Proposition 33 show that K is of class C^1, as at each point on the boundary there is a unique supporting hyperplane. To prove K is of class $C^{1,\alpha}$ we will show the inner unit normal vectors are α-Hölder continuous and apply Proposition 39. We first consider the case $x, y \in \partial K$ such that $|x-y|^\alpha < \sqrt{2}$, with inner normal vectors $\nu(x), \nu(y)$ of K. Let $\delta = |x-y|^\alpha$ so that $\delta < \sqrt{2}$ and define

$$\eta := \min\left\{ \left(\frac{\delta^2}{2-\delta^2}\right)^{\frac{1+\alpha}{2\alpha}}, \varepsilon_0 \right\}.$$

Choose $M > 2$ such that $(\frac{2}{M})^{\frac{1+\alpha}{2\alpha}} < \varepsilon_0$. Thus if we fix $\varepsilon > 0$ to be,

$$\varepsilon := \frac{|x-y|^{1+\alpha}}{M^{\frac{1+\alpha}{2\alpha}}} \tag{49}$$

we can then see, as $|x-y|^{1+\alpha} < 2^{\frac{1+\alpha}{2\alpha}}$, that $\varepsilon \leq \varepsilon_0$ and

$$\varepsilon = \frac{|x-y|^{1+\alpha}}{M^{\frac{1+\alpha}{2\alpha}}} = \left(\frac{|x-y|^{2\alpha}}{M}\right)^{\frac{1+\alpha}{2\alpha}} = \left(\frac{\delta^2}{M}\right)^{\frac{1+\alpha}{2\alpha}} \leq \left(\frac{\delta^2}{M-\delta^2}\right)^{\frac{1+\alpha}{2\alpha}} \leq \left(\frac{\delta^2}{2-\delta^2}\right)^{\frac{1+\alpha}{2\alpha}}$$

so that $\varepsilon < \eta$. By Lemma 58 we know for every $\varepsilon < \eta$ and any $B_1 \in \mathfrak{B}^{(\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)}(x)$, $B_2 \in \mathfrak{B}^{(\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)}(y)$ we have

$$|\nu(x) - \nu^{B_1}| \leq \delta = |x-y|^\alpha \quad \text{and} \quad |\nu(y) - \nu^{B_2}| \leq \delta = |x-y|^\alpha. \tag{50}$$

Thus by Lemma 60 we can choose $B_1 \in \mathfrak{B}^{(\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)}(x)$ and $B_2 \in \mathfrak{B}^{(\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)}(y)$ such that

$$|\nu^{B_1} - \nu^{B_2}| \leq \frac{2|x-y|}{\varepsilon^{\frac{1-\alpha}{1+\alpha}}}.$$

Substituting (49) yields,

$$|\nu^{B_1} - \nu^{B_2}| \leq \frac{2|x-y|}{\left(\frac{|x-y|^{1+\alpha}}{M^{\frac{1+\alpha}{2\alpha}}}\right)^{\frac{1-\alpha}{1+\alpha}}} = 2M^{\frac{1-\alpha}{2\alpha}}|x-y|^\alpha \tag{51}$$

and combining (50) and (51) gives us,

$$\begin{aligned}|\nu(x)-\nu(y)| &\leq |\nu(x)-\nu^{B_1}|+|\nu^{B_1}-\nu^{B_2}|+|\nu^{B_2}-\nu(y)|\\ &\leq |x-y|^\alpha + 2M^{\frac{1-\alpha}{2\alpha}}|x-y|^\alpha + |x-y|^\alpha\\ &= \left(2+2M^{\frac{1-\alpha}{2\alpha}}\right)|x-y|^\alpha\end{aligned}$$

showing the inner unit normals are α-Hölder continuous for $|x-y|^\alpha < \sqrt{2}$. If $|x-y|^\alpha \geq \sqrt{2}$, then as $|\nu(x)| = |\nu(y)| = 1$ we have,

$$|\nu(x)-\nu(y)| \leq 2 = \sqrt{2}\sqrt{2} \leq \sqrt{2}|x-y|^\alpha.$$

Therefore $\nu(x)$ is α-Hölder continuous for all $x, y \in \partial K$ and hence K is of class $C^{1,\alpha}$. **QED**

In Section 3.2, for a general convex body K we used Theorem 40 to show that there existed a $C^{1,1}$ convex body $K(R)$ contained in K. Similarly, we can now use Theorem 51 to find a $C^{1,\alpha}$ convex body $K^\alpha(C,\varepsilon_0)$ contained in K. Recall by (27), that

$$K(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) = \bigcup\{\overline{B}^{n+1}(y, C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) : \overline{B}^{n+1}(y, C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset K\}. \tag{52}$$

We now further define,

$$K^\alpha(C,\varepsilon_0) := \{x \in K : \operatorname{dist}(x, K(C\varepsilon^{\frac{1-\alpha}{1+\alpha}})) \leq \varepsilon \text{ for all } \varepsilon \leq \varepsilon_0\}. \tag{53}$$

and if K contains a ball of radius $C\varepsilon_0^{\frac{1-\alpha}{1+\alpha}}$, then $K^\alpha(C,\varepsilon_0)$ is a $C^{1,\alpha}$ convex body by (33) and Theorem 51.

Given Theorem 51 we can now establish a more general version of Lemma 55, with the added assumption that f is convex, though convexity is only needed for the sufficient condition. The following corollary is a version of Theorem 51 for functions.

Corollary 61. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then $f \in C^{1,\alpha}_{\text{loc}}(\mathbb{R}^n)$ if and only if for every $x \in \mathbb{R}^n$ and $\delta > 0$ there exist constants $\varepsilon_0, C > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,* epi$(f)$ *satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition on $B^n(x,\delta)$.*

The idea of the proof of Corollary 61 is to first construct a convex body using the epigraph of f and then we can find a $C^{1,\alpha}$ convex body, contained in the epigraph of f, that intersects the graph of f in a neighborhood.

Proof of Corollary 61. The forward direction follows from Lemma 55, where convexity is not needed. For the opposite direction, fix $x \in \mathbb{R}^n$ and $\delta > 0$ and define

$$K = \Big(B^n(x, 2R) \times (-\infty, M]\Big) \cap \operatorname{epi}(f)$$

where $R, M \in \mathbb{R}$ are chosen to be,

$$R = \max\{\delta\,, 2C\varepsilon_0^{\frac{1-\alpha}{1+\alpha}} + \varepsilon_0\} \quad \text{and} \quad M = \sup_{y \in B(x,2R)} f(y) + R.$$

Note K is obviously a convex body. We then consider the $C^{1,\alpha}$ convex body given in (53) $K^\alpha(C, \varepsilon_0)$ and it remains to show that for all $y \in B^n(x, \delta)$, $\Gamma_f(y) \in \partial K^\alpha(C, \varepsilon_0)$.

By the assumption, for each $y \in B^n(x, \delta)$, there exists $\overline{B}^{n+1}(h_\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset \operatorname{epi}(f)$ such that

$$\operatorname{dist}(\Gamma_f(y), \overline{B}^{n+1}(h_\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}})) \leq \varepsilon. \tag{54}$$

Then by our choice of R and M, for all $y \in B^n(x, \delta)$,

$$\overline{B}^{n+1}(h_\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset B^{n+1}(\Gamma_f(y), R) \subset B^n(y, R) \times (-\infty, f(y) + R] \subset B^n(x, 2R) \times (-\infty, M]$$

and hence $\overline{B}^{n+1}(h_\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset K$. By definition of $K(C\varepsilon^{\frac{1-\alpha}{1+\alpha}})$ this shows that

$$\overline{B}^{n+1}(h_\varepsilon(y), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset K(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}).$$

Hence, by (54), for all $y \in B^n(x, \delta)$,

$$\operatorname{dist}(\Gamma_f(y), K(C\varepsilon^{\frac{1-\alpha}{1+\alpha}})) \leq \varepsilon$$

showing that, by (53), $\Gamma_f(y) \in \partial K^\alpha(C, \varepsilon_0)$. Therefore the graph of f restricted to $B^n(x, \delta)$ coincides with $K^\alpha(C, \varepsilon_0)$ showing $f \in C^{1,\alpha}(B^n(x, \delta)$. As this is true for any $x \in \mathbb{R}^n$ and $\delta > 0$, this implies $f \in C^{1,\alpha}_{\text{loc}}(\mathbb{R}^n)$. **QED**

3.4 Regularity of the Convex Envelope

Let $f : \mathbb{R}^n \to \mathbb{R}$ be bounded below by an affine function. We define the convex envelope of f to be

$$\text{conv}(f)(x) := \sup\{g(x) : g \leq f,\ g \text{ is convex}\}.$$

This is well defined and finite as f is bounded below by an affine function. Another way of defining the convex envelope of f is by making use of the convex hull of epi(f). Recall co(epi(f)) is the intersection of all convex sets containing epi(f). Thus we have the following equivalence of definitions for the convex envelope of f:

Lemma 62. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be bounded below by an affine function. Then,*

$$\text{conv}(f)(x) = \inf\{y \in \mathbb{R} : (x, y) \in \text{co}(\text{epi}(f))\}$$

Proof. Let $g : \mathbb{R}^n \to \mathbb{R}$ be defined by,

$$g(x) = \inf\{y \in \mathbb{R} : (x, y) \in \text{co}(\text{epi}(f))\},$$

and let $\ell(x)$ be an affine function satisfying $f(x) \geq \ell(x)$ for all $x \in \mathbb{R}^n$. Then co(epi(f)) $\subset$ epi(ℓ) implies that $g(x)$ is well defined and finite for each $x \in \mathbb{R}^n$. Moreover, by definition of g, $\Gamma_g(x) \in \partial\,\text{co}(\text{epi}(f))$. Given that co(epi$(f)$) is a convex set of dimension $n+1$ it is clear that $g(x)$ defines a convex function (see 31) and $g(x) \leq f(x)$ for all $x \in \mathbb{R}^n$. Thus we have by definition, $\text{conv}(f)(x) \geq g(x)$. Also, by definition, epi$(f) \subset$ epi(conv(f)), and as epi(conv(f)) is a closed convex set, co(epi(f)) $\subset$ epi(conv(f)), so that $g(x) \geq \text{conv}(f)(x)$ showing that $g(x) = \text{conv}(f)(x)$. **QED**

A useful characterization of the convex envelope follows from the Carathéodory theorem (Theorem 4). The proof given here is a more detailed version of the one presented in [27, Corollary 17.1.5].

Theorem 63. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\ell(x)$ an affine function satisfying $f(x) \geq \ell(x)$. Then,*

$$\text{conv}(f)(x) = \inf\left\{\sum_{i=1}^{n+1} \lambda_i f(x_i) : \lambda_i \geq 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1, \quad \sum_{i=1}^{n+1} \lambda_i x_i = x\right\} \tag{55}$$

Remark 64. We note that co(epi(f)) is an $n+1$ dimensional convex set so the Carathéodory theorem and Lemma 62 easily imply that conv(f)(x) can be written as the convex combination of $n+2$ elements of epi(f), yet Theorem 63 states we can write conv(f)(x) as the convex combination of $n+1$ elements of epi(f).

Proof. Fix $x \in \mathbb{R}^n$ and let $\lambda_1, \ldots, \lambda_{n+1} \geq 0$ and $x_1, \ldots, x_{n+1} \in \mathbb{R}^n$ be such that

$$x = \sum_{i=1}^{n+1} \lambda_i x_i, \quad \text{and} \quad \sum_{i=1}^{n+1} \lambda_i = 1.$$

Then it is clear that,

$$(x, y) := \sum_{i=1}^{n+1} \lambda_i (x_i, f(x_i)) \in \mathrm{co}(\mathrm{epi}(f))$$

and thus by Lemma 62, conv(f)(x) $\leq y$. As this is true for any y of this form, we have

$$\mathrm{conv}(f)(x) \leq \inf \left\{ \sum_{i=1}^{n+1} \lambda_i f(x_i) : \lambda_i \geq 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1, \quad \sum_{i=1}^{n+1} \lambda_i x_i = x \right\}$$

Thus, we need only show that for every $\varepsilon > 0$, there exists some $z \in \mathbb{R}$ satisfying,

$$z = \sum_{i=1}^{n+1} \lambda_i f(x_i), \quad \lambda_i \geq 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1, \quad \sum_{i=1}^{n+1} \lambda_i x_i = x \tag{56}$$

such that $z < \mathrm{conv}(f)(x) + \varepsilon$. By Lemma 62, let $(x, y) \in \mathrm{co}(\mathrm{epi}(f))$ be such that $y < \mathrm{conv}(f)(x) + \varepsilon$. By applying the Carathéodory theorem to epi(f) there exists $\{(x_i, y_i)\}_{i=1}^{n+2} \in$ epi(f) such that

$$(x, y) \in \Lambda := \mathrm{co}\left(\{(x_i, y_i)\}_{i=1}^{n+2}\right).$$

Then we choose $z^* := \inf\{t \in \mathbb{R} : (x, t) \in \Lambda\}$ which shows that z^* satisfies $z^* \leq y < \mathrm{conv}(f)(x) + \varepsilon$. Note, Λ is a closed convex polyhedra whose vertices are contained in the set $\{(x_i, y_i)\}_{i=1}^{n+2}\}$. Thus the point (x, z^*) is contained on the boundary of Λ, and in particular $(x, z^*) \in \Lambda'$ where Λ' is a face of Λ. Moreover Λ' is a closed convex polyhedra of dimension less than or equal to n, with at most $n+1$ vertices contained in $\{(x_i, y_i)\}_{i=1}^{n+2}$. By relabeling, we may assume that $\Lambda' \subset \mathrm{co}(\{(x_i, y_i)\}_{i=1}^{n+1}\}$ and thus,

$$(x, z^*) = \sum_{i=1}^{n+1} \lambda_i (x_i, y_i), \quad \sum_{i=1}^{n+1} \lambda_i = 1, \quad \lambda_i \geq 0$$

which shows that z^* is of the form,

$$z^* = \sum_{i=1}^{n+1} \lambda_i y_i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1, \quad \sum_{i=1}^{n+1} \lambda_i x_i = x.$$

As $\{(x_i, y_i)\}_{i=1}^{n+1} \in \text{epi}(f)$, this implies $f(x_i) \leq y_i$ so that we can choose $z := \sum_{i=1}^{n+1} \lambda_i f(x_i)$, and thus

$$z \leq z^* < \text{conv}(f)(x) + \varepsilon$$

as desired. **QED**

It is convenient to place certain restrictions on the coefficients used when writing x as the convex combination of points in $\mathbb{R}^n$. We define the set of convex combinations of $x \in \mathbb{R}^n$, with decreasing coefficients, as

$$C(x) = \left\{ \{(\lambda_i, x_i)\}_{i=1}^{n+1} : 1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n+1} \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1, \text{ and } \sum_{i=1}^{n+1} \lambda_i x_i = x \right\}$$

so that we can rewrite Theorem 63 as,

$$\text{conv}(f)(x) = \inf \left\{ \sum_{i=1}^{n+1} \lambda_i f(x_i) : \{(\lambda_i, x_i)\}_{i=1}^{n+1} \in C(x) \right\}.$$

By properties of the infimum, for all $x \in \mathbb{R}^n$, there is a sequence $\{(\lambda_i^{(k)}, x_i^{(k)})\}_{i=1}^{n+1} \in C(x)$ such that,

$$\sum_{i=1}^{n+1} \lambda_i^{(k)} f(x_i^{(k)}) \to \text{conv}(f)(x) \quad \text{as } k \to \infty. \tag{57}$$

In 2001 Kirchheim and Kristensen showed in [18] that a $C^{1,\alpha}_{\text{loc}}(\mathbb{R}^n)$ function satisfying $f \to \infty$ as $|x| \to \infty$ has convex envelope of class $C^{1,\alpha}_{\text{loc}}(\mathbb{R}^n)$. Furthermore in [19] a weaker version of this result is proved using the infimal convolution, assuming that f grows faster than any linear function. Here we provide a novel proof for this theorem applying the geometric characterization of $C^{1,\alpha}$ convex bodies given in Theorem 51.

