Generalized convexities and semismoothness of nonsmooth functions

Ville Rinne

Licentiate Thesis
February 2023

Reviewed by:
Yury Nikulin
Antonio Fuduli

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF TURKU

Contents

1 Introduction

Convexity is a basic concept of geometry, but it is also used in other areas of mathematics. It is used, for example, in optimization, functional analysis, complex analysis, graph theory, partial differential equations, discrete mathematics, algebral mathematics, probability theory and coding theory. Convexity is also used outside mathematics, like in biology, physics and chemistry.

The first definition of convexity was given by Archimedes: If any two points on a bent line are taken, then either all the straight lines connecting the points fall on the same side of the line, or some fall on one and the same side while others fall on the line itself, but none on the other side.[3]

This thesis discusses the generalizations of convexity for nonsmooth functions, in other words functions that are not continuously differentiable. First we define some basic concepts of set theory. Then we define a convex set and a convex function and discuss their properties. To define generalized convexity for nonsmooth functions we need Clarke's generalized directional derivative and generalized subgradient. We define generalized pseudo- and quasiconvexity for nonsmooth locally Lipschitz continuous functions. A smooth convex function reaches its global minimum at points, where the function gradient is zero. This property extends for pseudoconvex and generalized pseudoconvex functions, but not for quasiconvex or generalized quasiconvex functions. Finally we define well-behaved semismooth and weakly semismooth functions and discuss their relations with generalized convex functions.

In mathematical research it is often required to make several, possibly overlapping, assumptions because the relations between different concepts are not known. This may severely limit the generality of the results. The purpose of this thesis is to find out if there are any causal relations between these concepts, which would make some of the assumptions made in literature redundant.

2 Definitions and results

This thesis discusses the Euclidean space R^n, whose elements are column-vectors $\mathbf{x}$. Let us give some basic definitions.

Definition 2.1. The *norm* of a vector $\mathbf{x} \in R^n$ is $\|\mathbf{x}\| = \sqrt{x_1^2 + ... + x_n^2}$. The space of $m \times n$ matrices is endowed with the norm

$$||A||_{m\times n} = \left(\sum_{i=1}^{m} ||A_i||^2\right)^{\frac{1}{2}},$$

where $A_i \in R^n$ is the i:th row of matrix A.

Definition 2.2. The *inner product* of vectors $\mathbf{x}$ and $\mathbf{y} \in R^n$ is $\mathbf{x}^T\mathbf{y} = x_1y_1 + ... + x_ny_n$.

Definition 2.3. The *spherical neighbourhood* $B(\mathbf{x},\ \delta)$ of an element $\mathbf{x}$ is a set which consists of center $\mathbf{x}$ and all elements within distance $\delta > 0$ of the centre, i.e. $B(\mathbf{x},\ \delta) = \{\mathbf{y} \in R^n \mid ||\mathbf{x} - \mathbf{y}|| < \delta\}$. Elements with a spherical neighbourhood within a set S are called *interior elements.* Interior elements of the complement space of the set S are called *exterior elements* of the set.

Definition 2.4. The *interior* $\text{int}(S)$ of a set S is the set of its interior elements. The *closure* $\text{cl}(S)$ of a set S consists of elements $\mathbf{x}$ which have a spherical neighbourhood $B(\mathbf{x},\ \delta)$ such that $\text{int}(S) \cap B(\mathbf{x},\ \delta) \neq \varnothing,\ \forall \delta > 0$.

Definition 2.5. The *boundary* $\delta(S)$ of a set S is the set of elements that are neither in the interior of the set nor the interior of its complement.

Definition 2.6. A set S is *open* if for every element $\mathbf{x}$ in the set can be given a radius δ such that every element in the spherical neighbourhood belongs to the set.

Definition 2.7. A set S is *closed* if its complement space in R^n is open.

Definition 2.8. A set S is *bounded* if there exists a positive real number r such that $||\mathbf{x} - \mathbf{y}|| < r,\ \forall\ \mathbf{x},\ \mathbf{y} \in S$.

Definition 2.9. A subset S of R^n is *compact* if it is closed and bounded.

Definition 2.10. Function $f : S \to R$ is *continuous* if for every element $\mathbf{x}$ and $\mathbf{y}$ of set S and for every positive real number ϵ there exists a positive real number δ such that from $||\mathbf{x} - \mathbf{y}|| < \delta$ it follows $|f(\mathbf{x}) - f(\mathbf{y})| < \epsilon$.

A continuous function is a function that does not have abrupt changes in its value. A real-valued function is continuous if, roughly speaking, its graph is a single unbroken graph [12]. A rigorous definition of continuity of real functions can be given in terms of the idea of a limit. A function f is said to be continuous at the point c on the real line, if the limit of $f(x)$, as x approaches the point c, is equal to the value $f(c)$. A function is said to be continuous if it is continuous at every point in its domain. A function is said to have a discontinuity at a point c when it is not continuous there. For example, the function $f(x) = \frac{1}{x}$ is continuous on the domain $R \backslash \{0\}$, but it is not continuous over the domain R, because it is undefined at $x = 0$ (Figure 1).

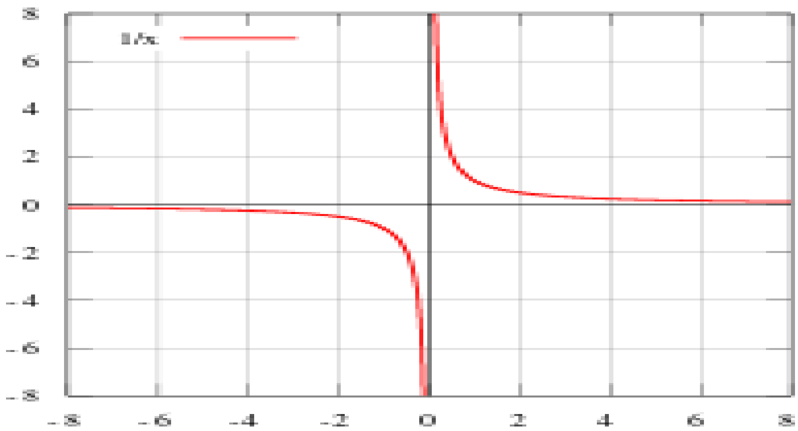


Figure 1: Function $1/x$ is not continuous

Definition 2.11. Function $f : S \to R$ is *upper bounded* if there exists a real number M such that $f(\mathbf{x}) < M$, $\forall\ \mathbf{x} \in S$ (see Figure 2).

Definition 2.12. Function $f : S \to R$ is *lower bounded* if there exists a real number M such that $f(\mathbf{x}) > M$, $\forall\ \mathbf{x} \in S$ (see Figure 3).

Definition 2.13. Function $f : S \to R$ is *upper semi-continuous* at $\mathbf{x}$ if for every sequence $\{\mathbf{x}_t\} \in S$ which approaches element $\mathbf{x}$ we have $\limsup f(\mathbf{x}_t) \leq f(\mathbf{x})$.

Definition 2.14. Function $f : S \to R$ is *lower semi-continuous* at $\mathbf{x}$ if for every sequence $\{\mathbf{x}_t\} \in S$ which approaches element $\mathbf{x}$ we have $\liminf f(\mathbf{x}_t) \geq f(\mathbf{x})$. A continuous function is both upper and lower semi-continuous.