We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is *coercive* if $f(x) \to \infty$ as $|x| \to \infty$. Namely, continuous coercive functions achieve a minimum in $\mathbb{R}^n$. The proof of Theorem 70 uses the sequences as in (57). The following lemma tells us that for $y \in B(x, \delta)$ we can select such a sequence in $C(y)$ so that both $\mu_1^{(k)}$ and $y_1^{(k)}$ converge. The importance of this lemma is that $\mu_1^{(k)}$ converges to something positive and that the $y_1^{(k)}$ remains bounded.

Lemma 65. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous and coercive, $x \in \mathbb{R}^n$, and $\delta > 0$. Then for every $y \in B(x,\delta)$ there exists a sequence $\{(\mu_i^{(k)}, y_i^{(k)})\}_{i=1}^{n+1} \in C(y)$ for all $k \in \mathbb{N}$ such that $\mu_1^{(k)} \to \mu_1 \in \left[\frac{1}{n+1}, 1\right]$, $y_1^{(k)} \to y_1$ and $\{(\mu_i^{(k)}, y_i^{(k)})\}_{i=1}^{n+1}$ satisfies (57). Moreover there exists $M > 0$ and $N > 0$ such that,*

$$B(x,\delta) \subset B(0,M) \quad \text{and} \quad y_1^{(k)} \subset B(0,M) \quad \text{for all } k \geq N.$$

Remark 66. In Lemma 65, M depends only on the choice of $x \in \mathbb{R}^n$ and $\delta > 0$.

Proof. As f is continuous and coercive, we can assume $f \geq 0$ and we fix $y \in B(x,\delta)$. Applying Theorem 63 there exists $\{(\mu_i^{(k)}, y_i^{(k)})\}_{i=1}^{n+1} \in C(y)$ satisfying (57). If $\mu_1^{(k)} < \frac{1}{n+1}$, we would have $\mu_i^{(k)} < \frac{1}{n+1}$ for all $i = 1, \ldots, n+1$, which contradicts $\sum_{i=1}^{n+1} \mu_i^{(k)} = 1$. Therefore for all $k \in \mathbb{N}$, we have $\mu_1^{(k)} \in \left[\frac{1}{n+1}, 1\right]$. Moreover by the coercivity of f, we know $y_1^{(k)}$ is bounded. Indeed, if not, then there exists a subsequence $|y_1^{(k_j)}| \to \infty$ and by coercivity, $f(y_1^{(k_j)}) \to \infty$. As $\mu_1^{(k)} \geq \frac{1}{n+1}$, $\mu_i^{(k)} \geq 0$ for all $i = 2, \ldots, n+1$, and $f \geq 0$, we have for each $j \in \mathbb{N}$,

$$\sum_{i=1}^{n+1} \mu_i^{(k_j)} f_i^{(k_j)} \geq \mu_1^{(k_j)} f(y_1^{(k_j)}) \geq \frac{f(y_1^{(k_j)})}{n+1}.$$

Letting $j \to \infty$ the left side approaches $\operatorname{conv}(f)(y)$ and the right side approaches ∞, a contradiction as f is finite everywhere. Therefore, up to subsequences, we may assume that $\mu_1^{(k)} \to \mu_1 \in \left[\frac{1}{n+1}, 1\right]$ and $y_1^{(k)} \to y_1$.

Also by coercivity, there exists $M > 0$ such that $B(x,\delta) \subset B(0,M)$ and,

$$f(z) \geq (n+1)\left(\sup_{y \in B(x,\delta)} f(y) + 1\right) \quad \text{for all } |z| > M. \tag{58}$$

We must have for k large enough, $y_1^{(k)} \in B(0,M)$ as if not there exists a subseqeunce such that, $|y_1^{(k_j)}| > M$ for all $j \in \mathbb{N}$. Then (58) implies,

$$\sum_{i=1}^{n+1} \mu_i^{(k_j)} f(y_i^{(k_j)}) \geq \mu_1^{(k_j)} f(y_1^{(k_j)}) \geq \mu_1^{(k_j)}(n+1)(f(y)+1) \geq f(y)+1$$

and letting $j \to \infty$ we have $\operatorname{conv}(f)(y) \geq f(y) + 1$, a contradiction. Moreover this implies $y_1 \in \overline{B}(0,M)$. Thus for k large enough, we have $y_1^{(k)} \subset B(0,M)$. **QED**

Remark 67. In the proof of the following lemma, we will need to consider the translations and dilations of balls. Let $a, r > 0$ and $x, h \in \mathbb{R}^n$. Then,

$$aB(h,r) + \{x\} = \{ay + x : y \in B(h,r)\} = B(ah + x, ar).$$

Lemma 68. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be coercive and continuous. If for every $x \in \mathbb{R}^n$ and $\delta > 0$ there exists $R, \varepsilon > 0$ such that* $\operatorname{epi}(f)$ *satisfies the (R, ε)-approximate inner ball condition on $B^n(x, \delta)$, then for any $x \in \mathbb{R}^n$ and $\delta > 0$ there exists $R_0, \delta_0 > 0$ such that* $\operatorname{epi}(\operatorname{conv}(f))$ *satisfies the (R_0, ε_0)-approximate inner ball condition on $B^n(x, \delta)$.*

Remark 69. In the proof we specifically show that for fixed $x \in \mathbb{R}^n$ and $\delta > 0$, if f satisfies the (R, ε)-approximate inner ball condition on $B^n(0, M)$ for M large enough, then $\operatorname{conv}(f)$ satisfies the $(\frac{R}{n+1}, \varepsilon)$-approximate inner ball condition on $B^n(x, \delta)$.

Proof. Fix $x \in \mathbb{R}^n$ and $\delta > 0$. By Lemma 65, for any $y \in B^n(x, \delta)$ we can find a sequence $\{\mu_i^{(k)}, y_i^{(k)}\}_{i=1}^{n+1} \in C(y)$ for all $k \in \mathbb{N}$, such that $\mu_1^{(k)} \to \mu_1 \geq \frac{1}{n+1}$, $y_1^{(k)} \to y_1$ and

$$\sum_{i=1}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)}) \to \Gamma_{\operatorname{conv}(f)}(y) \quad \text{as } k \to \infty. \tag{59}$$

Lemma 65 also implies there exists $M > 0$, depending only on the choice of x and δ, such that $B^n(x, \delta) \cup \{y_1^{(k)}\}_{k=1}^{\infty} \subset B^n(0, M)$. By assumption, there exist $R, \varepsilon > 0$ such that $\operatorname{epi}(f)$ satisfies the (R, ε)-approximate inner ball condition on $B^n(0, M)$. Thus for each $y_1^{(k)}$, there exists $h^{(k)}$ such that

$$\overline{B}^{n+1}(h^{(k)}, R) \subset \operatorname{epi}(f) \quad \text{and } \operatorname{dist}(\Gamma_f(y_1^{(k)}), \overline{B}(h^{(k)}, R)) \leq \varepsilon. \tag{60}$$

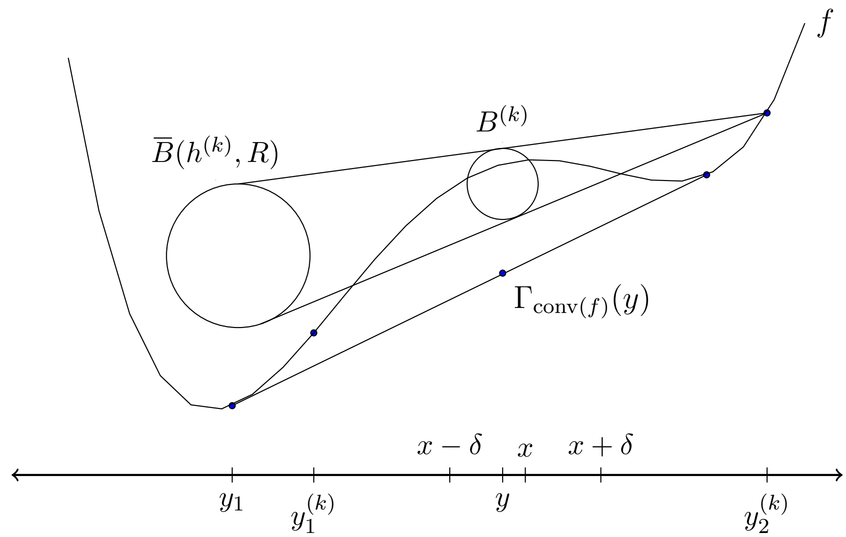
We now consider the closed ball $B^{(k)}$ given by,

$$B^{(k)} := \mu_1^{(k)} \overline{B}^{n+1}(h^{(k)}, R) + \left\{ \sum_{i=2}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)}) \right\}.$$

As $\overline{B}^{n+1}(h^{(k)}, R) \subset \operatorname{epi}(f)$ and $\Gamma_f(y_i^{(k)}) \in \operatorname{epi}(f)$ for all $k \in \mathbb{N}$, we have $B^{(k)} \subset \operatorname{epi}(\operatorname{conv}(f))$ for all $k \in \mathbb{N}$, being the convex combination of elements in $\operatorname{epi}(f)$. By (59), as $\sum_{i=1}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)})$ converges, there exists $L > 0$ such that $|\sum_{i=1}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)})| \leq L$ for all $k \in \mathbb{N}$ and (60) implies

$$|\Gamma_f(y_1^{(k)}) - h^{(k)}| \leq R + \varepsilon.$$

Figure 2: Proof of Lemma 68



Thus the centers of $B^{(k)}$, given by $\mu_1^{(k)} h^{(k)} + \sum_{i=2}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)})$, are bounded. Indeed,

$$\begin{aligned} \left| \mu_1^{(k)} h^{(k)} + \sum_{i=2}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)}) \right| &\leq \left| \mu_1^{(k)} h^{(k)} - \mu_1^{(k)} \Gamma_f(y_1^{(k)}) \right| + \left| \sum_{i=1}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)}) \right| \\ &\leq \mu_1^{(k)}(R+\varepsilon) + L \leq R + \varepsilon + L. \end{aligned}$$

Therefore up to a subsequence, we may assume that $\mu_1^{(k)} h^{(k)} + \sum_{i=2}^{n+1} \mu_i^{(k)} \Gamma_f(y_i^{(k)}) \to h$.

We now claim that the closed ball $\overline{B}^{n+1}(h, \mu_1 R)$ satisfies,

$$\overline{B}^{n+1}(h, \mu_1 R) \subset \mathrm{epi}(\mathrm{conv}(f)) \quad \text{and } \mathrm{dist}(\Gamma_{\mathrm{conv}(f)}(y), \overline{B}^{n+1}(h, \mu_1 R)) \leq \mu_1 \varepsilon. \tag{61}$$

As the centers of $B^{(k)}$ converge to h and clearly the radius of $B^{(k)}$, given by $\mu_1^{(k)} R$, converges to $\mu_1 R$, we can see that $B^{(k)} \to \overline{B}^{n+1}(h, \mu_1 R)$ as $k \to \infty$, in the sense that for every $x \in \overline{B}^{n+1}(h, \mu_1 R)$ there exists $x^{(k)} \in B^{(k)}$ such that $x^{(k)} \to x$. Thus, as epi(conv(f)) is closed and $B^{(k)} \subset \mathrm{epi}(\mathrm{conv}(f))$ we have $\overline{B}^{n+1}(h, \mu_1 R) \subset \mathrm{epi}(\mathrm{conv}(f))$. Also by (60), there exists some $a^{(k)} \in \overline{B}^{n+1}(h^{(k)}, R)$ such that

$$|\Gamma_f(y_1^{(k)}) - a^{(k)}| \leq \varepsilon. \tag{62}$$

By the same argument used to show the centers of $B^{(k)}$ are bounded, we see that $\mu_1^{(k)}a^{(k)} + \sum_{i=2}^{n+1}\mu_i^{(k)}\Gamma_f(y_i^{(k)})$ are bounded and up to a subsequence we have,

$$B^{(k)} \ni \mu_1^{(k)}a^{(k)} + \sum_{i=2}^{n+1}\mu_i^{(k)}\Gamma_f(y_i^{(k)}) \to a \in \overline{B}(h, \mu_1 R).$$

Therefore (62) shows,

$$\left|\sum_{i=1}^{n+1}\mu_i^{(k)}\Gamma_f(y_i^{(k)}) - \left(\mu_1^{(k)}a^{(k)} + \sum_{i=2}^{n+1}\mu_i^{(k)}\Gamma_f(y_i^{(k)})\right)\right| \le \mu_1^{(k)}\varepsilon$$

and letting $k \to \infty$ yields,

$$|\Gamma_{\mathrm{conv}(f)}(y) - a| \le \mu_1\varepsilon.$$

As $a \in \overline{B}^{n+1}(h, \mu_1 R)$ this proves (61). Given that $\mu_1 \ge \frac{1}{n+1}$ we can find a smaller ball and $h' \in \mathrm{epi}(\mathrm{conv}(f))$ such that

$$\overline{B}^{n+1}\left(h', \frac{R}{n+1}\right) \subset \mathrm{epi}(\mathrm{conv}(f)) \quad \text{and} \quad \mathrm{dist}\left(\Gamma_{\mathrm{conv}(f)}(y), \overline{B}^{n+1}\left(h', \frac{R}{n+1}\right)\right) \le \varepsilon$$

showing that epi(conv(f)) satisfies the $(\frac{R}{n+1}, \varepsilon)$-approximate inner ball condition on $B^n(x, \delta)$.

QED

Theorem 70. *Let $f : \mathbb{R}^n \to \mathbb{R}$ satisfy $f \in C^{1,\alpha}_{\mathrm{loc}}(\mathbb{R}^n)$ and be coercive. Then* $\mathrm{conv}(f) \in C^{1,\alpha}_{\mathrm{loc}}(\mathbb{R}^n)$.

Proof. By Corollary 61, for any $x \in \mathbb{R}^n$ and $\delta > 0$ there exists $\varepsilon_0, C > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, epi$(f)$ satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition on $B(x, \delta)$. Specifically, choosing $M > 0$ as in the proof of Lemma 68, we have that there exists $\varepsilon_0, C > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, epi$(f)$ satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition on $B(0, M)$. As f is coercive, the proof of Lemma 68 implies that for all $0 < \varepsilon < \varepsilon_0$, epi(conv$(f)$) satisfies the $\left(\frac{C}{n+1}\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon\right)$-approximate inner ball condition on $B(x, \delta)$. The result follows by again applying Corollary 61. **QED**

Remark 71. In [7, Example 4.1] is given a counterexample to Theorem 70 if we remove the coercivity of f. The function $f : \mathbb{R}^2 \to \mathbb{R}$, defined as,

$$f(x, y) = \sqrt{x^2 + e^{-y^2}}$$

is clearly smooth, and it can be seen that $\text{conv}(f)(x, y) = |x|$. The value $\text{conv}(f)(0, y)$, for any $y \in \mathbb{R}$, can be approximated by a sequence of convex combinations of points of the form $f(0, y_i^{(k)})$, such that $y_i^{(k)} \to \infty$ as $k \to \infty$. Moreover, as $\text{epi}(f)$ satisfies the (R, ε)-approximate inner ball condition, the graph of f can be approximated by balls of radius R. But as $k \to \infty$ we have $R \to 0$.