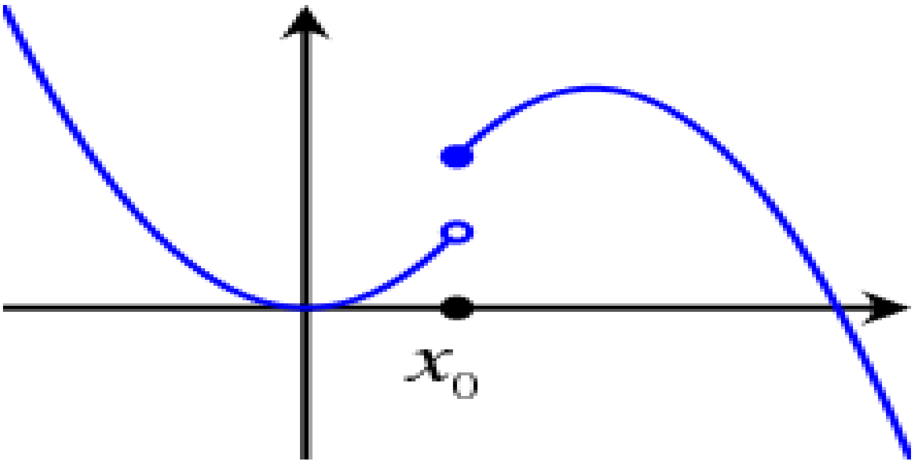


Figure 2: An upper semi-continuous function

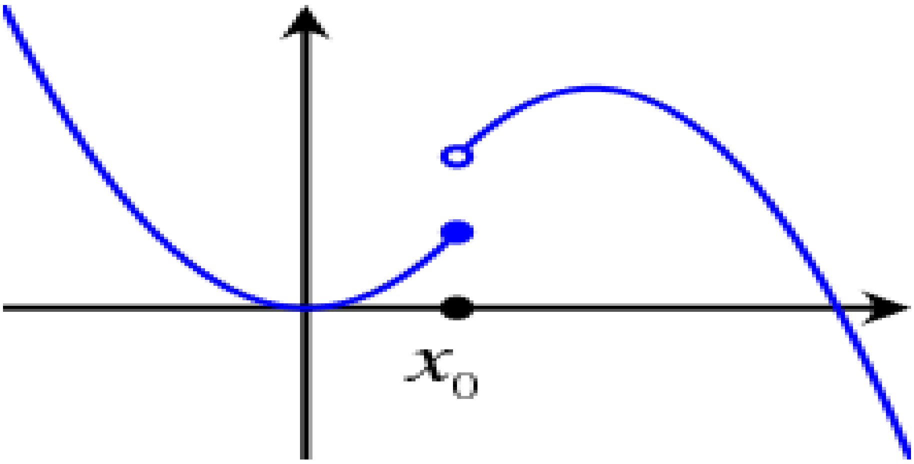


Figure 3: A lower semi-continuous function

Definition 2.15. A tangent line to a plane curve at a given point is the straight line that just touches the curve at that point.

Definition 2.16. [1] Let function f be defined in an open subset S of R^n. The function is *differentiable* at $\mathbf{x}^* \in S$, if there exists a vector $\nabla f(\mathbf{x}^*) \in R^n$ such that for every element $\mathbf{d} \in R^n$ satisfying $\mathbf{x}^* + \mathbf{d} \in S$

$$f(\mathbf{x}^* + \mathbf{d}) = f(\mathbf{x}^*) + \mathbf{d}^T \nabla f(\mathbf{x}^*) + \alpha(\mathbf{x}^*,\ \mathbf{d})\|\mathbf{d}\|,$$

where α is a real function such that $\lim_{\mathbf{d}\to 0} \alpha(\mathbf{x}^*,\ \mathbf{d})\|\mathbf{d}\| = 0$. If function f is differentiable at every $\mathbf{x}^* \in S$, it is said to be differentiable on set S.

A differentiable function of one real variable is a function whose derivative exists at every point of its domain. As a result, the graph of a differentiable function has a non-vertical tangent line at each interior point in its domain and it cannot contain any breaks or angles. More generally, if $\mathbf{x}$ is an interior point in the domain of a function f, then f is said to be differentiable at $\mathbf{x}$ if the function derivative exists at $\mathbf{x}$. That means that the function curve has a non-vertical tangent line at the point $(\mathbf{x}, f(\mathbf{x}))$.

If f is differentiable at a point $\mathbf{x}$, then f must also be continuous at $\mathbf{x}$. In particular, any differentiable function is continuous at every point in its domain. The converse does not hold: a continuous function does not need to be differentiable. An example of such function is $f(x) = |x|$, which contains an angle at $x = 0$.

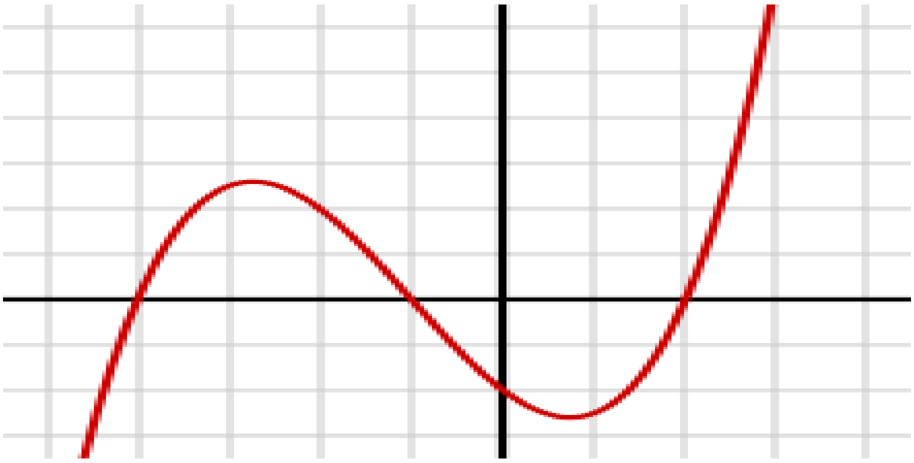


Figure 4: A differentiable function

Definition 2.17. Function $f : S \to R$ is *positively homogenous* if $f(\lambda \mathbf{x}) = \lambda f(\mathbf{x}),\ \forall\ \mathbf{x} \in S,\ \lambda > 0$.

Definition 2.18. Function $f : S \to R$ is *subadditive* if $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y}),\ \forall\ \mathbf{x},\ \mathbf{y} \in S$.

Next we define a convex set and give some examples.