3.5 Sum of Convex Bodies

We define the *Minkowski sum* of two sets as,

$$A + B = \{x + y : x \in A \text{ and } y \in B\}.$$

We also define a convex domain as a convex set with nonempty interior. It is easy to verify that the sum of two convex domain is itself a convex domain. To apply the results from the previous sections to the sum of two sets we will need to work with the boundary of $A+B$ and see how this connects to the boundaries of A and B individually. While $\partial(A+B) \neq \partial A+\partial B$, we do have the following inclusion,

Lemma 72. *For convex body $A \subset \mathbb{R}^n$ and convex domain $B \subset \mathbb{R}^n$, $\partial(A + B) \subset \partial A + \partial B$.*

We first prove this simple lemma,

Lemma 73. *For $A \subset \mathbb{R}^n$ compact and $B \subset \mathbb{R}^n$ we have $\overline{A + B} \subset \overline{A} + \overline{B}$.*

Proof. Let $z \in \overline{A + B}$. This implies that there exists a sequence $z_n \in A + B$ such that $z_n \to z$. As $z_n \in A + B$, there exist $a_n \in A$ and $b_n \in B$ such that $z_n = a_n + b_n$. By compactness of A, there exists $a_{n_k} \to a \in A$. Thus, if we define $b := z - a$, we have,

$$b_{n_k} = z_{n_k} - a_{n_k} \to z - a = b$$

so that $z = a + b$ where $a \in \overline{A}$ and $b \in \overline{B}$. **QED**

Proof of Lemma 72. Let $z \in \partial(A+B)$ and by Lemma 73 we have $\partial(A+B) \subset \overline{A+B} \subset \overline{A} + \overline{B}$ so that there exists $a \in \overline{A}$ and $b \in \overline{B}$ such that $z = a + b$. The claim is that $a \in \partial A$ and $b \in \partial B$. If $a \in \operatorname{int} A$, there exists $\delta > 0$ such that $B(a,\delta) \subset A$. Thus $B(a,\delta) + \{b\} \subset B(a,\delta) + B \subset A + B$. This implies that $B(a+b,\delta) = B(z,\delta) \subset A + B$, and thus $z \in \operatorname{int}(A+B)$, a contradiction, proving $a \in \partial A$. A similar argument shows that $b \in \partial B$ **QED**

Remark 74. In general $\partial(A+B) \neq \partial A + \partial B$. Consider the sets $A = [0,2] \times [0,2]$ and $B = \overline{B}^2((0,0),1)$. Then $(0,0) \in \partial A$ and $(1,0) \in \partial B$, but

$$(1,0) + (0,0) = (1,0) \in B^2((1,0),1/2) \subset (1,0) + B^2((0,0),1/2) \subset A + B$$

so that $(1,0) + (0,0) \notin \partial(A+B)$.

Theorem 75. *If $A \subset \mathbb{R}^n$ is a convex body of class C^1 and B is any convex domain, then $A + B$ is of class C^1.*

Proof. We will be making use of the characterization of C^1 convex domains given in Proposition 33, that is, the goal is to show at each $z \in \partial(A+B)$ there exists a unique supporting hyperplane. We first note that for each $b \in B$, the set $A + \{b\}$ is a C^1 convex body. Thus for each $a \in \partial A$, there exists a unique supporting hyperplane of $A + \{b\}$ at $a + b$. Fix $z_0 \in \partial(A+B)$. By Lemma 72, there exists $a_0 \in \partial A$ and $b_0 \in \partial B$ such that $z_0 = a_0 + b_0$. Moreover by the convexity of $A + B$ there exists a supporting hyperplane of $A + B$ at z_0. It is clear that any supporting hyperplane of $A + B$ at z_0 is also a supporting hyperplane of $A + \{b_0\}$ at $z_0 = a_0 + b_0$. If there exists more than one supporting hyperplane of $A + B$ at z_0, then there exists more than one supporting hyperplane of $A + \{b\}$ at z_0, a contradiction. Therefore at each point $z \in \partial(A+B)$ there is a unique supporting hyperplane, showing that $A + B$ is of class C^1. **QED**

In 1991, Krantz and Parks proved in [20] that the sum of a $C^{1,\alpha}$ convex body and a general convex domain is a $C^{1,\alpha}$ convex body. Their proof relies on writing the boundaries of the convex bodies using coordinate systems and is difficult. Another proof is given by Kiselman in [19], which relies on the use of the infimal convolution. Here we will present a simple and new proof of this result that follows from Theorem 51.

Remark 76. The proof for the opposite direction of both Theorem 40 and Theorem 51 does not require the set K to be bounded. The $\varepsilon_0 > 0$ included in the theorem is a global condition to ensure the closed balls remain in K. Thus, if there exists $\varepsilon_0 > 0$ and $C > 0$ such that the convex domain K satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition for all $0 < \varepsilon < \varepsilon_0$, then K is of class $C^{1,\alpha}$.

Theorem 77. *If $A \subset \mathbb{R}^n$ is a convex body of class $C^{1,\alpha}$, $\alpha \in (0, 1]$, and $B \subset \mathbb{R}^n$ is any convex domain, then $A + B$ is of class $C^{1,\alpha}$.*

Proof. We first present the proof for the case $\alpha = 1$, since it is an easy application of Theorem 40. Consider any $z \in \partial(A + B)$ and by Lemma 72 let $a \in \partial A$ and $b \in \partial B$ satisfy $z = a + b$. By Theorem 40, A satisfies the uniform inner ball condition, so there exists $h(a) \in A$ and $r > 0$ such that $a \in \overline{B}(h(a), r) \subset A$. This implies

$$a + b \in \overline{B}(h(a), r) + \{b\} \subset A + B$$

where $\overline{B}(h(a), r) + \{b\}$ is a closed ball of radius r. Therefore $A + B$ satisfies the uniform inner ball condition and hence is of class $C^{1,1}$.

In general, for $\alpha \in (0, 1]$, we know by Theorem 51 there exists constants $\varepsilon_0, C > 0$ such that for each $0 < \varepsilon < \varepsilon_0$, A satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition. Hence for each $a \in \partial A$ and for each $0 < \varepsilon < \varepsilon_0$ there exists $h_\varepsilon(a) \in A$ such that $\overline{B}(h_\varepsilon(a), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) \subset A$ and

$$\text{dist}(a, \overline{B}(h_\varepsilon(a), C\varepsilon^{\frac{1-\alpha}{1+\alpha}})) \leq \varepsilon.$$

Let $z \in \partial(A + B)$ and by Lemma 72, let $a \in \partial A$ and $b \in \partial B$ such that $z = a + b$. Then consider the ball $\overline{B}(h_\varepsilon(a), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) + \{b\} \subset A + B$, and note, for all $0 < \varepsilon < \varepsilon_0$,

$$\text{dist}(a + b, \overline{B}(h_\varepsilon(a), C\varepsilon^{\frac{1-\alpha}{1+\alpha}}) + \{b\}) = \text{dist}(a, \overline{B}(h_\varepsilon(a), C\varepsilon^{\frac{1-\alpha}{1+\alpha}})) \leq \varepsilon$$

so that $A + B$ satisfies the $(C\varepsilon^{\frac{1-\alpha}{1+\alpha}}, \varepsilon)$-approximate inner ball condition for all $0 < \varepsilon < \varepsilon_0$. Therefore, by Theorem 51, $A + B$ is of class $C^{1,\alpha}$. **QED**

4.0 Second Differentiability of Convex Functions

We established in Chapter 2 that a convex function is differentiable almost everywhere, by showing it is locally Lipschitz and applying the Rademacher theorem. But in fact a much more impressive result is true for convex functions. In this chapter we will prove the Alexandrov theorem which states that a convex function is in fact second differentiable almost everywhere.

4.1 Approximating Convex Bodies with Lipschitz Outer Normal Vectors

Recall for a convex body K, the inner parallel body K_r, defined in (26), is given by

$$K_r = \{x \in K : \operatorname{dist}(x, \partial K) \geq r\}.$$

and $K(r)$, defined in (27), is given by

$$K(r) = \bigcup \{\overline{B}(x,r) : \overline{B}(x,r) \subset K\}.$$

Moreover, choosing r small enough, both K_r and $K(r)$ are convex bodies.

Lemma 78. *If a convex body K contains a ball of radius r_0, then*

$$\mathcal{H}^{n-1}(\partial K_r) \leq \mathcal{H}^{n-1}(\partial K \cap \partial K(r)). \tag{63}$$

Proof. By Lemma 34, K_r is a convex body. Observe that

$$\pi_{K_r}(\partial K \cap \partial K(r)) = \partial K_r. \tag{64}$$

Indeed, if $z \in \partial K_r$, then there is $x \in \partial K$, such that $|x - z| = r$. Therefore, $x \in \overline{B}(z,r) \subset K$, and hence $x \in K(r)$. Thus, $x \in \partial K \cap \partial K(r)$, $|x - z| = r \geq \operatorname{dist}(x, K_r)$, and hence $z = \pi_{K_r}(x)$. Now (63) follows from (64) and the fact that π_{K_r} is 1-Lipschitz (Lemma 5). **QED**

The next beautiful result is due to McMullen [22]. While it can be concluded from Alexandrov's theorem, we present here a direct and surprisingly elementary proof which is a small modification of McMullen's argument. In fact, Lemma 79 will play an important role in our proof of Alexandrov's theorem.

Lemma 79. *If $K \subset \mathbb{R}^n$ is a convex body, then* $\lim_{r\to 0^+} \mathcal{H}^{n-1}(\partial K \setminus \partial K(r)) = 0$.

Remark 80. Lemma 79 has the following geometric interpretation: for almost all $x \in \partial K$, there is a closed ball $\overline{B} \subset K$ touching the boundary of K at x, i.e., $x \in \overline{B}$.

Remark 81. The proof presented here of Lemma 79 will follow from the use of dilating the inner parallel bodies of K, but a famous result in convex geometry can also be used. Steiner's formula shows that for a convex body K, the surface area can be found using this derivative style limit:

$$\mathcal{H}^{n-1}(\partial K) = \lim_{\delta\to 0^+} \frac{\mathcal{H}^n(K + \delta\overline{B}(0,1)) - \mathcal{H}^n(K)}{\delta}.$$

where we have $K + \delta\overline{B}(0,1) = \{x \in \mathbb{R}^n : \operatorname{dist}(x,K) \leq \delta\}$, the *outer parallel body* of K.

Proof. Without loss of generality we may assume that $\overline{B}(0,r_0) \subset K$. If $r \in (0,r_0)$, then 0 belongs to the interior of K_r. For $\lambda > 0$ we define

$$\lambda K_r := \{\lambda z : \; z \in K_r\},$$

that is, λK_r is a dilation of K_r. For $r \in (0,r_0)$, let

$$\lambda(r) := \inf\{\lambda > 0 : \; K \subset \lambda K_r\}.$$

Clearly, $K \subset \lambda(r)K_r$. It is easy to see that the function $r \mapsto \lambda(r)$ is non-decreasing and $\lambda(r) \to 1$ as $r \to 0^+$. Indeed, for any $\varepsilon > 0$, $(1+\varepsilon)^{-1}K \subset \operatorname{int} K$, and hence $\delta := \operatorname{dist}((1+\varepsilon)^{-1}K, \partial K) > 0$, so for all $r \in (0,\delta]$

$$(1+\varepsilon)^{-1}K \subset K_r, \qquad \text{i.e.,} \qquad K \subset (1+\varepsilon)K_r.$$

In other words $1 \leq \lambda(r) \leq 1+\varepsilon$ for all $0 < r \leq \delta$ proving that $\lambda(r) \to 1$ as $r \to 0^+$.

It is easy to see that $\pi_K(\partial(\lambda(r)K_r)) = \partial K$. Indeed, if $x \in \partial K$ and $\nu(x)$ is the outer unit normal vector to a supporting hyperplane of K at x, then there is $t \geq 0$ such that

$z := x + t\nu(x) \in \partial(\lambda(r)K_r)$ and it easily follows that $\pi_K(z) = x$. Since π_K is 1-Lipschitz and it maps $\partial(\lambda(r)K_r)$ onto ∂K, we have by Lemma 78, that

$$\begin{aligned}\mathcal{H}^{n-1}(\partial K) &\leq \mathcal{H}^{n-1}(\partial(\lambda(r)K_r)) = \lambda(r)^{n-1}\mathcal{H}^{n-1}(\partial K_r) \leq \lambda(r)^{n-1}\mathcal{H}^{n-1}(\partial K \cap \partial K(r))\\ &\leq \lambda(r)^{n-1}\mathcal{H}^{n-1}(\partial K) \to \mathcal{H}^{n-1}(\partial K) \quad \text{as } r \to 0^+.\end{aligned}$$

Therefore, $\mathcal{H}^{n-1}(\partial K \cap \partial K(r)) \to \mathcal{H}^{n-1}(\partial K)$, as $r \to 0^+$. This completes the proof of Lemma 79. **QED**

Corollary 82. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then it is differentiable a.e. Moreover*

$$f(y) = f(x) + Df(x)(y-x) + O(|y-x|^2) \qquad \text{for almost all } x \in \mathbb{R}^n. \tag{65}$$

Proof. Since the boundary of a ball is parameterized by a smooth convex function, Lemma 44 implies (65) whenever there is a ball in the epigraph of f that touches the graph of f at $(x, f(x))$ and it follows from Lemma 79 that it is true for almost all x. **QED**

Remark 83. Note that the proof of Corollary 82 does not use Rademacher's theorem. Moreover, the estimate (65), is stronger than the a.e. differentiability of f that would follow from an application of Rademacher's theorem.

4.2 Alexandrov's Theorem

The first part of Alexandrov's theorem states that a convex function is twice differentiable a.e. in the sense of Taylor's theorem with the Peano remainder.

Theorem 84. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then it is differentiable a.e. and at almost every point where f is differentiable, there is a symmetric matrix denoted by $D^2f(x)$ such that*

$$\lim_{y\to x} \frac{f(y) - f(x) - Df(x)(y-x) - \frac{1}{2}(y-x)^T D^2 f(x)(y-x)}{|y-x|^2} = 0. \tag{66}$$

To prove this result we will build upon the previous section, where we showed that we can approximate a convex body K with a $C^{1,1}$ convex body $K(R) \subset K$ such that surface area of $\partial K \setminus \partial K(R)$ can be made arbitrarily small for R small. The idea of the proof goes as follows: take a convex function f and intersect the epigraph with a cylinder bounded from above, making a convex body. Then this convex body can be approximated by a $C^{1,1}$ convex body which intersects the boundary outside a set of small measure, and locally this convex body is the graph of a $C^{1,1}$ convex function, let's call it g. Note that on the set of points where $f = g$, ∇f exists and $\nabla f = \nabla g$. Applying the Rademacher theorem to ∇g at almost every point will give us the desired symmetric matrix.

The next result is a direct consequence of Lemma 79 and will be used to prove both Alexandrov's theorem and Theorem 90.

Corollary 85. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then for every $R > 0$ and $\varepsilon > 0$, there is a convex function $g \in C^{1,1}(B^n(0,R))$ such that $g \geq f$ and*

$$\mathcal{L}^n(\{x \in B^n(0,R) :\ f(x) \neq g(x)\}) < \varepsilon. \tag{67}$$

Proof. Let $M := \sup_{\overline{B}^n(0,2R)} f(x)$ and define

$$K := \{(x,y) \in \overline{B}^n(0,2R) \times \mathbb{R} :\ f(x) \leq y \leq M + 2R\}.$$

That is, K is an $(n+1)$-dimensional convex body bounded by the graph of f, the cylinder $\partial B^n(0,2R) \times \mathbb{R}$ and the hyperplane $y = M + 2R$. According to Lemma 79, there is $\delta < R$ such that

$$\mathcal{H}^n(\partial K \setminus \partial K(\delta)) < \varepsilon.$$

Since $K(\delta)$ is the union of closed balls of radius $\delta < R$ that are contained in K, it follows that

$$\overline{B}^n(0,2R) \times \{M + R\} \subset K(\delta),$$

i.e., the intersection of $K(\delta)$ with the hyperplane $y = M + R$ is an n-dimensional closed ball of radius $2R$. Thus, if $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is the orthogonal projection, $\pi(K(\delta)) = \overline{B}^n(0,2R)$, and hence for $x \in \overline{B}^n(0,2R)$, we can define, as in (24)

$$g(x) := \inf\{y :\ (x,y) \in K(\delta)\}.$$

That is, the function $g : \overline{B}^n(0, 2R) \to \mathbb{R}$ parametrizes the bottom part of the boundary of $K(\delta)$. According to Theorem 40, the boundary of $K(\delta)$ is of class $C^{1,1}$ so $g \in C^{1,1}_{\text{loc}}(B^n(0, 2R))$ and hence g is a convex function in $C^{1,1}(B^n(0, R))$. Since $K(\delta)$ is contained in K and hence in the epigraph of f, it follows that $g \geq f$.

Observe that

$$\{x \in B^n(0, R) : f(x) \neq g(x)\} \subset \pi(\partial K \setminus \partial K(\delta))$$

and hence

$$\mathcal{L}^n(\{x \in B^n(0, R) : f(x) \neq g(x)\}) \leq \mathcal{L}^n(\pi(\partial K \setminus \partial K(\delta))) \leq \mathcal{H}^n(\partial K \setminus \partial K(\delta)) < \varepsilon,$$

because the orthogonal projection does not increase the Hausdorff measure and $\mathcal{H}^n$ coincides with the Lebesgue measure in $\mathbb{R}^n$. **QED**

Lemma 86. *Suppose that $f, g : B^n(0, R) \to \mathbb{R}$ are convex, $f \leq g$, and $g \in C^{1,1}(B^n(0, R))$. Then for almost all $x_0 \in \{x \in B^n(0, R) : f(x) = g(x)\}$ we have*

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T D^2 g(x_0)(x - x_0) + o(|x - x_0|^2). \tag{68}$$

While the main tool of the proof of this theorem is the Rademacher theorem, the simplicity of it follows from the use of density points in the set $\{f = g\}$. We will need the following property of density points for the proof of Lemma 86.

Lemma 87. *Let $A \subset \mathbb{R}^n$ be a closed set and 0 a density point of A. Then for any $x \in \mathbb{R}^n$, there exists $y \in A$ such that $|x - y| = o(|x|)$.*

Proof. We first want to show

$$\lim_{|x| \to 0} \frac{\text{dist}(x, A)}{|x|} = 0.$$

Suppose to the contrary that the limit does not equal 0, that is, there exist $\varepsilon_0 > 0$ and $|x_k| \to 0$ such that for all $k \in \mathbb{N}$,

$$\frac{\text{dist}(x_k, A)}{|x_k|} \geq \varepsilon_0.$$

Let $\delta_k := \text{dist}(x_k, A)$, so that we have for all k, $\frac{\delta_k}{|x_k|} \geq \varepsilon_0$. Also note that $\delta_k > 0$ implies $B(x_k, \delta_k/2) \cap A = \emptyset$. Now consider the balls $B(0, |x_k|)$, and note by the density of 0,

$$\lim_{k\to\infty} \frac{\mathcal{L}^n(A\cap B(0,|x_k|))}{\mathcal{L}^n(B(0,|x_k|))} = 1. \tag{69}$$

We can also note that there exists $y_k \in B(0,|x_k|)\cap B(x_k,\delta_k/2)$ such that

$$B(y_k,\delta_k/4) \subset B(0,|x_k|)\cap B(x_k,\delta_k/2).$$

As $B(y_k,\delta_k/4)\cap A = \emptyset$ implies $A\cap B(0,|x_k|) \subset B(0,|x_k|)\setminus B(y_k,\delta_k/4)$ we have,

$$\mathcal{L}^n(A\cap B(0,|x_k|)) \le \mathcal{L}^n(B(0,|x_k|)\setminus B(y_k,\delta_k/4)) = \mathcal{L}^n(B(0,|x_k|)) - \mathcal{L}^n(B(y_k,\delta_k/4))$$

and hence,

$$\frac{\mathcal{L}^n(A\cap B(0,|x_k|))}{\mathcal{L}^n(B(0,|x_k|)} \le 1 - \frac{\mathcal{L}^n(B(y_k,\delta_k/4))}{\mathcal{L}^n(B(0,|x_k|))} = 1 - \frac{\delta_k^n}{4^n|x_k|^n} \le 1 - \frac{\varepsilon_0^n}{4^n}$$

contradicting the definition of density point. The result then follows by noting, as A is closed, for each $x\in\mathbb{R}^n$ there exists a $y\in A$ such that

$$\operatorname{dist}(x,A) = |x-y|$$

QED

Remark 88. To find the y_k explicitly, draw a line segment from 0 to x_k and call z_k the point where this line segment intersects the boundary of $B(x_k,\delta_k/2)$. Then choose y_k to be the midpoint on the line segment between x_k and z_k.

Proof of Lemma 86. It follows from Lemma 44 that f is differentiable at every point of the set $\{f=g\}$ and that $Df = Dg$ in $\{f=g\}$. Since Dg is Lipschitz continuous, Dg is differentiable a.e. by Rademacher's theorem. Therefore, it suffices to prove the result whenever $x_0\in\{f=g\}$ is a density point and Dg is differentiable at x_0.

To simplify notation, without loss of generality, we may assume that $x_0 = 0$, and we need to prove that

$$f(x) - f(0) - Df(0)x - \frac{1}{2}x^T D^2 g(0)x = o(|x|^2).$$

Since $f(0) = g(0)$ and $Df(0) = Dg(0)$, the left hand side equals

$$(f(x) - g(x)) + \left(g(x) - g(0) - Dg(0)x - \frac{1}{2}x^T D^2 g(0)x\right) = (f(x) - g(x)) + o(|x|^2).$$

We used here the fact that g is twice differentiable at 0 (Taylor's theorem with the Peano remainder). Thus it remains to show that $g(x) - f(x) = o(|x|^2)$.

Since 0 is a density point of the set $\{f = g\}$, by Lemma 87 for any $x \in B^n(0, r)$ we can find $y \in \{f = g\}$ such that $|x - y| = o(|x|)$.

Clearly, $f(y) = g(y)$ and $Df(y) = Dg(y)$ by Lemma 44. Therefore,

$$f(x) \geq f(y) + Df(y)(x - y) = g(y) + Dg(y)(x - y),$$

where the inequality is a consequence of Theorem 22. Since $f \leq g$, the above inequality and Lemma 41 yield

$$0 \leq g(x) - f(x) \leq g(x) - g(y) - Dg(y)(x - y) \leq M|x - y|^2 = o(|x|^2).$$

completing the proof. **QED**

We are now ready to prove the first version of the Alexandrov theorem:

Proof of Theorem 84. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Let $R > 0$ and $\varepsilon > 0$ and let g be as in Corollary 85. It follows from Lemma 86 that for almost all $x \in \{f = g\}$, (66) is satisfied with $D^2 f(x) := D^2 g(x)$. Hence (66) holds true in $B(0, R)$ outside a set of measure less than ε. Since it is true for any $R > 0$ and $\varepsilon > 0$, it follows that (66) is satisfied almost everywhere. **QED**

4.3 The Differentiability of the Subdifferential of a Convex Function

There is also a second version of the Alexandrov theorem which says that the subdifferential ∂f is differentiable a.e.

Theorem 89. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then for all $x \in \mathbb{R}^n$ where f is twice differentiable as in* (66)*, we have*

$$\lim_{y \to x} \sup_{\sigma_y \in \partial f(y)} \frac{|\sigma_y - Df(x) - D^2 f(x)(y-x)|}{|y-x|} = 0. \tag{70}$$

The usual way to prove the first version of Alexandrov's theorem (Theorem 84) is to show Theorem 89 first and conclude Theorem 84 from it. In this thesis we will prove Theorem 89 directly from Theorem 84. If f is twice differentiable at 0 as in (66), then we have

$$f(x) = f(0) + Df(0)x + \frac{1}{2} x^T D^2 f(0) x + R(x) = f(0) + Df(0)x + \langle Ax, x \rangle + R(x),$$

where $A = \frac{1}{2} D^2 f(0)$ and $R(x) = o(|x|^2)$. Note that

$$a(r) := \sup_{0<|x| \le 2r} \frac{|R(x)|}{|x|^2} \to 0 \qquad \text{as } r \to 0^+.$$

Moreover,

$$|R(x)| \le a\Big(\frac{|x|}{2}\Big)\,|x|^2 \le a(|x|)|x|^2.$$

Proof of Theorem 89. Let f be twice differentiable at x as in (66). We need to prove (70). Without loss of generality we may assume that $x = 0$, and hence we need to prove that

$$\lim_{x \to 0} \frac{\sigma_x - Df(0) - D^2 f(0) x}{|x|} = 0 \quad \text{for any } \sigma_x \in \partial f(x).$$

For $x, y \neq 0$, we have

$$f(x) = f(0) + Df(0)x + \langle Ax, x \rangle + R(x), \quad f(y) = f(0) + Df(0)y + \langle Ay, y \rangle + R(y).$$

Since $f(x) + \langle \sigma_x, y - x \rangle \le f(y)$, we have

$$\langle \sigma_x, y-x \rangle \le f(y) - f(x) = Df(0)(y-x) + \langle A(x+y), y-x \rangle + R(y) - R(x).$$

We used here the fact that A is symmetric and hence $\langle Ax, y\rangle = \langle Ay, x\rangle$. Let

$$y = x + w, \quad \text{where} \quad w = \sqrt{a(|x|)}\,|x|z,\ |z| = 1.$$

Then

$$\langle \sigma_x, w\rangle \le Df(0)w + \langle A(2x + w), w\rangle + R(y) - R(x),$$

$$\langle \sigma_x - Df(0) - 2Ax, w\rangle \le \langle Aw, w\rangle + R(y) - R(x).$$

If $|x|$ is sufficiently small, then $a(|x|) \le 1$ and hence $|w| \le |x|$, so $|y| \le 2|x|$. Therefore,

$$|R(y)| \le a\Big(\frac{|y|}{2}\Big)\,|y|^2 \le 4a(|x|)|x|^2, \qquad |R(y) - R(x)| \le 5a(|x|)|x|^2.$$

Taking the supremum over all z with $|z| = 1$ we get

$$|\sigma_x - Df(0) - 2Ax|\sqrt{a(|x|)}|x| \le |A|a(|x|)|x|^2 + 5a(|x|)|x|^2,$$

and hence

$$\frac{|\sigma_x - Df(0) - 2Ax|}{|x|} \le (|A| + 5)\sqrt{a(|x|)} \to 0 \quad \text{as } x \to 0.$$

Since $2A = D^2f(0)$, the result follows. **QED**

4.4 Approximating Convex Functions Globally with Continuously Differentiable Convex Functions with Lipschitz Gradient

The following result was originally proved by Hajłasz and Azagra in [4], but the proof is technical. The previous results regarding $C^{1,1}$ convex bodies and functions allows us to find a simpler and more intuitive proof.

Theorem 90. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, for every $R > 0$ and for every $\varepsilon > 0$, there exists a convex function $g \in C^{1,1}(\mathbb{R}^n)$, $g \ge f$, such that*

$$\mathcal{L}^n(\{x \in B^n(0, R) :\ f(x) \ne g(x)\}) < \varepsilon.$$

The only difference between Corollary 85 and Theorem 90 is that the function g in Corollary 85 is only defined on the ball $B^n(0,R)$. The main step in the proof of Theorem 90 will be to show that the function g can be extended from a ball $B^n(0,R-\delta)$ to a convex function of class $C^{1,1}(\mathbb{R}^n)$. We will do it by gluing the function g with a quadratic function of the form $a|x|^2-b$ and we need to know how to glue convex functions while maintaining their smoothness.

The maximum of two convex functions

$$\max\{u,v\}=\frac{u+v+|u-v|}{2}$$

is convex, but even if $u,v\in C^\infty$, the maximum $\max\{u,v\}$ need not be C^1. To overcome this difficulty, we will use the so called smooth maximum that was introduced in [2].

Let $\theta\in C^\infty(\mathbb{R})$ be such that $\theta(t)=|t|$ if and only if $|t|\geq 1$, θ is convex, $\theta(t)=\theta(-t)$ for all t, and 1-Lipschitz.

It easily follows that $\theta(t)>0$ for all t and $|\theta'(t)|<1$ if and only if $|t|<1$. Then, we define the *smooth maximum* function $\mathcal{M}:\mathbb{R}^2\to\mathbb{R}$ as,

$$\mathcal{M}(x,y):=\frac{x+y+\theta(x-y)}{2}.$$

It is easy to see that $\mathcal{M}$ is smooth, convex and

$$\mathcal{M}(x,y)=\max\{x,y\}\quad\text{whenever}\quad |x-y|\geq 1. \tag{71}$$

It is also not difficult to prove that $\mathcal{M}(x,y)$ is non-decreasing in x and y, because partial derivatives of $\mathcal{M}$ are non-negative, see [2, Lemma 2.1(viii)]. This observation and convexity of $\mathcal{M}$ yield (see [2, Proposition 2.2(i)])

Lemma 91. *If $u,v:U\to\mathbb{R}$ are convex functions defined in an open convex set $U\subset\mathbb{R}^n$, then $\mathcal{M}(u,v):U\to\mathbb{R}$ is convex.*

It is also obvious that if $u,v\in C^{1,1}_{\rm loc}(U)$, then $\mathcal{M}(u,v)\in C^{1,1}_{\rm loc}(U)$.

We will use the smooth maximum to prove the following extension result and Theorem 90 follows immediately from Proposition 92.

Proposition 92. *Let $h \in C^{1,1}_{\rm loc}(B^n(0,R))$ be a convex function. Then, for every $r \in (0,R)$, there is a convex function $H \in C^{1,1}(\mathbb{R}^n)$, such that*

$$H(x) = h(x) \quad \text{whenever } |x| \le r. \tag{72}$$

Remark 93. If $h \in C^k$, $k \in \mathbb{N} \cup \{\infty\}$, then $H \in C^k(\mathbb{R}^n)$. The proof remains the same.

Proof. Choose $\rho \in (r,R)$ and let

$$m := \inf_{|x|\le r} h, \qquad M := \sup_{|x|=\rho} h.$$

Then, we can find $a, b > 0$ such that the function $q(x) := a|x|^2 - b$ satisfies

$$q(x) < m - 1 \qquad \text{if } |x| \le r \tag{73}$$

$$q(x) > M + 1 \qquad \text{if } |x| = \rho, \tag{74}$$

and we define

$$H(x) := \begin{cases} \mathcal{M}(h(x), q(x)) & \text{if } |x| \le \rho, \\ q(x) & \text{if } |x| > \rho. \end{cases}$$

It follows from (73) if $|x| \le r$, then $h(x) > q(x) + 1$ so by (71), we have

$$H(x) = \mathcal{M}(h(x), q(x)) = h(x) \quad \text{if } |x| \le r$$

and the condition (72) is satisfied. It follows from (74) that there is $\varepsilon > 0$ such that $q(x) > h(x) + 1$ if $\rho \le |x| \le \rho + \varepsilon$ and hence by (71),

$$\mathcal{M}(h(x), q(x)) = q(x) \quad \text{when } \rho \le |x| \le \rho + \varepsilon.$$

Therefore, the convex functions $q(x) \in C^{1,1}(\mathbb{R}^n)$ and $\mathcal{M}(h(x), q(x)) \in C^{1,1}_{\rm loc}(B^n(0,R))$ coincide in the annulus $\rho \le |x| \le \rho + \varepsilon$ and hence H is convex in $\mathbb{R}^n$ with $H \in C^{1,1}_{\rm loc}(\mathbb{R}^n)$. Since $H = q \in C^{1,1}$ outside the compact ball $\overline{B}^n(0,\rho)$, it follows that $H \in C^{1,1}(\mathbb{R}^n)$. **QED**

5.0 Support Functions and Boundary Properties of Convex Bodies

5.1 The Support Function

Previously when discussing convex bodies we noted that at each point on the boundary of the body there is a supporting hyperplane (Theorem 8). While this description can be thought of as an intrinsic view of the boundary structure of K we now want to consider an extrinsic view of the boundary of K. We recall the hyperplane with normal vector u is denoted

$$H_b(u) := \{x \in \mathbb{R}^n : \langle x, u \rangle = b\}$$

and the half space with boundary $H_b(u)$ and u outer normal vector as

$$H_b^-(u) := \{x \in \mathbb{R}^n : \langle x, u \rangle \leq b\}.$$

Geometrically, for a fixed u, increasing b translates the hyperplane $H_b(u)$ in the direction of u and decreasing b translates $H_b(u)$ in the direction of $-u$.