Definition 2.19. A nonempty set S is said to be *convex* if the line segment of two elements $\mathbf{x}_1$ and $\mathbf{x}_2$ of set S also belong to the set, in other words

$$\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2 \in S,\ \forall\ \mathbf{x}_1,\ \mathbf{x}_2 \in S, \lambda \in [0, 1].$$

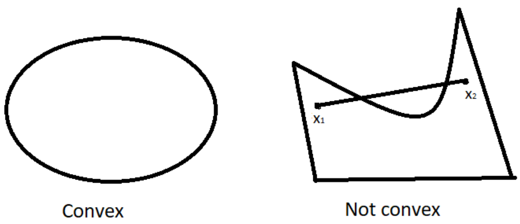


Figure 5: Illustration of convex and nonconvex sets

Here are some examples of convex sets:

1. Empty set.
2. Hyperplane $S = \{\mathbf{x} \mid \mathbf{p}^T\mathbf{x} = \alpha,\ \mathbf{p} \in R^n,\ \alpha \in R\}$.
3. Half space $S = \{\mathbf{x} \mid \mathbf{p}^T\mathbf{x} \geq \alpha,\ \mathbf{p} \in R^n,\ \alpha \in R\}$.
4. Convex polygon $S = \{\mathbf{x} \mid A\mathbf{x} \leq \mathbf{b},\ A \in R^{m\times n},\ \mathbf{b} \in R^m\}$.

Lemma 2.20. [2] Let S_1 and S_2 be convex sets in R^n. Then the sets

1. $S_1 \cap S_2$
2. $S_1 + S_2 = \{\mathbf{x}_1 + \mathbf{x}_2 \mid \mathbf{x}_1 \in S_1,\ \mathbf{x}_2 \in S_2\}$
3. $S_1 - S_2 = \{\mathbf{x}_1 - \mathbf{x}_2 \mid \mathbf{x}_1 \in S_1,\ \mathbf{x}_2 \in S_2\}$

are also convex.

Proof. 1. Let $\mathbf{x}_1$ and $\mathbf{x}_2 \in S_1 \cap S_2$. The line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$ is $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. Because $\mathbf{x}_1$ and $\mathbf{x}_2$ are elements of S_1, by the convexity of the set their line segment belongs to S_1. Because $\mathbf{x}_1$ and $\mathbf{x}_2$ are elements of S_2, their line segment also belongs to S_2. Thus it belongs to $S_1 \cap S_2$.

2. Let $\mathbf{x}$ and $\mathbf{y}$ be two elements of set S_1+S_2. Then they can be presented in form $\mathbf{x} = \mathbf{x}_1+\mathbf{x}_2,\ \mathbf{x}_1 \in S_1,\ \mathbf{x}_2 \in S_2$ and $\mathbf{y} = \mathbf{y}_1+\mathbf{y}_2,\ \mathbf{y}_1 \in S_1,\ \mathbf{y}_2 \in S_2$. The line segment of $\mathbf{x}$ and $\mathbf{y}$ is

$$\lambda(\mathbf{x}_1 + \mathbf{x}_2) + (1-\lambda)(\mathbf{y}_1 + \mathbf{y}_2) = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{y}_1 + \lambda\mathbf{x}_2 + (1-\lambda)\mathbf{y}_2.$$

   By the convexity of S_1 and S_2 $\lambda\mathbf{x}_1+(1-\lambda)\mathbf{y}_1 \in S_1$ and $\lambda\mathbf{x}_2+(1-\lambda)\mathbf{y}_2 \in S_2$. Therefore $S_1 + S_2$ is convex.

3. Proving the convexity of $S_1 - S_2$ is analogical.

□

Definition 2.21. [2] The *convex hull* $\text{conv}(S)$ of set S is the collection of all convex combinations of S. In other words $\mathbf{x} \in \text{conv}(S)$ if and only if $\mathbf{x}$ can be represented as

$$\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j$$
$$\sum_{j=1}^{k} \lambda_j = 1$$
$$\lambda_j \geq 0,\ \text{for } j = 1, ..., k,$$

where k is a positive integer and $\mathbf{x}_1, ..., \mathbf{x}_2 \in S$. $\text{conv}(S)$ is the minimal convex set that contains S. It is also the intersection of all convex sets containing S.

Definition 2.22. An *r-simplex* of set $P \subseteq R^n$ is a convex hull of its $r+1$ vertices, where $r \leq n$. More formally, if the vectors $\mathbf{x}_0 - \mathbf{x}_1, ..., \mathbf{x}_0 - \mathbf{x}_r$ are linearly independent, the simplex determined by the vertices is the set

$$C = \left\{ \lambda_0\mathbf{x}_0 + ... + \lambda_r\mathbf{x}_r \;\middle|\; \sum_{i=0}^{r} \lambda_i = 1,\ \lambda_i \geq 0\ \forall i = 0, ...r \right\}.$$

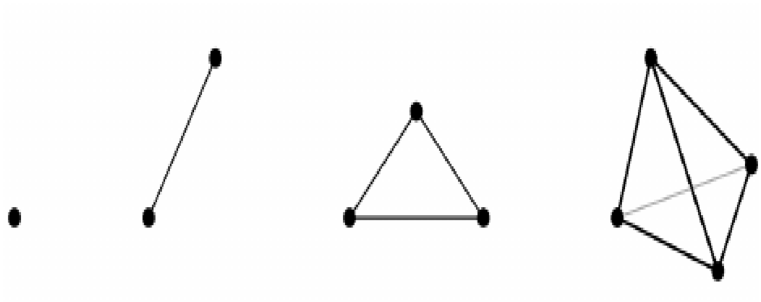


Figure 6: Examples of 0-, 1-, 2- and 3-simplexes

Caratheodory's theorem 2.23. [2,7] Let S be a compact set in R^n. If $\mathbf{x} \in \text{conv}(S)$, then $x \in \text{conv}(\mathbf{x}_1, ..., \ \mathbf{x}_{n+1})$, where $\mathbf{x}_j \in S$ for $j = 1, ..., n+1$. In other words, $\mathbf{x}$ can be represented as

$$\mathbf{x} = \sum_{j=1}^{n+1} \lambda_j \mathbf{x}_j$$

$$\sum_{j=1}^{n+1} \lambda_j = 1$$

$$\lambda_j \geq 0, \ j = 1, ..., n+1$$

$$\mathbf{x}_j \in S, \ j = 1, ..., n+1.$$

Proof. Since $\mathbf{x} \in \text{conv}(S)$, then $\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j$, where $\lambda_j > 0, \ j = 1, ..., k$ and $\sum_{j=1}^{k} = 1$. If $k \leq n+1$, the result is at hand. Now suppose that $k > n+1$. Note that $\mathbf{x}_2 - \mathbf{x}_1, \ \mathbf{x}_3 - \mathbf{x}_1, ..., \mathbf{x}_k - \mathbf{x}_1$ are linearly dependent. Thus there exist scalars $\mu_2, \ \mu_3, ..., \mu_k$ not all zero such that $\sum_{j=2}^{k} \mu_j(\mathbf{x}_j - \mathbf{x}_1) = \mathbf{0}$. Letting $\mu_1 = -\sum_{j=2}^{k} \mu_j$, it follows that $\sum_{j=1}^{k} \mu_j \mathbf{x}_j = \mathbf{0}$, $\sum_{j=1}^{k} \mu_j = 0$, and not all the μ_j's are equal to zero. Note that at least one $\mu_j > 0$. Then

$$\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j + \mathbf{0} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j - \alpha \sum_{j=1}^{k} \mu_j \mathbf{x}_j = \sum_{j=1}^{k} (\lambda_j - \alpha \mu_j) \mathbf{x}_j$$

for any real α. Now choose α as follows:

$$\alpha = \min_{1 \leq j \leq k} \left\{ \frac{\lambda_j}{\mu_j} : \mu_j > 0 \right\} = \frac{\lambda_i}{\mu_i}, \text{ where index } i \text{ is the argmin.}$$