Let K be a convex body. We say that $H(u, K)$ is the *supporting hyperplane of K in the direction of u* if there exists $b \in \mathbb{R}$ that satisfies

$$K \subset H_b^-(u) \quad \text{and} \quad H_b(u) \cap \partial K \neq \emptyset.$$

It is clear, by the convexity of K, for a fixed u, the choice of $b \in \mathbb{R}$ is unique. Thus we can define a mapping that sends the outer normal vector u to the corresponding b such that $H_b(u) = H(u, K)$. If $H_b(u)$ is the supporting hyperplane of K in the direction of u, then we define the function $\sigma_K : \mathbb{R}^n \to \mathbb{R}$ by $\sigma_K(u) = b$ and we call σ_K the *support function of K*. Hence we can write the supporting hyperplane in the direction of u as,

$$H(u, K) := \{x \in \mathbb{R}^n : \langle x, u \rangle = \sigma_K(u)\} = H_{\sigma_K(u)}(u).$$

We now seek to provide an explicit definition for the support function of a convex body K. The standard construction goes as follows: pick any $u \in \mathbb{R}^n$ and choose b_0 so that,

$K \subset H_{b_0}^-(u)$. This is possible by the compactness of K. Note that u is pointing outwards of $H_{b_0}^-(u)$ and thus also pointing outwards of K. If we define the set

$$S_K(u) := \{\langle x, u\rangle : x \in K\} \tag{75}$$

then we see that b_0 is an upper bound of $S_K(u)$ implying the supremum of $S_K(u)$ exists and is finite. Moreover, by continuity of the mapping $x \mapsto \langle x, u\rangle$ and the compactness of K we see that the supremum is achieved at some point in K, and specifically some point on ∂K. Let $x_u \in \partial K$ be a point satisfying $\langle x_u, u\rangle = \sup S_K(u)$. So if we consider the hyperplane,

$$H_{\sup S_K(u)}(u) = \{x \in \mathbb{R}^n : \langle x, u\rangle = \sup S_K(u)\}$$

then we see that $K \subset H^-_{\sup S_K(u)}(u)$ and $x_u \in K \cap H_{\sup S_K(u)}(u)$ so that $H_{\sup S_K(u)}(u)$ is the supporting hyperplane of K in the direction of u. Therefore it is clear that $\sigma_K(u) = \sup S_K(u)$, i.e. the support function of K can be explicitly defined as,

$$\sigma_K(u) = \sup_{x\in K} \langle x, u\rangle.$$

While this construction provides a clear motivation for an explicit equation of the support function we present an alternative construction that shows the reader the geometric uses of the support function. Consider, as before, the half space $H_{b_0}^-(u)$, where $K \subset H_{b_0}^-(u)$ and suppose K does not intersect $H_{b_0}(u)$. Then we can consider $b < b_0$ so that $K \subset H_b^-(u)$ and decrease the value of b, translating the hyperplane $H_b(u)$ in the direction of $-u$, until we find some value b_u such that $H_{b_u}(u)$ is the supporting hyperplane of K for some point $x \in K$. This value b_u is precisely the support function of K defined at u, i.e. $b_u = \sigma_K(u)$.

With this in mind we can establish some new notation to make the uses of the support function clearer. We define the *face of K in the direction of u* as,

$$F(u, K) := \{x \in K : \langle x, u\rangle = \sigma_K(u)\} = H(u, K) \cap K.$$

Remark 94. It is interesting to note the faces of convex bodies are preserved under addition and dilation, that is, if $K, L \subset \mathbb{R}^n$ are convex bodies and $c > 0$ we have,

$$F(u, K + L) = F(u, K) + F(u, L) \quad \text{and} \quad F(u, cK) = cF(u, K).$$

This shows that there is in fact some additional structure to the boundaries of convex bodies. See [28] for more about the structure of convex bodies and their faces.

As before we can also consider the half space bounded by $H(u, K)$, with outer unit normal u, and denote this by

$$H^{-}(u, K) := \{x \in \mathbb{R}^n : \langle x, u \rangle \leq \sigma_K(u)\}.$$

As $K \subset H^{-}(u, K)$ for all $u \in \mathbb{R}^n \setminus \{0\}$, it is then easy to see that these half spaces define K.

Proposition 95. *For a convex body K,*

$$K = \bigcap_{u \in \mathbb{R}^n \setminus \{0\}} H^{-}(u, K) = \bigcap_{u \in \mathbb{R}^n \setminus \{0\}} \{x : \langle x, v \rangle \leq \sigma_K(u)\}.$$

Proof. $K \subset \bigcap_{u \in \mathbb{R}^n \setminus \{0\}} H^{-}(u, K)$, follows immediately from the previous statement of $K \subset H^{-}(u, K)$ for all $u \in \mathbb{R}^n \setminus \{0\}$. If $x \in \bigcap_{u \in \mathbb{R}^n \setminus \{0\}} H^{-}(u, K)$ but $x \notin K$, then by the compactness and convexity of K, and Corollary 10, there would exist a separating hyperplane between x and K. Namely there exists an $\alpha \in \mathbb{R}$ and $v_0 \in \mathbb{R}^n \setminus \{0\}$ such that $\langle k, v_0 \rangle < \alpha < \langle x, v_0 \rangle$ for all $k \in K$. As α is an upper bound for the set $S_K(v_0)$, defined in (75), we have that $\sigma_K(v_0) \leq \alpha$ so that,

$$\sigma_K(v_0) \leq \alpha < \langle x, v_0 \rangle$$

which contradicts $x \in H^{-}(v_0, K)$. **QED**

Remark 96. Proposition 95 is an analogous result to Corollary 9, where both seek to describe the convex body K in terms of supporting hyperplanes.

By construction, the support function is useful to identify and characterize the boundary structure of convex bodies. Specifically the faces of a convex body can now be identified with a unique outer unit normal vector allowing us to move between the geometry of the convex body and analysis on the support function seamlessly.

5.2 Properties of the Support Function

As the support function σ_K is defined using the supremum and inner product, much of the basic properties of σ_K are easy to establish. Recall the explicit equation for the support function is given as

$$\sigma_K(u) = \sup_{k \in K} \langle k, u \rangle.$$

Theorem 97 (Properties of the Support Function)**.** *For a convex body K, we have*

(1) *σ_K is subadditive i.e. $\sigma_K(u+v) \leq \sigma_K(u) + \sigma_K(v)$;*

(2) *σ_K is 1-homogeneous i.e. for $t > 0$, $\sigma_K(tu) = t\sigma_K(u)$;*

(3) *σ_K is convex;*

(4) *$-\sigma_K(-u) = \inf_{k \in K} \langle k, u \rangle$.*

Proof. Recall that the supremum satisfies the following properties for sets $A, B \subset \mathbb{R}$ which are bounded;

(i) $\sup(A + B) \leq \sup A + \sup B$;

(ii) $\sup(tA) = t \sup(A)$ for $t \geq 0$;

(iii) $\sup(-A) = -\inf(A)$.

Thus (1) and (2) follow from (i) and (ii), respectively, and the bilinearity of $\langle \cdot, \cdot \rangle$, (3) is an obvious consequence of (1) and (2), and (4) follows from (iii). **QED**

Recall the subdifferential of a function at x is given by,

$$\partial f(x) = \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x \rangle \text{ for all } y \in \mathbb{R}^n\}.$$

Moreover recall, by Theorem 15, if f is convex, then for all $x \in \mathbb{R}^n$, $\partial f(x) \neq \emptyset$. The support function, being a geometrically inspired function, has a geometrically beautiful subdifferential.

Theorem 98. *For a convex body K, and $u \in \mathbb{R}^n \setminus \{0\}$,*

$$\partial \sigma_K(u) = F(u, K).$$

Proof. Let $x \in F(u, K)$. Then by definition, $x \in K$ and $\langle x, u \rangle = \sigma_K(u)$. Also by the definition of the support function we have for any $v \in \mathbb{R}^n \setminus \{0\}$ and all $k \in K$, $\sigma_K(v) \geq \langle k, v \rangle$, and namely, as $x \in K$,

$$\sigma_K(v) \geq \langle x, v \rangle = \langle x, v \rangle + \sigma_K(u) - \langle x, u \rangle = \sigma_K(u) + \langle x, v - u \rangle.$$

This is precisely the definition of the subdifferential and hence $x \in \partial \sigma_K(u)$.

Now suppose that $x \in \partial \sigma_K(u)$. This implies for any $v \in \mathbb{R}^n$,

$$\sigma_K(v) \geq \sigma_K(u) + \langle x, v - u \rangle.$$

Specifically choosing $v = 2u$ and noting $\sigma_K(2u) = 2\sigma_K(u)$ yields,

$$2\sigma_K(u) = \sigma_K(2u) \geq \sigma_K(u) + \langle x, 2u - u \rangle \implies \sigma_K(u) \geq \langle x, u \rangle.$$

Similarly, choosing $v = 0$, and noting that $\sigma_K(0) = 0$, yields,

$$0 = \sigma_K(0) \geq \sigma_K(u) + \langle x, 0 - u \rangle \implies \langle x, u \rangle \geq \sigma_K(u).$$

Therefore we have that $\sigma_K(u) = \langle x, u \rangle$. This implies that $x \in H(u, K)$. We need only show that $x \in K$. As we have for all $v \in \mathbb{R}^n$,

$$\sigma_K(v) \geq \sigma_K(u) + \langle x, v - u \rangle = \langle x, v \rangle$$

This shows for all $v \in \mathbb{R}^n \setminus \{0\}$, that $x \in H^-(v, K)$, so that by Proposition 95 we have $x \in K$. Therefore $x \in H(u, K) \cap K = F(u, K)$. **QED**

Recall by Theorem 27 a convex function f is differentiable at x if and only if $\partial f(x)$ is a singleton, which implies $\partial f(x) = \{\nabla f(x)\}$. Given Theorem 98, this implies,

Theorem 99. *For a convex body K, σ_K is differentiable at u if and only if $F(u, K) = \{x\}$ for some $x \in \partial K$. That is, σ_K is only differentiable at the outer normals of supporting hyperplanes that intersect K at a single point.*

5.3 Rectifiability of the Normal Directions

We define the set

$$\Sigma^d(\partial f) := \{x \in \mathbb{R}^n : \dim \partial f(x) \geq d\}$$

where we recall, by Theorem 16, $\partial f(x)$ is a compact convex set so the dimension is well defined to be the dimension of the affine hull of $\partial f(x)$. The main tool we will need for this section comes from a theorem of Zajíček [29] regarding coverings of the set $\Sigma^k(\partial f)$ for convex f. Before we state the result we define some new objects.

We say that a set $G \subset \mathbb{R}^n$ is a $(c-c)$-graph in the direction of x_i, if there exists a function $f : \mathbb{R}^{n-1} \to \mathbb{R}$, convex functions $g, h : \mathbb{R}^{n-1} \to \mathbb{R}$ such that $f = g - h$, and

$$G = \{x \in \mathbb{R}^n : x_i = f(x_1, \dots, x_{i-1}, x_{i+1}, \dots x_n)\}.$$

Moreover, we call f a $(c-c)$-function.

Similarly we can define a $(c-c)_k$-graph. Let π be a permutation of $\{1, \dots, n\}$. If there exist functions $f_{\pi(j)} : \mathbb{R}^{n-k} \to \mathbb{R}$, convex functions $g_{\pi(j)}, h_{\pi(j)} : \mathbb{R}^{n-k} \to \mathbb{R}$ such that $f_{\pi(j)} = g_{\pi(j)} - h_{\pi(j)}$, for $j = 1, \dots, k$, and $F : \mathbb{R}^{n-k} \to \mathbb{R}^k$ defined by

$$F(x_{\pi(k+1)}, \dots, x_{\pi(n)}) = (f_{\pi(1)}(x_{\pi(k+1)}, \dots, x_{\pi(n)}), \dots, f_{\pi(k)}(x_{\pi(k+1)}, \dots, x_{\pi(n)}))$$

then we call F a $(c-c)_k$-function and if we define

$$G := \{x \in \mathbb{R}^n : (x_{\pi(1)}, \dots, x_{\pi(k)}) = F(x_{\pi(k+1)}, \dots, x_{\pi(n)})\}$$

then we call G a $(c-c)_k$-graph in the directions of $x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(k)}$.

Remark 100. We call F a $(c-c)_k$-function because F has k component functions all of which are $(c-c)$. Also we note that a $(c-c)_k$-graph is a surface of dimension $n-k$.

The theorem of Zajíček can be stated as follows,

Theorem 101. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then $\Sigma^k(\partial f)$ can be covered by a countable number of $(c-c)_k$-graphs.*

The proof presented in this thesis is a direct extension of the one given in [15]. A function is *strongly convex* if there exists some $\mu > 0$ such that $f(x) - \mu|x|^2$ is convex. It is obvious that a strongly convex function is convex by adding $\mu|x|^2$ and noting the sum of convex functions is convex. Recall a function is *coercive* if $|x| \to \infty$ implies $f(x) \to \infty$. Thus the following lemma says that the difference of a strongly convex and linear function is coercive.

Proposition 102. *Given a strongly convex function,* $f : \mathbb{R}^n \to \mathbb{R}$, *and any* $\ell : \mathbb{R}^n \to \mathbb{R}$ *linear,*

$$\lim_{|x|\to\infty} (f(x) - \ell(x)) = \infty. \tag{76}$$

Proof. As f is strongly convex, there exists $\mu > 0$ such that $f(x) = \mu|x|^2$ is convex. By (17) every convex function is bounded below by an affine function and thus $f(x) - \mu|x|^2$ is bounded below by some affine function $\alpha(x)$. The result follows by noting that f is bounded below by the paraboloid $\mu|x|^2 + \alpha(x) - \ell(x)$ which is coercive. **QED**

The next result is a generalization of Proposition 31 for convex functions.

Lemma 103. *If* $f : \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R}$ *is convex and coercive, then* $F(x) := \inf_{y\in\mathbb{R}^\ell} f(x, y)$ *defines a convex function* $F : \mathbb{R}^k \to \mathbb{R}$.

Proof. This follows by noting that for any $y_1, y_2 \in \mathbb{R}^\ell$,

$$\begin{aligned} F(\lambda x_1 + (1-\lambda)x_2) &\leq f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\ &= f(\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)) \leq \lambda f(x_1, y_1) + (1-\lambda) f(x_2, y_2) \end{aligned}$$

and taking the infima over $y_1 \in \mathbb{R}^\ell$ and $y_2 \in \mathbb{R}^\ell$. **QED**

Denote by M_k the set of increasing k-multi-indices. Thus, we can write the set

$$M_k := \{I = (i_1, \ldots, i_k) \in \mathbb{N}^k : i_1 < i_2 < \cdots < i_k\}.$$

Then for $I = (i_1, \ldots, i_k) \in M_k$, we define the projection map $\pi_I : \mathbb{R}^n \to \mathbb{R}^k$ by

$$\pi_I(x) = (x_{i_1}, x_{i_2}, \ldots, x_{i_k}).$$

For $I = (i_1, \ldots, i_k) \in M_k$, $\alpha = (\alpha_{i_1}, \ldots, \alpha_{i_k}) \in \mathbb{R}^k$ and $\beta = (\beta_{i_1}, \ldots, \beta_{i_k}) \in \mathbb{R}^k$ we consider the sets,

$$A^I_{\alpha,\beta} = \{x \in \mathbb{R}^n : [\alpha_{i_1}, \beta_{i_1}] \times \ldots \times [\alpha_{i_k}, \beta_{i_k}] \subset \pi_I(\partial f(x))\}.$$

We then define the set

$$A := \bigcup_{I \in M_k} \bigcup_{\substack{\alpha,\beta \in \mathbb{Q}^k \\ \alpha < \beta}} A^I_{\alpha,\beta}. \tag{77}$$

where we say $\alpha < \beta$ if $\alpha_{i_j} < \beta_{i_j}$ for all $j \in \{1, \ldots, k\}$. Moreover, for $I \in M_k$, we can denote $[\alpha, \beta]_I = [\alpha_{i_1}, \beta_{i_1}] \times \ldots \times [\alpha_{i_k}, \beta_{i_k}]$ and $\mathbb{R}^I = \operatorname{span}(e_{i_1}, \ldots, e_{i_k})$.

Lemma 104. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then $A = \Sigma^k(\partial f)$.*

Proof. If $x \in A$, then for some multi index $I \in M_K$ and some $\alpha < \beta \in \mathbb{Q}^k$ we have $x \in A^I_{\alpha,\beta}$. Hence $[\alpha, \beta]_I \subset \pi_I(\partial f(x))$, which implies that $\partial f(x)$, being a convex set, must be contained in the affine hull of dimension at least k, proving that $\dim \partial f(x) \geq k$, i.e. $x \in \Sigma^k(\partial f)$. Now similarly if $x \in \Sigma^k(\partial f)$, then $\partial f(x)$ has affine hull of dimension at least k, denoted by $\operatorname{aff}(\partial f(x))$. Hence there exists some multi index $I \in M_k$ such that $\pi_I(\operatorname{aff}(\partial f(x))) = \mathbb{R}^I$ meaning that $\pi_I(\partial f(x))$ has dimension of k in $\mathbb{R}^I$ so that there exists some k-dimensional box, $[\alpha, \beta]_I \subset \pi_I(\partial f(x))$. Therefore $x \in A$. **QED**

To prove Theorem 101, the idea is to cover each $A^I_{\alpha,\beta}$ by a $(c-c)_k$-graph. Note by Theorem 28

$$\Sigma^k(\partial f) = \Sigma^k(\partial(f + |\cdot|^2)),$$

where $f(x) + |x|^2$ is strongly convex, and the result will follow from Lemma 105.

Lemma 105. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strongly convex function. Then each $A^I_{\alpha,\beta}$ is contained in a $(c-c)_k$-graph.*

Proof of Lemma 105. Without loss of generality we will show $A^{(1,\ldots,k)}_{\alpha,\beta}$ is contained in a $(c-c)_k$ graph. For $s := (s_1, \ldots, s_k, 0, \ldots 0) \in \mathbb{R}^n$, let $f_s(x) = f(x) - s \cdot x$, which is convex being the sum of two convex functions, and note that by Proposition 102, $f_s(x)$ is coercive. Thus $g_s : \mathbb{R}^{n-k} \to \mathbb{R}$ defined by:

$$g_s(x_{k+1}, \ldots, x_n) = \inf_{(x_1,\ldots,x_k) \in \mathbb{R}^k} f_s(x_1, \ldots, x_n)$$

is well defined and convex by Lemma 103. Consider $a \in A_{\alpha,\beta}^{(1,\dots,k)}$, implying $[\alpha,\beta]_{(1,\dots,k)} \subset \pi_{(1,\dots,k)}(\partial f(a))$. In particular $\alpha = (\alpha_1,\dots,\alpha_k) \in \pi_{(1,\dots,k)}(\partial f(a))$ and hence there exists some $\gamma_{k+1},\dots,\gamma_n \in \mathbb{R}$ such that $\alpha' := (\alpha_1,\dots,\alpha_k,\gamma_{k+1},\dots,\gamma_n) \in \partial f(a)$. Thus,

$$\begin{aligned} f(a+s) &\geq f(a) + \langle(\alpha_1,\dots,\alpha_k,\gamma_{k+1},\dots,\gamma_n),(s_1,\dots,s_k,0\dots,0)\rangle \\ &= f(a) + \langle(\alpha_1,\dots,\alpha_k),(s_1,\dots,s_k)\rangle \end{aligned}$$

which can be rewritten as,

$$f(a+s_1e_1+\dots+s_ke_k) \geq f(a) + \alpha_1 s_1 + \dots + \alpha_k s_k. \tag{78}$$

As the values $\gamma_{k+1},\dots,\gamma_n$ have no impact on the above inequalities, we define a new vector $\widetilde{\alpha} = (\alpha_1,\dots,\alpha_k,0\dots,0)$. Thus (78) is equivalent to,

$$f_{\widetilde{\alpha}}(a+s_1e_1+\ldots+s_ke_k) \geq f_{\widetilde{\alpha}}(a).$$

As this is true for arbitrary $(s_1,\dots,s_k,0,\dots 0) \in \mathbb{R}^n$ we have the function

$$(x_1,\dots,x_k) \mapsto f_\alpha(x_1,\dots,x_k,a_{k+1},\dots,a_n)$$

achieves its minimum at $(a_1,\dots,a_k)$. Therefore,

$$g_{\widetilde{\alpha}}(a_{k+1},\dots,a_n) = f_{\widetilde{\alpha}}(a_1,\dots,a_n) = f(a) - \alpha_1 a_1 - \dots - \alpha_k a_k \tag{79}$$

By a similar argument we can also produce the same result for the vector

$$\widetilde{\beta} := (\beta_1,\alpha_2,\dots,\alpha_k,0,\dots,0),$$

i.e. $(\beta_1,\alpha_2,\dots,\alpha_k) \in \pi_I(\partial(f(a))$, and thus,

$$g_{\widetilde{\beta}}(a_{k+1},\dots,\alpha_n) = f_{\widetilde{\beta}}(a_1,\dots,a_n) = f(a) - \beta_1 a_1 - \alpha_2 a_2 - \dots - \alpha_k a_k. \tag{80}$$

Subtracting (79) and (80) and noting $\alpha_1 < \beta_1$ we have,

$$a_1 = \frac{1}{\alpha_1 - \beta_1}(g_{\widetilde{\alpha}}(a_{k+1},\dots,a_n) - g_{\widetilde{\beta}}(a_{k+1},\dots,a_n)).$$

Therefore we can define $h_1 : \mathbb{R}^{n-k} \to \mathbb{R}$ by $h_1(x_{k+1},\dots,x_n) = \frac{1}{\alpha_1-\beta_1}(g_{\widetilde{\alpha}}(x_{k+1},\dots,x_n) - g_{\widetilde{\beta}}(x_{k+1},\dots,x_n)$ and we see h_1 is a $(c{-}c)$-function satisfying $h_1(a_{k+1},\dots,a_n) = a_1$. Repeating this process yields, $h_2, \ldots, h_k : \mathbb{R}^{n-k} \to \mathbb{R}$ such that each h_i is a $(c-c)$-function and $h_i(a_{k+1}, \ldots, a_n) = a_i$. Defining $h : \mathbb{R}^{n-k} \to \mathbb{R}^k$ by

$$h(x_{k+1}, \ldots, x_n) = (h_1(x_{k+1}, \ldots, x_n), \ldots, h_k(x_{k+1}, \ldots, x_n))$$

we have that h is a $(c-c)_k$-function and $a \in \{x \in \mathbb{R}^n : (h(x_1, \ldots, x_k), x_{k+1}, \ldots, x_n)\}$. This shows

$$A_{\alpha,\beta}^{(1,\ldots,k)} \subset \{x \in \mathbb{R}^n : (h(x_1, \ldots, x_k), x_{k+1}, \ldots, x_n)\}$$

where $\{x \in \mathbb{R}^n : (h(x_{k+1}, \ldots, x_n), x_{k+1}, \ldots, x_n)\}$ is a $(c-c)_k$-graph. **QED**

Remark 106. The $(c-c)_k$-graph we constructed in the proof of Lemma 105 can be written in the form

$$\{x \in \mathbb{R}^n : x = (h(x_{k+1}, \ldots, x_n), x_{k+1}, \ldots, x_n)\} = H(\mathbb{R}^{n-k})$$

where $H : \mathbb{R}^{n-k} \to \mathbb{R}^n$ is defined as

$$H(x_{k+1}, \ldots, x_n) = (h_1(x_{k+1}, \ldots, x_n), \ldots, h_k(x_{k+1}, \ldots, x_n), x_{k+1}, \ldots, x_n).$$

By the convexity of σ_K we can apply Theorem 101 directly to σ_K to show that $\Sigma^d(\partial\sigma_K)$ can be covered by countably many $(c-c)_d$-graphs. But given σ_K is 1-homogeneous and convex we can in fact restrict $\Sigma^d(\partial\sigma_K)$ to $\mathbb{S}^{n-1}$ and show that $\Sigma^d(\partial\sigma_K) \cap \mathbb{S}^{n-1}$ can be covered by countably many $(c-c)_{d+1}$-graphs. By Theorem 98, this has a remarkable geometrical implication. If we define the set,

$$N_d(K) := \mathbb{S}^{n-1} \cap \Sigma^d(\partial\sigma_K) = \{u \in \mathbb{S}^{n-1} : \dim(\partial\sigma_K(u)) \geq d\}$$

then Theorem 98 allows us to write this as,

$$N_d(K) = \{u \in \mathbb{S}^{n-1} : \dim F(u, K) \geq d\}.$$

This means we will show that the normal directions associated with d-dimensional faces on a convex body can be covered by countably many surfaces of dimension $n-d-1$.

We first establish that the subdifferential of a 1-homogeneous convex function is invariant under dilations.

Lemma 107. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a 1-homogeneous convex function and $c > 0$. Then for every $x \in \mathbb{R}^n$, $\partial f(x) = \partial f(cx)$.*

Proof. By the 1-homogeneity of f we have for $v \in \partial f(cx)$, and for any $y \in \mathbb{R}^n$,

$$cf(y) = f(cy) \geq f(cx) + \langle v, cy - cx\rangle = cf(x) + c\langle v, y - x\rangle = c(f(x) + \langle v, y - x\rangle)$$

implying that $v \in \partial f(x)$ and hence $\partial f(cx) \subset \partial f(x)$. Similarly for $v \in \partial f(x)$, and for any $y \in \mathbb{R}^n$,

$$f(cy) = cf(y) \geq cf(x) + c\langle v, y - x\rangle = f(cx) + \langle v, cy - cx\rangle.$$

As this is true for any $y \in \mathbb{R}^n$, we have for any z, there exists $y \in \mathbb{R}^n$ such that $z = cy$ and

$$f(z) \geq f(cx) + \langle v, z - cx\rangle.$$

Therefore $v \in \partial f(cx)$. **QED**

Theorem 108. *Let $K \subset \mathbb{R}^n$ be a convex body. Then there exists countably many functions $h_i : \mathbb{R}^{n-d-1} \to \mathbb{R}^n$ such that*

$$N_d(K) \subset \bigcup_{i=1}^{\infty} (h_i)(\mathbb{R}^{n-d-1}),$$

where each h_i is the composition of a Lipschitz function and a $(c-c)_{d+1}$-function.

Remark 109. The idea of the proof is to project hemispheres of $\mathbb{S}^{n-1}$ onto hyperplanes and apply Theorem 101. In the following proof we consider only the case of projecting $\{x \in \mathbb{S}^{n-1} : x_n < 0\}$ onto the hyperplane $\{x \in \mathbb{R}^n : x_n = -1\}$. Using the notation of this thesis we have

$$H_{-1}(e_n) = \{x \in \mathbb{R}^n : x_n = -1\}.$$

Proof. Let $y \in N_d(K)$. Define the bottom hemisphere of $\mathbb{S}^{n-1}$ as $\mathbb{S}^{n-1}_- = \{x \in \mathbb{S}^{n-1} : x_n < 0\}$. Then assume that $y \in \mathbb{S}^{n-1}_-$ and define the map, $p : \mathbb{S}^{n-1}_- \to H_{-1}(e_n)$ by

$$p(x_1, \ldots, x_n) = \left(\frac{x_1}{-x_n}, \ldots, \frac{x_{n-1}}{-x_n}, -1\right) = \frac{-1}{x_n}(x_1, \ldots, x_n).$$

Note that p is a bijection, as for $y' = (y_1, \ldots, y_{n-1})$, we have

$$p^{-1}(y_1, \ldots, y_{n-1}, -1) = \frac{-1}{\sqrt{1 + |y'|^2}}(y_1, \ldots, y_{n-1}, -1)$$

(i.e. $p^{-1}(z) = -z/|z|$ restricted to $H_{-1}(e_n)$). We now show that p^{-1} is Lipschitz. For $x, y \in H_{-1}(e_n)$,

$$|p^{-1}(x) - p^{-1}(y)| = \left|\frac{x}{|x|} - \frac{y}{|y|}\right| = \left|\frac{|y|x - |x|y}{|x||y|}\right| = \left|\frac{|y|x - |x|x + |x|x - |x|y}{|x||y|}\right|$$

As $x, y \in H_{-1}(e_n)$, this implies that $|x|, |y| \geq 1$ and by the triangle inequality we have,

$$\begin{aligned} |p^{-1}(x) - p^{-1}(y)| \leq \frac{|x|\big||y| - |x|\big| + |x||x - y|}{|x||y|} &\leq \frac{\big||y| - |x|\big|}{|y|} + \frac{|x - y|}{|y|} \\ &\leq \big||y| - |x|\big| + |x - y| \leq 2|x - y|. \end{aligned}$$

As $y \in N_d(K)$, and given that $y_n < 0$ implies that $\frac{-1}{y_n} > 0$, we have by Theorem 97 and Lemma 107, that $\partial\sigma_K(y) = \partial\sigma_K(p(y))$. Thus for any $z \in p(N_d(K) \cap \mathbb{S}^{n-1})$ there exists $z_0 \in N_d(K) \cap \mathbb{S}^{n-1}$ such that $p(z_0) = z$ and $\partial\sigma_K(p(z_0)) = \partial\sigma_K(z_0)$. Since $z_0 \in N_d(K)$, $\dim(\partial\sigma_K(z)) = \dim(\partial\sigma_K(z_0)) \geq d$ we have,

$$z \in \Sigma^d\left(\partial\sigma_K\big|_{H_{-1}(e_n)}\right) := \{x \in H_{-1}(e_n) : \dim(\partial\sigma_K(x)) \geq d\}.$$

This shows that,

$$p(N_d(K) \cap \mathbb{S}^{n-1}_-) \subset \Sigma^d\left(\partial\sigma_K\big|_{H_{-1}(e_n)}\right).$$

Given that $\sigma\big|_{H_{-1}(e_n)}$ is convex, Theorem 101 implies there exists countably many $g_i : \mathbb{R}^{n-d-1} \to H_{-1}(e_n)$ such that $g_i(\mathbb{R}^{n-d-1})$ is a $(c-c)_{d+1}$-graph, and

$$p(N_d(K) \cap \mathbb{S}^{n-1}_-) \subset \bigcup_{i=1}^{\infty} g_i(\mathbb{R}^{n-d-1}).$$

Therefore, we have

$$N_d(K) \cap \mathbb{S}^{n-1}_- \subset \bigcup_{i=1}^{\infty}(p^{-1} \circ g_i)(\mathbb{R}^{n-d-1}).$$

The result follows by repeating the above argument for the other $2n-1$ hemispheres of $\mathbb{S}^{n-1}$. **QED**

We say that a set $X \subset \mathbb{R}^n$ is *countably k-rectifiable* if there is a family of measurable sets $E_i \subset \mathbb{R}^k$ and Lipschitz functions $f_i : E_i \to \mathbb{R}^n$ such that

$$\mathcal{H}^k\left(X \setminus \bigcup_{i=1}^{\infty} f_i(E_i)\right) = 0.$$

Corollary 110. *The set $N_d(K)$ is countably $(n-d-1)$-rectifiable.*

Proof. We note that by Theorem 14, restricting any convex function $f : \mathbb{R}^n \to \mathbb{R}$ to a bounded subset is Lipschitz, and hence if $F : \mathbb{R}^{n-d} \to \mathbb{R}^d$ is a $(c-c)_d$-function, then F is Lipschitz when restricted to any bounded set. Therefore by Theorem 108, restricted to bounded sets, each h_i is Lipschitz and the result follows by noting

$$N_d(K) \subset \bigcup_{i=1}^{\infty} \left(\bigcup_{k=1}^{\infty} h_i(B^{n-d-1}(0,k)) \right).$$

QED

6.0 Equivalence of Continuously Differentiable Convex Functions with Hölder Continuous Gradient

6.1 Convex Functions with Lipschitz Continuous Gradient

There is a fascinating set of equivalent statements for $C^{1,1}$ convex functions that is difficult to find compiled together. We give here the statements and proofs so that they can be a resource to anyone in the future looking for a characterization of $C^{1,1}$ convex functions.

Theorem 111. *Let $f \in C^1(\mathbb{R}^n)$ be convex and let $L > 0$. Then for all $x, y \in \mathbb{R}^n$ the following are equivalent:*

(1) $|\nabla f(x) - \nabla f(y)| \leq L|x - y|$;

(2) $g(x) = \frac{L}{2}|x|^2 - f(x)$ *is convex;*

(3) $f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}|y - x|^2$;

(4) $\langle \nabla f(x) - \nabla f(y), x - y\rangle \leq L|x - y|^2$;

(5) $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2L}|\nabla f(x) - \nabla f(y)|^2$;

(6) $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{L}|\nabla f(x) - \nabla f(y)|^2$;

(7) $|\langle \nabla f(x) - \nabla f(y), x - y\rangle \leq L|x - y|^2$.

Remark 112. We will prove the following implications

$$(1) \Rightarrow (4) \Leftrightarrow (2) \Leftrightarrow (3) \Rightarrow (5) \Rightarrow (6) \Rightarrow (1) \Leftrightarrow (7).$$

The only implication that requires convexity of f is (3) $\Rightarrow$ (5). All other implications are true for any $f \in C^1(\mathbb{R}^n)$.