Note that $\alpha > 0$. If $\mu_j \leq 0$, then $\lambda_j - \alpha\mu_j > 0$, and if $\mu_j > 0$, then $\lambda_j/\mu_j \geq \lambda_i/\mu_i = \alpha$, and hence $\lambda_j - \alpha\mu_j \geq 0$. In other words, $\lambda_j - \alpha\mu_j \geq 0, \ \forall\ j = 1, ..., k$. In particular, $\lambda_i - \alpha\mu_i = 0$ by definition of α. Therefore, $\mathbf{x} = \sum_{j=1}^{k}(\lambda_j - \alpha\mu_j)\mathbf{x}_j$, where $\lambda_j - \alpha\mu_j \geq 0,\ j = 1, ..., k,\ \sum_{j=1}^{k}(\lambda_j - \alpha\mu_j) = 1$, and furthermore, $\lambda_i - \alpha\mu_i = 0$. In other words, $\mathbf{x}$ is represented as a convex combination of at most $k - 1$ points in S. The process is repeated until $\mathbf{x}$ is represented as a convex combination of $n + 1$ points in S. □

Carathéodory's Theorem states that if a point $\mathbf{x}$ of R^n lies in the convex hull of a set P, then $\mathbf{x}$ can be written as the convex combination of at most $n + 1$ points in P. Namely, there is a subset P' of P consisting of $n + 1$ or fewer points such that $\mathbf{x}$ lies in the convex hull of P'. Equivalently, $\mathbf{x}$ lies in an r-simplex with vertices in P, where $r \leq n$. The smallest r that makes the last statement valid for each $\mathbf{x}$ in the convex hull of P is defined as the *Carathéodory's number* of P. Depending on the properties of P, upper bounds lower than the one provided by Carathéodory's Theorem can be obtained.

Definition 2.24. A *cone* is a shape that tapers smoothly from a flat base to a point called the apex. It is formed by a set of lines connecting a common point, the apex, to all of the points on a base in a plane that does not contain the apex. When the cone extends infinitely in both directions of the apex, it is called a *double cone*.

3 Convex functions

In this chapter we define convex functions and discuss their properties.

3.1 Definition and properties

Definition 3.1. [1] Let function $f : S \to R$ be defined on a nonempty convex set $S \subset R^n$. Function f is *convex* if the line segment from $f(\mathbf{x})$ to $f(\mathbf{y})$ of is higher than the function curve between $\mathbf{x}$ and $\mathbf{y}$, or

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}),\ \forall\ \mathbf{x},\ \mathbf{y} \in S,\ \lambda \in (0,\ 1).$$

Function f is *strictly convex* if the above inequality is strict.

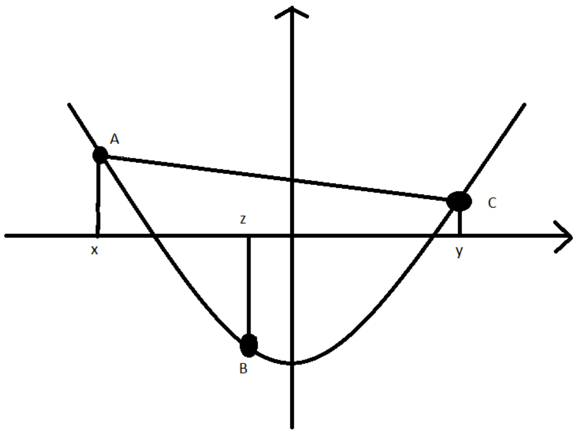


Figure 7: Convex function

The definition of convex functions can be illustrated in the case of a single variable function. The left-hand side of the inequality is the expression for the values of f between the points x and y, while the right-hand side represents the values of the line segment of $f(x)$ and $f(y)$. Equivalently, if A, B C are any three points on the graph of f such that B is between A and C, then B is on or below the line segment AC. Let $A = f(x),\ B = f(z)$ and $C = f(y)$. It is easy to verify the following relationship for a convex function that holds for $x \leq z \leq y$:

$$\frac{f(z)-f(x)}{z-x} \geq \frac{f(y)-f(x)}{y-x} \geq \frac{f(y)-f(z)}{y-z}.$$

A convex function and a convex set are related, because a convex function is defined on a convex set. Next we will show another relationship between a convex function and set, but first we will give a necessary definition of an epigraph.

Definition 3.2. [2] Let S be a nonempty set in R^n and let $f : S \to R$. The *epigraph* epi$f$ of f is a subset of R^{n+1} defined by

$$\{(\mathbf{x},\ y) : \mathbf{x} \in S,\ y \in R,\ y \geq f(\mathbf{x})\}.$$

The epigraph of f is the set of elements on and above the graph of f.

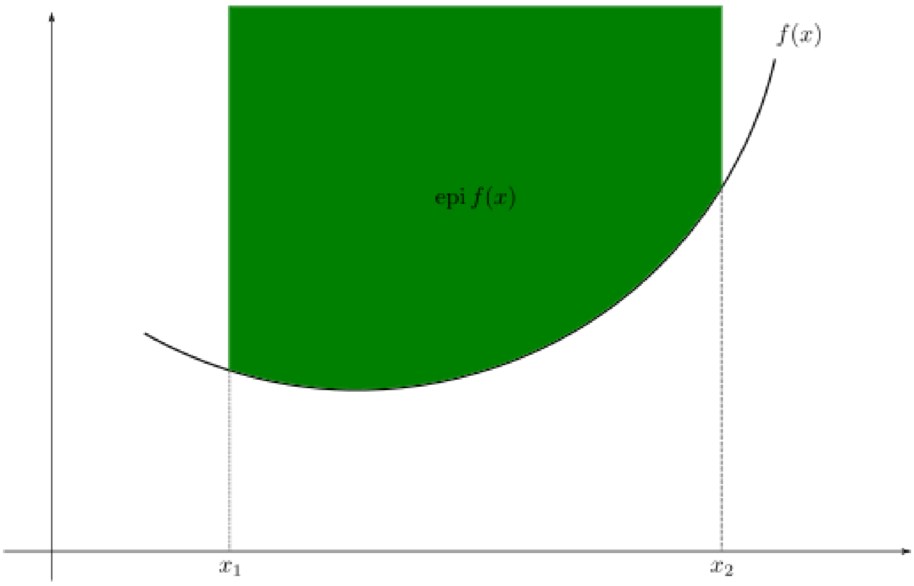


Figure 8: Epigraph of a convex function

Theorem 3.3. [2] Let S be a nonempty convex set in R^n and let $f : S \to R$. Then f is convex if and only if epi$f$ is a convex set.

Proof. Assume that f is convex, and let $(\mathbf{x}_1, y_1)$ and $(\mathbf{x}_2, y_2) \in$ epi$f$; that is, $\mathbf{x}_1,\ \mathbf{x}_2 \in S,\ y_1 \geq f(\mathbf{x}_1),\ y_2 \geq f(\mathbf{x}_2)$. Let $\lambda \in (0,\ 1)$. Then

$$\lambda y_1 + (1-\lambda)y_2 \geq \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) \geq f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2),$$

where the last inequality follows by convexity of f. Note that $\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2 \in S$. Thus $[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2,\ \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)] \in \text{epi}f$, so epi$f$ is convex. Conversely, assume that epi$f$ is convex, and let $\mathbf{x}_1,\ \mathbf{x}_2 \in S$. Then $(\mathbf{x}_1,\ f(\mathbf{x}_1))$ and $(\mathbf{x}_2,\ f(\mathbf{x}_2)) \in \text{epi}f$, and by convexity of epi$f$,

$$[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2,\ \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)] \in \text{epi}f,\ \lambda \in (0,\ 1).$$

In other words, $\lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) \geq f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2),\ \lambda \in (0,\ 1)$. Therefore f is convex. □

Definition 3.4. Let S be a nonempty convex set in R^n and let $f : S \to R$ be a convex function. The *level set* of f is

$$S_\alpha = \{\mathbf{x} \in S \mid f(\mathbf{x}) \leq \alpha\},$$

where $\alpha \in R$. The level set of f is the projection of epi$f$ onto S.

Theorem 3.5. Let S be a nonempty convex set in R^n and let $f : S \to R$ be a convex function. Then the level set S_α is convex.

Proof. Let $\mathbf{x},\ \mathbf{y} \in S_\alpha$. Thus $\mathbf{x},\ \mathbf{y} \in S$ and $f(\mathbf{x}) \leq \alpha$ and $f(\mathbf{y}) \leq \alpha$. Now let $\lambda \in (0,\ 1)$ and $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$. By convexity of S, $\mathbf{z} \in S$. Furthermore, by convexity of f,

$$f(\mathbf{z}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

Hence $\mathbf{x} \in S_\alpha$, and therefore S_α is convex. □

Definition 3.6. Function f defined on a convex set S is a *closed convex function*, if its epigraph is a closed convex set.

Theorem 3.7. [2] Let S be a nonempty convex set in R^n and let $f : S \to R$ be convex. Then f is continuous on the interior of S.

Proof. Let $\overline{\mathbf{x}} \in \text{int}(S)$. Let us prove that given $\epsilon > 0$, there exists a $\delta > 0$ such that $||\mathbf{x} - \overline{\mathbf{x}}|| \leq \delta$ implies that $|f(\mathbf{x}) - f(\overline{\mathbf{x}})| \leq \epsilon$. Since $\overline{\mathbf{x}} \in \text{int}(S)$, there exists a $\delta^{'} > 0$ such that $||\mathbf{x} - \overline{\mathbf{x}}|| \leq \delta^{'}$ implies that $\mathbf{x} \in S$. Construct θ as follows:

$$\theta = \max_{1 \leq i \leq n} \{\max[f(\overline{\mathbf{x}} + \delta^{'}\mathbf{e}_i) - f(\overline{\mathbf{x}}),\ f(\overline{\mathbf{x}} - \delta^{'}\mathbf{e}_i) - f(\overline{\mathbf{x}})]\}, \tag{1}$$

where $\mathbf{e}_i$ is a vector of zeroes except for one element at the ith position. By the convexity of f, $0 \leq \theta < \infty$. Let

$$\delta = \min\left(\frac{\delta^{'}}{n}, \frac{\epsilon\delta^{'}}{n\theta}\right). \tag{2}$$

Choose $\mathbf{x}$ with $\|\mathbf{x} - \overline{\mathbf{x}}\| \leq \delta$. If $x_i - \overline{x}_i \geq 0$, let $\mathbf{z}_i = \delta^{'}\mathbf{e}_i$; otherwise let $\mathbf{z}_i = -\delta^{'}\mathbf{e}_i$. Then $\mathbf{x} - \overline{\mathbf{x}} = \sum_{i=i}^{n} \alpha_i \mathbf{z}_i$, where $\alpha_i \geq 0$ for $i = 1,\ 2, ..., n$. Furthermore

$$\|\mathbf{x} - \overline{\mathbf{x}}\| = \delta^{'} \left(\sum_{i=1}^{n} \alpha_i^2\right)^{\frac{1}{2}} \tag{3}$$

.

From (2) and since $\|\mathbf{x} - \overline{\mathbf{x}}\| \leq \delta$, it follows that $\alpha_i \leq 1/n$ for $i = 1,\ 2, ..., n$. Hence, by the convexity of f and since $0 \leq n\alpha_i \leq 1$, we get

$$\begin{aligned}
f(\mathbf{x}) = f\left(\overline{\mathbf{x}} + \sum_{i=1}^{n} \alpha_i \mathbf{z}_i\right) &= f\left[\frac{1}{n}\sum_{i=1}^{n} (\overline{\mathbf{x}} + n\alpha_i \mathbf{z}_i)\right] \\
&\leq \frac{1}{n}\sum_{i=1}^{n} f(\overline{\mathbf{x}} + n\alpha_i \mathbf{z}_i) \\
&\leq \frac{1}{n}\sum_{i=1}^{n} f[(1 - n\alpha_i)\overline{\mathbf{x}} + n\alpha_i(\overline{\mathbf{x}} + \mathbf{z}_i)] \\
&\leq \frac{1}{n}\sum_{i=1}^{n} [(1 - n\alpha_1) f(\overline{\mathbf{x}}) + n\alpha_i f(\overline{\mathbf{x}} + \mathbf{z}_i)].
\end{aligned}$$

Therefore $f(\mathbf{x}) - f(\overline{\mathbf{x}}) \leq \sum_{i=1}^{n} \alpha_i [f(\overline{\mathbf{x}} + \mathbf{z}_i) - f(\overline{\mathbf{x}})]$. By (1) $f(\overline{\mathbf{x}} + \mathbf{z}_i) - f(\overline{\mathbf{x}}) \leq \theta,\ i = 1, ..., n$, and since $\alpha_i \geq 0$, it follows that

$$f(\mathbf{x}) - f(\overline{\mathbf{x}}) \leq \theta \sum_{i=1}^{n} \alpha_i. \tag{4}$$

From (3) and (2) it follows that $\alpha_i \leq \epsilon/n\theta$, and (4) implies that $f(\mathbf{x}) - f(\overline{\mathbf{x}}) \leq \epsilon$. So far we have showed that $\|\mathbf{x} - \overline{\mathbf{x}}\| \leq \delta$ implies that $f(\mathbf{x}) - f(\overline{\mathbf{x}}) \leq \epsilon$. To finish the proof let us show that $f(\overline{\mathbf{x}}) - f(\mathbf{x}) \leq \epsilon$. Let $\mathbf{y} = 2\overline{\mathbf{x}} - \mathbf{x}$ and note that $\|\mathbf{y} - \overline{\mathbf{x}}\| \leq \delta$. Therefore

$$f(\mathbf{y}) - f(\overline{\mathbf{x}}) \leq \epsilon, \tag{5}$$

But $\overline{\mathbf{x}} = \frac{1}{2}\mathbf{y} + \frac{1}{2}\mathbf{x}$, and by the convexity of f, we have

$$f(\overline{\mathbf{x}}) \leq \frac{1}{2} f(\mathbf{y}) + \frac{1}{2} f(\mathbf{x}). \tag{6}$$

Combining (5) and (6), it follows that $f(\overline{\mathbf{x}}) - f(\mathbf{x}) \leq \epsilon$. □

Next we will prove a stronger result concerning the local Lipschitz continuity of a function. First let us define local Lipschitz continuity and prove the necessary lemma.