Proof. $\boxed{(1) \implies (4)}$

Let $|\nabla f(x) - \nabla f(y)| \leq L|x - y|$. Then by Cauchy-Schwarz,

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \leq |\nabla f(x) - \nabla f(y)||x - y| \leq L|x - y|^2$$

$\boxed{(2) \Leftrightarrow (4)}$

Note $g(x) = \frac{L}{2}|x|^2 - f(x)$ implies $\nabla g(x) = Lx - \nabla f(x)$ and thus by Theorem 22 we have the following equivalence:

$$\begin{aligned} g \text{ is convex} &\Leftrightarrow \langle \nabla g(x) - \nabla g(y), x - y\rangle \geq 0 \\ &\Leftrightarrow \langle (Lx - \nabla f(x)) - (Ly - \nabla f(y)), x - y\rangle \geq 0 \\ &\Leftrightarrow \langle L(x-y) - (\nabla f(x) - \nabla f(y)), x - y\rangle \geq 0 \\ &\Leftrightarrow \langle L(x-y), x-y\rangle - \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0 \\ &\Leftrightarrow L|x-y|^2 \geq \langle \nabla f(x) - \nabla f(y), x - y\rangle \end{aligned}$$

$\boxed{(2) \Leftrightarrow (3)}$

Similarly, we consider by Theorem 22

$$\begin{aligned} g \text{ is convex} &\Leftrightarrow g(x) + \langle \nabla g(x), y - x\rangle \leq g(y) \\ &\Leftrightarrow \frac{L}{2}|x|^2 - f(x) + \langle Lx - \nabla f(x), y - x\rangle \leq \frac{L}{2}|y|^2 - f(y) \\ &\Leftrightarrow f(y) \leq f(x) + \frac{L}{2}(|y|^2 - |x|^2) - \langle Lx - \nabla f(x), y - x\rangle \\ &\Leftrightarrow f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}(|y|^2 - |x|^2) - \langle Lx, y - x\rangle \\ &\Leftrightarrow f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}(|y|^2 - 2\langle x, y\rangle + |x|^2) \\ &\Leftrightarrow f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}|y-x|^2 \end{aligned}$$

$\boxed{(3) \implies (5)}$

As mentioned earlier to prove (3) $\Rightarrow$ (5) we will need the convexity of f. By (3) we have for any y and z,

$$f(z) \leq f(y) + \langle \nabla f(y), z - y\rangle + \frac{L}{2}|z-y|^2 \tag{81}$$

Fix $x \in \mathbb{R}^n$. By convexity of f and Corollary 22 we know for all z,

$$f(x) + \langle \nabla f(x), z - x\rangle \leq f(z)$$

which implies,

$$f(x) - \langle \nabla f(x), x\rangle \leq f(z) - \langle \nabla f(x), z\rangle.$$

Combining this inequality with (81) yields,

$$f(x) - \langle \nabla f(x), x \rangle \leq f(y) + \langle \nabla f(y), z - y \rangle + \frac{L}{2}|z - y|^2 - \langle \nabla f(x), z \rangle$$

and thus,

$$f(x) - \langle \nabla f(x), x - y \rangle \leq f(y) + \langle \nabla f(y), z - y \rangle + \frac{L}{2}|z - y|^2 - \langle \nabla f(x), z \rangle + \langle \nabla f(x), y \rangle.$$

This gives us,

$$f(x) - f(y) - \langle \nabla f(x), x - y \rangle \leq \langle \nabla f(y) - \nabla f(x), z - y \rangle + \frac{L}{2}|z - y|^2.$$

Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be defined by $\psi(z) = \langle \nabla f(y) - \nabla f(x), z - y \rangle + \frac{L}{2}|z - y|^2$. Thus we can rewrite the above equation as,

$$f(x) - f(y) - \langle \nabla f(x), x - y \rangle \leq \psi(z) \tag{82}$$

and note that as $|z| \to \infty$, $\psi(z) \to \infty$. Thus the function ψ attains a minimum at some point z_0, where $\nabla \psi(z_0) = 0$, i.e.

$$\nabla f(y) - \nabla(x) + L(z_0 - y) = 0 \tag{83}$$

implying that $z_0 = y - \frac{1}{L}(\nabla f(y) - \nabla(x))$. Substituting z_0 into (83), yields

$$\psi(z_0) = -\frac{1}{2L}|\nabla f(y) - \nabla(x)|^2 \tag{84}$$

and substituting (84) into (82) gives the desired result as,

$$f(x) - f(y) - \langle \nabla f(x), x - y \rangle \leq -\frac{1}{2L}|\nabla f(y) - \nabla(x)|^2$$

which implies,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}|\nabla f(y) - \nabla(x)|^2.$$

$\boxed{(5) \implies (6)}$

By (5),

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}|\nabla f(y) - \nabla(x)|^2$$

and by switching x and y we also have,

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2L}|\nabla f(x) - \nabla f(y)|^2.$$

Thus adding these two inequalities together yields,

$$f(x) + f(y) \geq f(y) + f(x) + \langle \nabla f(y), x - y \rangle - \langle \nabla f(x), x - y \rangle + \frac{1}{L}|\nabla f(x) - \nabla f(y)|^2$$

which can be simplified to,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L}|\nabla f(x) - \nabla f(y)|^2.$$

$\boxed{(6) \implies (1)}$

Consider,

$$\begin{aligned} \frac{1}{L}|\nabla f(x) - \nabla f(y)|^2 &\leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \\ &\leq |\nabla f(x) - \nabla f(y)||x - y| \end{aligned} \tag{85}$$

which implies, $|\nabla f(x) - \nabla f(y)| \leq L|x - y|$ as desired.

$\boxed{(1) \Leftrightarrow (7)}$

We first recall the definitions of convolution and mollifiers. Given $f, g \in L^1(\mathbb{R}^n)$, we define the convolution $f * g : \mathbb{R}^n \to \mathbb{R}$ as,

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x - y)\, dy.$$

Consider, a smooth function $\phi : \mathbb{R}^n \to \mathbb{R}$ with compact support such that

$$\int_{\mathbb{R}^n} \phi(x)\, dx = 1.$$

Moreover for every $\varepsilon > 0$ we define $\phi_\varepsilon(x) = \varepsilon^{-n}\phi(x/\varepsilon)$. We call ϕ_ε the standard mollifier. For a full description of the properties of the standard mollifier see [13, Theorem 4.1].

The implication (1) $\Rightarrow$ (7) follows immediately from the Cauchy-Schwarz inequality so we will prove (7) $\Rightarrow$ (1). We first assume that $f \in C^\infty(\mathbb{R}^n)$. Thus for $|u| = 1$ condition (1) yields,

$$\left| \left\langle \frac{\nabla f(x + tu) - \nabla f(x)}{t}, u \right\rangle \right| \leq \frac{|\nabla f(x + tu) - \nabla f(x)|}{|tu|} \leq L$$

for any $t \in \mathbb{R}$. Letting $t \to 0$ gives us, $|\langle D^2 f(x)u, u\rangle| \leq L$. Since $D^2 f(x)$ is a symmetric matrix the spectral theorem implies that the operator norm of the matrix $D^2 f(x)$ satisfies,

$$\|D^2 f(x)\| = \sup_{|u|=1} |\langle D^2 f(x)u, u\rangle| \leq L. \tag{86}$$

Using (86) allows us to prove the result for $f \in C^\infty(\mathbb{R}^n)$ as, by the Mean value theorem,

$$\begin{aligned}|\nabla f(x) - \nabla f(y)| &= \left| \int_0^1 \frac{d}{dt} \nabla f(y + t(x-y))\, dt \right| \\ &\leq |x-y| \int_0^1 \|D^2 f(y + t(x-y))\|\, dt \leq L|x-y|.\end{aligned}$$

Now let us assume that $f \in C^1(\mathbb{R}^n)$ and let $f_\varepsilon = f * \phi_\varepsilon$ be a standard approximation by convolution. Recall that $f_\varepsilon \in C^\infty(\mathbb{R}^n)$ and $\nabla f_\varepsilon = (\nabla f) * \phi_\varepsilon$. Thus using condition (1) and the fact that $\int_{\mathbb{R}^n} \phi_\varepsilon(z)\, dz = 1$ we have,

$$\begin{aligned}|\langle \nabla f_\varepsilon(x) - \nabla f_\varepsilon(y), x-y\rangle| &= \left| \left\langle \int_{\mathbb{R}^n} (\nabla f(x-z) - \nabla f(y-z))\phi_\varepsilon(z)\, dz, x-y \right\rangle \right| \\ &\leq \int_{\mathbb{R}^n} |\langle \nabla f(x-z) - \nabla f(y-z), (x-z)-(y-z)\rangle|\, \phi_\varepsilon(z)\, dz \\ &\leq L|x-y|^2\end{aligned}$$

Since $f_\varepsilon \in C^\infty(\mathbb{R}^n)$ and satisfies (1), by our previous result,

$$|\nabla f_\varepsilon(x) - \nabla f_\varepsilon(y)| \leq L|x-y|$$

and the equivalence (1) $\Leftrightarrow$ (7) follows by letting $\varepsilon \to 0^+$. **QED**

In the above proof the convexity of f was not required for the proof of (1) $\Rightarrow$ (4) but was required for the converse, (4) $\Rightarrow$ (1). The next result given in [5, Proposition 2.2] provides one more characterization of convex functions with a Lipschitz continuous gradient.

Theorem 113. *For a convex function $f : \mathbb{R}^n \to \mathbb{R}$ the following conditions are equivalent*

(a) *There is $L > 0$ such that for all $x, h \in \mathbb{R}^n$*

$$f(x+h) + f(x-h) - 2f(x) \leq L|h|^2 \tag{87}$$

(b) *$f \in C^1$ and $|\nabla f(x) - \nabla f(y)| \leq L|x-y|$ for all $x, y \in \mathbb{R}^n$*

Proof. First we will prove the implication $(b) \Rightarrow (a)$ which is true for all C^1 functions and does not require convexity. Lemma 41 implies that

$$f(x+h) - f(x) - \langle \nabla f(x), h \rangle \leq \frac{L}{2}|h|^2$$

By replacing h with $-h$ we then have the inequality

$$f(x-h) - f(x) + \langle \nabla f(x), h \rangle \leq \frac{L}{2}|h|^2.$$

Adding these two inequalities together yields the desired inequality

$$f(x+h) + f(x-h) - 2f(x) \leq L|h|.$$

The proof of $(a) \Rightarrow (b)$ requires convexity of f. Recall, by Theorem 21 that convex functions have one sided partial derivatives at every point, denoted by

$$\frac{\partial^{\pm} f}{\partial x_i}(x) = \lim_{t \to 0^{\pm}} \frac{f(x + te_i) - f(x)}{t}.$$

Using (a) we can see that

$$\lim_{h \to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{|h|} = 0$$

and thus letting $h = te_i$ yields,

$$\begin{aligned} 0 &= \lim_{t \to 0^+} \left(\frac{f(x+te_i) - f(x)}{t} - \frac{f(x-te_i) - f(x)}{-t} \right) \\ &= \lim_{t \to 0^+} \frac{f(x+te_i) - f(x)}{t} - \lim_{t \to 0^-} \frac{f(x+te_i) - f(x)}{t} = \frac{\partial^+ f}{\partial x_i}(x) - \frac{\partial^- f}{\partial x_i}(x) \end{aligned}$$

implying that $\frac{\partial^+ f}{\partial x_i}(x) = \frac{\partial^- f}{\partial x_i}(x)$, which proves that each partial derivative exists for all $x \in \mathbb{R}^n$. As f is convex, we have $f \in C^1$. Lastly we will show

$$g(x) = \frac{L}{2}|x|^2 - f(x)$$

is convex and by Theorem 111 this will complete the proof. As g is continuous, by Theorem 12, convexity of g is equivalent to

$$g\left(\frac{x+y}{2}\right) \leq \frac{g(x) + g(y)}{2} \tag{88}$$

for all $x, y \in \mathbb{R}^n$. Thus we have,

$$g\left(\frac{x+y}{2}\right) = \frac{g(x)+g(y)}{2} + \frac{1}{2}\left(f(x)+f(y)-2f\left(\frac{x+y}{2}\right) - L\left|\frac{x-y}{2}\right|^2\right)$$

and we can see (88) follows by replacing x with $\frac{x+y}{2}$ and h with $\frac{x-y}{2}$ in (87), which implies

$$f(x)+f(y)-2f\left(\frac{x+y}{2}\right) - L\left|\frac{x-y}{2}\right|^2 \leq 0.$$

Therefore g is convex, completing the proof. **QED**

6.2 Convex Functions with Hölder Continuous Gradient

There is a similar set of equivalent statements for $C^{1,\alpha}$ convex functions which have been collected and adapted from sources such as [12], [6], and [5]. Moreover more statements are likely able to be added to this list, though proofs are needed.

Theorem 114. *Let $f \in C^1(\mathbb{R}^n)$ and $x, y \in \mathbb{R}^n$, then the following are equivalent:*

(1) $f \in C^{1,\alpha}(\mathbb{R}^n)$;

(2) $f(y) \leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{1+\alpha}|y-x|^{1+\alpha}$ *for some* $L > 0$;

(3) $f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle + \frac{\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x) - \nabla f(y)|^{\frac{1+\alpha}{\alpha}}$ *for some* $L > 0$;

(4) $\langle \nabla f(x) - \nabla f(y), x-y\rangle \geq \frac{2\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x) - \nabla f(y)|^{\frac{1+\alpha}{\alpha}}$ *for some* $L > 0$.

Proof. $\boxed{(1) \implies (2)}$

This follows from Lemma 41.

$\boxed{(2) \implies (3)}$

Suppose to the contrary that there exists x_0, y_0 such that,

$$f(y_0) - f(x_0) - \langle \nabla f(x_0), y_0 - x_0\rangle < \frac{\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(y_0) - \nabla f(x_0)|^{\frac{1+\alpha}{\alpha}}. \tag{89}$$

Define the function $g : \mathbb{R}^n \to \mathbb{R}$ by,

$$g(x) = \frac{1}{L}\left(f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0\rangle\right)$$

and note that $g(x_0) = 0$ and $\nabla g(x) = \frac{1}{L}\left(\nabla f(x) - \nabla f(x_0)\right)$. Thus we can rewrite (89) as,

$$Lg(y_0) < \frac{L\alpha}{1+\alpha}|L\nabla g(y_0)|^{\frac{1+\alpha}{\alpha}}.$$

This then yields the following inequality,

$$g(y_0) < \frac{\alpha}{1+\alpha}|\nabla g(y_0)|^{\frac{1+\alpha}{\alpha}}. \tag{90}$$

Lemma 115. *$g : \mathbb{R}^n \to \mathbb{R}$, as defined above, also satisfies* (2) *with $L = 1$, that is,*

$$g(y) \leq g(z) + \langle \nabla g(z), y - z\rangle + \frac{1}{1+\alpha}|y-z|^{1+\alpha} \tag{91}$$

Proof of Lemma 115. By (2) we have,

$$f(y) \leq f(z) + \langle f(z), y - z\rangle + \frac{L}{1+\alpha}|y-z|^{1+\alpha}$$

which is equivalent to,

$$Lg(y) \leq f(z) - f(x_0) - \langle \nabla f(x_0), y - x_0\rangle + \langle \nabla f(z), y - z\rangle + \frac{L}{1+\alpha}|y-z|^{1+\alpha}.$$

As,

$$\begin{aligned}
&-\langle \nabla f(x_0), y - x_0\rangle + \langle \nabla f(z), y - z\rangle \\
&\qquad\qquad = -\langle \nabla f(x_0), y\rangle + \langle \nabla f(x_0), x_0\rangle + \langle \nabla f(z), y\rangle - \langle \nabla f(z), z\rangle \\
&\qquad\qquad = \langle L\nabla g(z), y\rangle + \langle \nabla f(x_0), x_0\rangle - \langle \nabla f(z), z\rangle
\end{aligned}$$

and

$$\begin{aligned}
\langle \nabla f(x_0), x_0\rangle - \langle \nabla f(z), z\rangle &= \langle \nabla f(x_0), x_0\rangle - \langle \nabla f(x_0), z\rangle + \langle \nabla f(x_0), z\rangle - \langle \nabla f(z), z\rangle \\
&= \langle \nabla f(x_0) - \nabla f(z), z\rangle + \langle \nabla f(x_0), x_0 - z\rangle. \\
&= \langle \nabla f(x_0), x_0 - z\rangle - \langle L\nabla g(z), z\rangle
\end{aligned}$$

we now have,

$$\begin{aligned}
Lg(y) &\leq f(z) - f(x_0) + \langle \nabla g(z), y\rangle + \langle \nabla f(x_0), x_0 - z\rangle - \langle \nabla g(z), z\rangle + \frac{L}{1+\alpha}|y-z|^{1+\alpha} \\
&= Lg(z) + \langle \nabla Lg(z), y - z\rangle + \frac{L}{1+\alpha}|y-z|^{1+\alpha}
\end{aligned}$$

proving the lemma. **QED**

Thus for every $t > 0$ and $|v| = 1$, we have by (91)

$$g(y_0 - tv) \le \frac{1}{1+\alpha}|tv|^{1+\alpha} + g(y_0) - \langle \nabla g(y_0), tv\rangle = \frac{t^{1+\alpha}}{1+\alpha} + g(y_0) - \langle \nabla g(y_0), tv\rangle.$$