Definition 3.8. [1] A function is *locally Lipschitz continuous at a point* $\mathbf{x} \in R^n$ if there exist scalars $K > 0$ and $\delta > 0$ such that

$$|f(\mathbf{y}) - f(\mathbf{z})| \leq K \|\mathbf{y} - \mathbf{z}\|, \ \forall\, \mathbf{y},\ \mathbf{z} \in B(\mathbf{x},\ \delta).$$

A function is *locally Lipschitz continuous on a set* $S \subseteq R^n$ if it is locally Lipschitz continuous at every point belonging to S. If $S = R^n$ the function is *locally Lipschitz continuous.* A function is *Lipschitz continuous on a set* $S \subseteq R^n$ if there exists a scalar K such that

$$|f(\mathbf{y}) - f(\mathbf{z})| \leq K \|\mathbf{y} - \mathbf{z}\|, \ \forall\, \mathbf{y},\ \mathbf{z} \in S.$$

If $S = R^n$ the function is *Lipschitz continuous.*

It follows from the definition of Lipschitz continuity that a function is Lipschitz continuous, if there exist a constant K such that the slope of the line segment between any two points of the function graph is smaller than K, or

$$\frac{|f(\mathbf{y}) - f(\mathbf{z})|}{\|\mathbf{y} - \mathbf{z}\|} \leq K, \ \forall\, \mathbf{y},\ \mathbf{z} \in R^n,\ \mathbf{y} \neq \mathbf{z}.$$

For a Lipschitz continuous function at a point $\mathbf{x}^*$ there exists a double cone such that the function graph stays outside of the double cone. If the function is Lipschitz continuous, the origin of the double cone can be moved along the graph so that the whole graph always stays outside of the double cone.

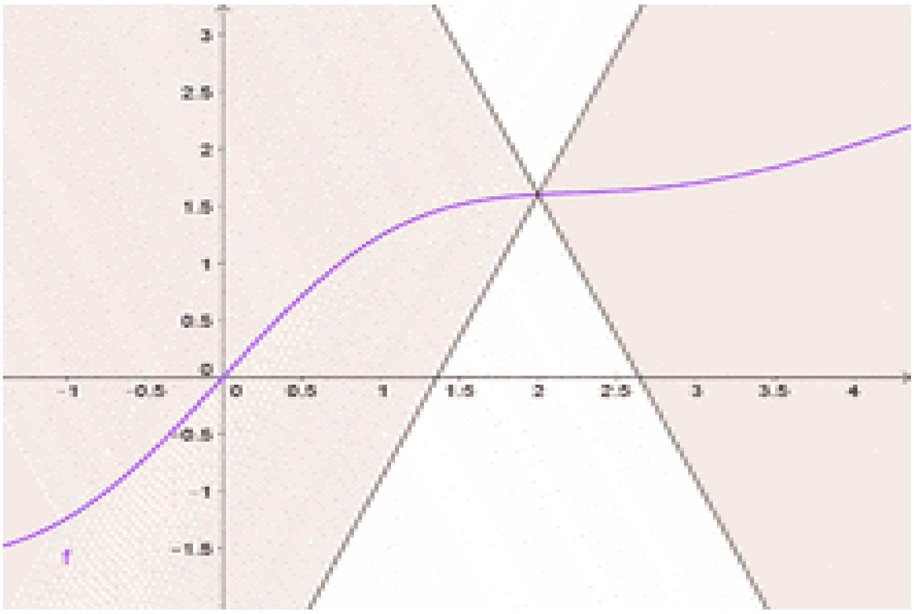


Figure 9: A Lipschitz continuous function

Lemma 3.9. [1] Let f be a convex function on the open convex set $S \subset R^n$. If f is bounded from above in a neighborhood of one point $\mathbf{x}^*$ of S, then it is locally bounded, that is, each $\mathbf{x}$ of S has a neighborhood on which f is bounded.

Proof. Let us first show that if f is bounded from above in a neighborhood of $\mathbf{x}^*$, it is also bounded from below in the same neighborhood. Suppose that f is bounded from above by a number M in $B(\mathbf{x}^*,\ \epsilon)$. Let us express every $\mathbf{z} \in B(\mathbf{x}^*,\ \epsilon)$ as $\mathbf{z} = \mathbf{x}^* + \theta\mathbf{y}$, where $\mathbf{y} \in R^n$ is a vector such that $\|\mathbf{y}\| = 1$ and θ is a sufficiently small positive number. Then

$$\mathbf{x}^* = \frac{1}{2}(\mathbf{x}^* + \theta\mathbf{y}) + \frac{1}{2}(\mathbf{x}^* - \theta\mathbf{y}).$$

By the definition of a convex function

$$f(\mathbf{x}^*) \le \frac{1}{2} f(\mathbf{x}^* + \theta \mathbf{y}) + \frac{1}{2} f(\mathbf{x}^* - \theta \mathbf{y}),$$

and thus

$$2f(\mathbf{x}^*) - f(\mathbf{x}^* - \theta \mathbf{y}) \le f(\mathbf{x}^* + \theta \mathbf{y}).$$

By the hypothesis f is bounded from above by a number M, or $f(\mathbf{x}^* - \theta \mathbf{y}) \le M$, hence

$$2f(\mathbf{x}^*) - M \le f(\mathbf{x}^* + \theta \mathbf{y}) = f(\mathbf{z}),$$

and f is bounded from below for every $\mathbf{z} \in B(\mathbf{x}^*,\ \epsilon)$.

Let $\mathbf{x} \in S,\ \mathbf{x} \neq \mathbf{x}^*$. Then $\mathbf{x} = \mathbf{x}^* + \alpha \mathbf{y}$, where $\mathbf{y} \in R^n,\ ||\mathbf{y}|| = 1$ and α is a positive number. Choose $\rho > \alpha$ such that $\mathbf{u} = \mathbf{x}^* + \rho \mathbf{y} \in S$ and let $\lambda = \alpha/\rho$. Then

$$B(\mathbf{x},\ \delta) = \{\mathbf{v} \mid \mathbf{v} \in S,\ \mathbf{v} = (1-\lambda)\mathbf{z} + \lambda \mathbf{u},\ \mathbf{z} \in B(\mathbf{x}^*,\ \epsilon)\}$$

is a neighborhood of $\mathbf{x}$ with radius $\delta = (1-\lambda)\epsilon$. Also, for $\mathbf{v} \in B(\mathbf{x},\ \delta)$

$$f(\mathbf{v}) \le (1-\lambda) f(\mathbf{z}) + \lambda f(\mathbf{u}) \le (1-\lambda) M + \lambda f(\mathbf{u}),$$

by the convexity of f. That is, f is bounded from above on $B(\mathbf{x},\ \delta)$, and by the first part of the proof, f is also bounded from below on $B(\mathbf{x},\ \delta)$. □

Theorem 3.10. [1] Let f be a convex function on the open convex set $S \subset R^n$. If f is bounded from below in a neighborhood of one point of S, then f if locally Lipschitz continuous in S, hence continuous in S.

Proof. By the previous lemma, for every $\mathbf{x}^* \in S$ there is a neighborhood $B(\mathbf{x}^*,\ 2\epsilon)$ on which f is bounded, or $x \in B(\mathbf{x}^*,\ 2\epsilon) \Rightarrow |f(\mathbf{x})| \le M$. We now show that f is locally Lipschitz continuous at $\mathbf{x}^*$, or that there exists a constant K for which $|f(\mathbf{x}) - f(\mathbf{y})| \le K\,||\mathbf{x} - \mathbf{y}||\,,\ \forall\ \mathbf{x},\ \mathbf{y} \in B(\mathbf{x}^*,\ \epsilon)$. Suppose, on the contrary, that there are points $\mathbf{x}$ and $\mathbf{y}$ in $B(\mathbf{x}^*,\ \epsilon)$ such that

$$\frac{f(\mathbf{x}) - f(\mathbf{y})}{||\mathbf{x} - \mathbf{y}||} > \frac{2M}{\epsilon}.$$

Let now $\mathbf{z} \in B(\mathbf{x}^*,\ 2\epsilon)$ be the point such that $\mathbf{y} = \lambda\mathbf{z} + (1-\lambda)\mathbf{x}$ and $||\mathbf{z} - \mathbf{y}|| = \epsilon$. Restricting f to the line segment of $\mathbf{y}$ and $\mathbf{z}$ it follows from the convexity of f that

$$\frac{f(\mathbf{y}) - f(\mathbf{z})}{||\mathbf{y} - \mathbf{z}||} \geq \frac{f(\mathbf{x}) - f(\mathbf{y})}{||\mathbf{x} - \mathbf{y}||} > \frac{2M}{\epsilon},$$

implying that $f(\mathbf{y}) - f(\mathbf{z}) > 2M$ and hence $|f(\mathbf{z})| > M$, which is a contradiction, and f must be locally Lipschitz continuous on S, hence continuous on S. □

Next we will show that every continuously differentiable function is locally Lipschitz continuous.

Theorem 3.11. [16] Every continuously differentiable function is locally Lipschitz continuous.

Proof. This proof is for vector valued functions, of which real valued functions are a special case. Let $F : R^n \to R^n$ be continuously differentiable. Fix any two points $\mathbf{z}, \mathbf{y} \in R^n$ and define the function $f : [0,1] \to R$ as $f(\theta) := F(\mathbf{z} + \theta(\mathbf{y} - \mathbf{z}))$. It is clear that

$$f(0) = F(\mathbf{z}) \text{ and } f(1) = F(\mathbf{y}). \tag{7}$$

Furthermore, by the chain rule, we know that f is differentiable and that

$$\frac{d}{d\theta} f(\theta) = \nabla F(\mathbf{z} + \theta(\mathbf{y} - \mathbf{z}))(\mathbf{y} - \mathbf{z}) \tag{8}$$

Here $\nabla F(\mathbf{w})$ is the Jacobian matrix of F at $\mathbf{w}$:

$$\begin{bmatrix} \frac{\partial f_1}{\partial w_1}(\mathbf{w}) & \cdots & \frac{\partial f_1}{\partial w_n}(\mathbf{w}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial w_1}(\mathbf{w}) & \cdots & \frac{\partial f_n}{\partial w_n}(\mathbf{w}) \end{bmatrix}, \tag{9}$$

where $\frac{\partial f_i}{\partial w_j}(\mathbf{w})$ is the partial derivative of the i-th component of f with respect to the j-th coordinate.

By the definition of f and the and the fundamental theorem of calculus, we have

$$F(\mathbf{y}) - F(\mathbf{z}) = f(1) - f(0) = \int_0^1 f'(\theta) d\theta$$

$$= \left(\int_0^1 \nabla F(\mathbf{z} + \theta(\mathbf{y} - \mathbf{z})) d\theta \right) (\mathbf{y} - \mathbf{z});$$

where the integral in the last line is a matrix whose i, j-th component is given by

$$\int_0^1 \frac{\partial f_i}{\partial w_j} (\mathbf{z} + \theta(\mathbf{y} - \mathbf{z})) d\theta.$$

It follows that

$$||F(\mathbf{y}) - F(\mathbf{z})|| \leq || \int_0^1 \nabla F(\mathbf{z}\theta(\mathbf{y} - \mathbf{z})) d\theta || || \mathbf{y} - \mathbf{z}||,$$

where we use the notation $|| \cdot ||$ for both the vector norm and its associated matrix norm.

Notice that, by the triangle inequality for integrals, we have

$$\left|\left| \int_0^1 \nabla F(\mathbf{z} + \theta(\mathbf{v} - \mathbf{z})) d\theta \right|\right| \leq \int_0^1 ||\nabla F(\mathbf{z} + \theta(\mathbf{v} - \mathbf{z}))|| d\theta$$

$$\sup_{\theta \in [0,1]} ||\nabla F(\mathbf{z} + \theta(\mathbf{v} - \mathbf{z}))|| \int_0^1 d\theta$$

$$\sup_{\theta \in [0,1]} ||\nabla F(\mathbf{z} + \theta(\mathbf{v} - \mathbf{z}))||.$$

Furthermore by the equivalence of matrix norms we have

$$\exists c_o > 0 : ||A|| \leq c_0 ||A||_\infty, \ \forall \ A \in R^{n \times n}, \tag{10}$$

where c_0 depends only on n and $|| \cdot ||_\infty$ is the maximum row norm sum $\max_{1 \leq i \leq n} \left(\sum_{j=1}^n |a_{ij}| \right)$.

Thus, to establish Lipschitz continuity, we fix an arbitrary point $\mathbf{x} \in R^n$ and we establish a bound for $\int_0^1 \nabla F(\mathbf{z} + \theta(\mathbf{y} - \mathbf{z})) d\theta$ in an approptiate neighbourhood B of $\mathbf{x}$.

Let $B = B(\mathbf{x}, \delta)$, with δ arbitrary. Since F is continuously differentiable on B, there exists K_0 such that

$$\sup_{\mathbf{w} \in B} \left|\left| \frac{\partial f_i}{\partial w_j} (\mathbf{w}) \right|\right| \leq K_0, \ \forall \ i, j \in [1 : n].$$

Here we have applied the Weierstrass Theorem which says that each continuous function $\left(\frac{\partial f_i}{\partial w_j} (\mathbf{w}) \right)$ is bounded on a closed and bounded set cl(B).

Now, given $\mathbf{z}, \mathbf{y} \in B$, it follows that $\mathbf{z} + \theta(\mathbf{z} - \mathbf{y}) \in B \; \forall \; \theta \in [0, 1]$ because B is a ball, so

$$\sup_{\theta \in [0,1]} ||\nabla F(\mathbf{z} + \theta(\mathbf{y} - \mathbf{z}))|| \leq c_0 m K_0 =: K.$$

It follows that

$$||F(\mathbf{y}) - F(\mathbf{z})|| \leq K||\mathbf{y} - \mathbf{z}||, \; \forall \; \mathbf{y}, \mathbf{z} \in B;$$

which is to say that F is Lipschitz continuous on $B(\mathbf{x}; \delta)$. Since $\mathbf{x}$ is arbitrary this means that F is locally Lipschitz continuous on R^n. □

Next let us define the directional derivative of a function. For smooth functions the directional derivative exists for every point in the domain of f.