Choosing $t = |\nabla g(y_0)|^{\frac{1}{\alpha}}$ and $v = \frac{-\nabla g(y_0)}{|\nabla g(y_0)|}$, yields,

$$\begin{aligned} g\left(y_0 + |\nabla g(y_0)|^{\frac{1-\alpha}{\alpha}} \nabla g(y_0)\right) &\le g(y_0) - |\nabla g(y_0)|^{\frac{1+\alpha}{\alpha}} + \frac{|\nabla g(y_0)|^{\frac{1+\alpha}{\alpha}}}{1+\alpha} \\ &= g(y_0) - \frac{\alpha}{1+\alpha}|\nabla g(y_0)|^{\frac{1+\alpha}{\alpha}} \\ &< 0 \quad \text{by (90)} \end{aligned}$$

But as g is convex, $g(x_0) = 0$ and $\nabla g(x_0) = 0$, we have for all $y \in \mathbb{R}^n$,

$$g(y) \ge g(x_0) + \langle \nabla g(x_0), y - x_0\rangle = 0$$

a contradiction. Thus for all $x, y \in \mathbb{R}^n$,

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(y) - \nabla f(x)|^{\frac{1+\alpha}{\alpha}}$$

(3) $\implies$ (4)

We have by (3),

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x) - \nabla f(y)|^{\frac{1+\alpha}{\alpha}}$$

and switching the roles of x and y, yields,

$$f(x) \ge f(y) + \langle \nabla f(y), x - y\rangle + \frac{\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x) - \nabla f(y)|^{\frac{1+\alpha}{\alpha}}.$$

Adding these two inequalities gives us,

$$f(y) + f(x) \ge f(y) + f(x) + \langle \nabla f(x), y - x\rangle - \langle \nabla f(y), y - x\rangle + \frac{2\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x) - \nabla f(y)|^{\frac{1+\alpha}{\alpha}}$$

which is equivalent to,

$$0 \ge \langle \nabla f(x) - \nabla f(y), y - x\rangle + \frac{2\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x) - \nabla f(y)|^{\frac{1+\alpha}{\alpha}}$$

and the result easily follows.

(4) $\implies$ (1)

Assuming (4),

$$\frac{2\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x)-\nabla f(y)|^{\frac{1+\alpha}{\alpha}} \leq \langle \nabla f(x)-\nabla f(y), x-y\rangle < |\nabla f(x)-\nabla f(y)||x-y|$$

which implies,

$$\frac{2\alpha}{L^{\frac{1}{\alpha}}(1+\alpha)}|\nabla f(x)-\nabla f(y)|^{\frac{1}{\alpha}} \leq |x-y|$$

or equivalently,

$$|\nabla f(x)-\nabla f(y)|^{\frac{1}{\alpha}} \leq \frac{L^{\frac{1}{\alpha}}(1+\alpha)}{2\alpha}|x-y|.$$

Thus,

$$|\nabla f(x)-\nabla f(y)| \leq \frac{L(1+\alpha)^{\frac{1}{\alpha}}}{(2\alpha)^{\frac{1}{\alpha}}}|x-y|^{\alpha}.$$

QED

We also now present a proof that in fact (1) and (2) are equivalent. The proof is a simplification of the one given in [12, Lemma 3.1].

$\boxed{(2) \implies (1)}$

Proof. Assuming that,

$$f(y) \leq f(x)+\nabla f(x)\cdot(y-x)+\frac{L}{1+\alpha}|y-x|^{1+\alpha}$$

implies,

$$0 \leq f(y)-f(x)-\nabla f(x)\cdot(y-x) \leq \frac{L}{1+\alpha}|y-x|^{1+\alpha}$$

where the left inequality follows from the convexity of f. Therefore,

$$|f(y)-f(x)-\nabla f(x)\cdot(y-x)| \leq \frac{L}{1+\alpha}|y-x|^{1+\alpha}$$

First we choose $z \in \mathbb{R}^n$ such that,

$$|z-y| = |y-x| \quad \text{and} \quad |(\nabla f(x)-\nabla f(y))\cdot(z-y)| = |z-y||\nabla f(y)-\nabla f(x)|.$$

In this case choosing

$$z = y+\frac{\nabla f(y)-\nabla f(x)}{|y-x|}$$

will suffice, though we will continue to write z. Thus,

$$\begin{aligned}
|y-x||\nabla f(y)-\nabla f(x)| &= |z-y||\nabla f(y)-\nabla f(x)| \\
&= |(\nabla f(x)-\nabla f(y))\cdot(z-y)| \\
&= |\nabla f(x)\cdot(z-y)-\nabla f(y)\cdot(z-y)| \\
&= |f(z)-f(y)-\nabla f(y)\cdot(z-y)-f(z)+f(x)+\nabla f(x)\cdot z \\
&\qquad -\nabla f(x)\cdot x+f(y)-f(x)-\nabla f(x)\cdot y+\nabla f(x)\cdot x| \\
&\le |f(z)-f(y)-\nabla f(y)\cdot(z-y)|+|f(z)-f(x)-\nabla f(x)\cdot(z-x)| \\
&\qquad +|f(y)-f(x)-\nabla f(x)\cdot(y-x)| \\
&\le \frac{L}{1+\alpha}|z-y|^{1+\alpha}+\frac{L}{1+\alpha}|z-x|^{1+\alpha}+\frac{L}{1+\alpha}|y-x|^{1+\alpha} \\
&= \frac{L}{1+\alpha}|y-x|^{1+\alpha}+\frac{L}{1+\alpha}|z-x|^{1+\alpha}+\frac{L}{1+\alpha}|y-x|^{1+\alpha} \\
&= \frac{2L}{1+\alpha}|y-x|^{1+\alpha}+\frac{L}{1+\alpha}|z-x|^{1+\alpha}
\end{aligned}$$

And we note that,

$$|z-x|\le|z-y|+|y-x|=2|y-x| \implies |z-x|^{1+\alpha}\le 2^{1+\alpha}|y-x|^{1+\alpha}$$

so that we have,

$$|y-x||\nabla f(y)-\nabla f(x)|\le\frac{2L}{1+\alpha}|y-x|^{1+\alpha}+\frac{L}{1+\alpha}|z-x|^{1+\alpha}\le\frac{2L+2^{1+\alpha}L}{1+\alpha}|y-x|^{1+\alpha}$$

which implies,

$$|\nabla f(y)-\nabla f(x)|\le\frac{2L+2^{1+\alpha}L}{1+\alpha}|y-x|^{\alpha}.$$

QED

The following statement is conjectured to fit into the above equivalence but at this time only one direction is known. We include it here regardless.

Theorem 116. *If* $f\in C^1(\mathbb{R}^n)$ *satisfies* $f(y)\le f(x)+\langle\nabla f(x),y-x\rangle+\frac{L}{1+\alpha}|y-x|^{1+\alpha}$, *then*

$$f(tx+(1-t)y)\ge tf(x)+(1-t)f(y)-\frac{2t(1-t)L}{1+\alpha}|y-x|^{1+\alpha}.$$

Proof. If $t = 0, 1$ then the result is obvious. Let $t \in (0, 1)$. For any $x, y, z \in \mathbb{R}^n$, by (2)

$$f(y) - \frac{L}{1+\alpha}|y - z|^{1+\alpha} \leq f(z) + \langle \nabla f(z), y - z\rangle$$

$$f(x) - \frac{L}{1+\alpha}|x - z|^{1+\alpha} \leq f(z) + \langle \nabla f(z), x - z\rangle.$$

Thus letting $z = tx + (1-t)y$ yields,

$$f(y) - \frac{Lt^{1+\alpha}}{1+\alpha}|y - x|^{1+\alpha} \leq f(z) + t\langle \nabla f(z), y - x\rangle$$

$$f(x) - \frac{L(1-t)^{1+\alpha}}{1+\alpha}|y - x|^{1+\alpha} \leq f(z) - (1-t)\langle \nabla f(z), y - x\rangle.$$

Therefore by dividing by t and $1 - t$, respectively, we have,

$$\frac{f(y)}{t} - \frac{Lt^{\alpha}}{1+\alpha}|y - x|^{1+\alpha} \leq \frac{f(z)}{t} + \langle \nabla f(z), y - x\rangle$$

$$\frac{f(x)}{(1-t)} - \frac{L(1-t)^{\alpha}}{1+\alpha}|y - x|^{1+\alpha} \leq \frac{f(z)}{1-t} - \langle \nabla f(z), y - x\rangle$$

and adding these two inequalities yields,

$$\frac{f(y)}{t} + \frac{f(x)}{(1-t)} - \frac{Lt^{\alpha}}{1+\alpha}|y - x|^{1+\alpha} - \frac{L(1-t)^{\alpha}}{1+\alpha}|y - x|^{1+\alpha} \leq \frac{f(z)}{t} + \frac{f(z)}{1-t}.$$

Multiplying through by $t(1-t)$,

$$(1-t)f(y) + tf(x) - \frac{Lt^{1+\alpha(1-t)}}{1+\alpha}|y - x|^{1+\alpha} - \frac{Lt(1-t)^{1+\alpha}}{|y - x|^{1+\alpha}} \leq (1-t)f(z) + tf(z)$$

or equivalently,

$$(1-t)f(y) + tf(x) - \frac{L}{1+\alpha}|y - x|^{1+\alpha}(t^{1+\alpha}(1-t) - t(1-t)^{1+\alpha}) \leq f(z).$$

We now claim that, $2t(1-t) \geq t^{1+\alpha}(1-t) - t(1-t)^{1+\alpha}$ and thus,

$$-\frac{2t(1-t)L}{1+\alpha}|y - x|^{1+\alpha} \leq -\frac{L}{1+\alpha}|y - x|^{1+\alpha}(t^{1+\alpha}(1-t) - t(1-t)^{1+\alpha})$$

proving the desired inequality, as we recall $z = tx + (1-t)y$,

$$(1-t)f(y) + tf(x) - \frac{2t(1-t)L}{1+\alpha}|y - x|^{1+\alpha} \leq f(tx + (1-t)y).$$

We now provide the proof of the claim:

$$2t(1-t) \geq t^{1+\alpha}(1-t) - t(1-t)^{1+\alpha}.$$

This inequality is equivalent to,

$$2 \geq t^{\alpha} - (1-t)^{\alpha}$$

and thus defining, $g(t) = t^{\alpha} - (1-t)^{\alpha}$, we have that

$$g'(t) = \alpha t^{\alpha-1} + \alpha(1-t)^{\alpha-1}$$

but as this quantity is positive for all $t \in [0, 1]$ and $\alpha \in (0, 1)$, then the absolute max must be attained at either $t = 0$ or $t = 1$. Therefore considering $g(0) = -1$ and $g(1) = 1$ we see that for all $t \in [0, 1]$,

$$-1 \leq g(t) \leq 1 < 2$$

QED

Bibliography

[1] A. D. Alexandroff. Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. *Leningrad State Univ. Annals [Uchenye Zapiski] Math. Ser.*, 6:3–35, 1939.

[2] Daniel Azagra. Global and fine approximation of convex functions. *Proc. Lond. Math. Soc. (3)*, 107(4):799–824, 2013.

[3] Daniel Azagra, Anthony Cappello, and Piotr Hajłasz. A geometric approach to second-order differentiability of convex functions. *Proc. Amer. Math. Soc. Ser. B*, 10:382–397, 2023.

[4] Daniel Azagra and Piotr Hajłasz. Lusin-type properties of convex functions and convex bodies. *J. Geom. Anal.*, 31(12):11685–11701, 2021.

[5] Daniel Azagra, E. Le Gruyer, and C. Mudarra. Explicit formulas for $C^{1,1}$ and $C^{1,\omega}_{\mathrm{conv}}$ extensions of 1-jets in Hilbert and superreflexive spaces. *J. Funct. Anal.*, 274(10):3003–3032, 2018.

[6] Daniel Azagra and Carlos Mudarra. Whitney extension theorems for convex functions of the classes C^1 and $C^{1,\omega}$. *Proc. Lond. Math. Soc. (3)*, 114(1):133–158, 2017.

[7] J. Benoist and J.-B. Hiriart-Urruty. What is the subdifferential of the closed convex hull of a function? *SIAM Journal on Mathematical Analysis*, 27(6):1661–1679, 1996.

[8] Gabriele Bianchi, Andrea Colesanti, and Carlo Pucci. On the second differentiability of convex surfaces. *Geom. Dedicata*, 60(1):39–48, 1996.

[9] Bogdan Bojarski, Piotr Hajłasz, and Paweł Strzelecki. Sard's theorem for mappings in Hölder and Sobolev spaces. *Manuscripta Math.*, 118(3):383–397, 2005.

[10] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[11] Bernard Dacorogna. *Direct methods in the calculus of variations*, volume 78 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2008.

[12] Guido De Philippis and Alessio Figalli. Optimal regularity of the convex envelope. *Trans. Amer. Math. Soc.*, 367(6):4407–4422, 2015.

[13] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition, 2015.

[14] Piotr Hajłasz. Analysis 1. https://sites.pitt.edu/ hajlasz/Notatki/Analysis

[15] Piotr Hajłasz. On an old theorem of Erdős about ambiguous locus. *Colloq. Math.*, 168(2):249–256, 2022.

[16] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Grundlehren Text Editions. Springer-Verlag, Berlin, 2001. Abridged version of *Convex analysis and minimization algorithms. I* [Springer, Berlin, 1993; MR1261420 (95m:90001)] and *II* [ibid.; MR1295240 (95m:90002)].

[17] Lars Hörmander. *Notions of convexity*, volume 127 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1994.

[18] Bernd Kirchheim and Jan Kristensen. Differentiability of convex envelopes. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(8):725–728, 2001.

[19] Christer O. Kiselman. Regularity classes for operations in convexity theory. *Kodai Math. J.*, 15(3):354–374, 1992.

[20] Steven G. Krantz and Harold R. Parks. On the vector sum of two convex sets in space. *Canad. J. Math.*, 43(2):347–355, 1991.

[21] K. Lucas. Submanifolds of dimension $n-1$ in $\mathbb{E}^n$ with normals satisfying a lipschitz condition. *Thesis (Ph.D.)–University of Kansas.*, Studies in Eigenvalue Problems, Technical Report 18, Dept of Mathematics, University of Kansas, 1957.

[22] P. McMullen. On the inner parallel body of a convex body. *Israel J. Math.*, 19:217–219, 1974.

[23] Boris S. Mordukhovich and Nguyen Mau Nam. *An easy path to convex analysis and applications*, volume 14 of *Synthesis Lectures on Mathematics and Statistics*. Morgan & Claypool Publishers, Williston, VT, 2014.

[24] Alec Norton. A critical set with nonnull image has large Hausdorff dimension. *Trans. Amer. Math. Soc.*, 296(1):367–376, 1986.

[25] B. G. Pachpatte. *Mathematical inequalities*, volume 67 of *North-Holland Mathematical Library*. Elsevier B. V., Amsterdam, 2005.

[26] A. Wayne Roberts and Dale E. Varberg. *Convex functions*, volume Vol. 57 of *Pure and Applied Mathematics*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1973.

[27] R. Tyrrell Rockafellar. *Convex analysis*, volume No. 28 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1970.

[28] Rolf Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, expanded edition, 2014.

[29] Luděk Zajíček. On the differentiation of convex functions in finite and infinite dimensional spaces. *Czechoslovak Math. J.*, 29(104)(3):340–348, 1979.