Definition 3.12. [2] Let S be a nonempty set in R^n and let $f : S \to R$. Let $\mathbf{x}^* \in S$ and $\mathbf{d}$ be a nonzero vector such that $\mathbf{x}^* + \lambda \mathbf{d} \in S$ for a sufficiently small positive number λ. The *directional derivative* $f^{'}(\mathbf{x}^*; \mathbf{d})$ of f at $\mathbf{x}^*$ along the vector $\mathbf{d}$ is given by

$$f^{'}(\mathbf{x}^*; \mathbf{d}) = \lim_{\lambda \to 0^+} \frac{f(\mathbf{x}^* + \lambda \mathbf{d}) - f(\mathbf{x}^*)}{\lambda}.$$

If f is differentiable, its directional derivative can be expressed with its gradient in the form $f^{'}(\mathbf{x}^*; \mathbf{d}) = \nabla f(\mathbf{x}^*)^T \mathbf{d}$. The more general variant of the directional derivative is known as the *Gateaux derivative* defined between any two locally convex topological vector spaces.

Lemma 3.13. [2] Let S be a nonempty convex set in R^n and $f : S \to R$ a convex function. Then the directional derivative $f^{'}(\mathbf{x}^*; \mathbf{d})$ exists for a point $\mathbf{x}^* \in \text{int}(S)$ for every direction $\mathbf{d} \neq \mathbf{0}$. Furthermore,

$$f^{'}(\mathbf{x}^*; \mathbf{d}) = \inf_{\lambda > 0} \frac{f(\mathbf{x}^* + \lambda \mathbf{d}) - f(\mathbf{x}^*)}{\lambda}.$$

Proof. Let $0 < \lambda_1 < \lambda_2$ be small enough numbers such that $\mathbf{x}^* + \lambda_1 \mathbf{d}$ and $\mathbf{x}^* + \lambda_2 \mathbf{d} \in \text{int}(S)$. By the convexity of f

$$f(\mathbf{x}^* + \lambda_1 \mathbf{d}) = f\left[\frac{\lambda_1}{\lambda_2}(\mathbf{x}^* + \lambda_2 \mathbf{d}) + \left(1 - \frac{\lambda_1}{\lambda_2}\right)\mathbf{x}^*\right]$$

$$\leq \frac{\lambda_1}{\lambda_2} f(\mathbf{x}^* + \lambda_2 \mathbf{d}) + \left(1 - \frac{\lambda_1}{\lambda_2}\right) f(\mathbf{x}^*).$$

This inequality implies that

$$\frac{f(\mathbf{x}^* + \lambda_1 \mathbf{d}) - f(\mathbf{x}^*)}{\lambda_1} \leq \frac{f(\mathbf{x}^* + \lambda_2 \mathbf{d}) - f(\mathbf{x}^*)}{\lambda_2}.$$

Therefore $[f(\mathbf{x}^* + \lambda \mathbf{d})]/\lambda$ is a nondecreasing function of $\lambda > 0$. Let λ be a positive real number, for which $\mathbf{x}^* + \lambda \mathbf{d} \in \text{int}(S)$. Then, by the convexity of f

$$f(\mathbf{x}^*) = f\left(\frac{\lambda}{1+\lambda}(\mathbf{x}^* - \mathbf{d}) + \frac{1}{1+\lambda}(\mathbf{x}^* + \lambda \mathbf{d})\right)$$
$$\leq \frac{\lambda}{1+\lambda} f(\mathbf{x}^* - \mathbf{d}) + \frac{\lambda}{1+\lambda} f(\mathbf{x}^* + \lambda \mathbf{d}),$$

and thus

$$\frac{f(\mathbf{x}^* - \mathbf{d}) - f(\mathbf{x}^*)}{\lambda} \geq f(\mathbf{x}^*) - f(\mathbf{x}^* - \mathbf{d}).$$

Therefore $[f(\mathbf{x}^* - \mathbf{d}) - f(\mathbf{x}^*)]/\lambda$ is lower bounded and it has a limit, as $\lambda \to 0$. Hence the limit in the theorem exists and is given by

$$\lim_{\lambda \to 0^+} \frac{f(\mathbf{x}^* + \lambda \mathbf{d}) - f(\mathbf{x}^*)}{\lambda} = \inf_{\lambda > 0} \frac{f(\mathbf{x}^* + \lambda \mathbf{d}) - f(\mathbf{x}^*)}{\lambda}.$$

□

3.2 Subgradients and extrema

In this section we will deal with the optimization of convex functions. First let us present necessary definitions and theorems.

Definition 3.14. [2] Let $S \subseteq R^n$ be a nonempty set, and let $\mathbf{z}$ be a point on the boundary of S, that is, $\mathbf{z} \in \delta(S)$. Then the hyperplane

$$H = \{\mathbf{x} \mid \mathbf{p}^T(\mathbf{x} - \mathbf{z}) = 0,\ \mathbf{p} \neq \mathbf{0}\}$$

is a *supporting hyperplane* of S at $\mathbf{z}$, if one of the following is true:

1. $\mathbf{p}^T(\mathbf{x} - \mathbf{z}) \geq 0,\ \forall\, \mathbf{x} \in S$
2. $\mathbf{p}^T(\mathbf{x} - \mathbf{z}) \leq 0,\ \forall\, \mathbf{x} \in S.$

Theorem 3.15. [2] Let $S \subseteq R^n$ be a nonempty convex set, and let $\mathbf{z} \in \delta(S)$. Then there exists a hyperplane that supports S at $\mathbf{z}$; that is, there exists a nonzero vector $\mathbf{p}$ such that

$$\mathbf{p}^T(\mathbf{x} - \mathbf{z}) \leq 0, \ \forall \ \mathbf{x} \in \text{cl}(S).$$

Proof. Since $\mathbf{z} \in \delta(S)$ there exists a sequence $\{\mathbf{y}_k\}$ not in cl(S) such that $\mathbf{y}_k \to \mathbf{z}$. By the Separation Theorem ([2], Theorem 2.3.4) there exists for each $\mathbf{y}_k$ a vector $\mathbf{p}_k$, for which $\|\mathbf{p}_k\| = 1$, such that

$$\mathbf{p}_k^T \mathbf{y}_k > \mathbf{p}_k^T \mathbf{x}, \ \forall \ \mathbf{x} \in \text{cl}(S).$$

Since the sequence $\{\mathbf{p}_k\}$ is bounded, it has a convergent subsequence $\{\mathbf{p}_k\}_\kappa$ with limit $\mathbf{p}$ whose norm is equal to one. Let us fix $\mathbf{x} \in$ cl(S) and take the limit as $k \in \kappa$ approaches infinity. Then we have

$$\mathbf{p}^T \mathbf{z} \geq \mathbf{p}^T \mathbf{x}.$$

□

Definition 3.16. [2] Let $S \subseteq R^n$ be a nonempty convex set, and let $f : S \to R$ be convex. Then $\boldsymbol{\xi}$ is called a *subgradient* (see Figure 10) of f at $\mathbf{x}^*$ if

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*), \ \forall \ \mathbf{x} \in S.$$

The set of subgradients of f at $\mathbf{x}^*$, $\partial f(\mathbf{x}^*)$, is called the *subdifferential.*

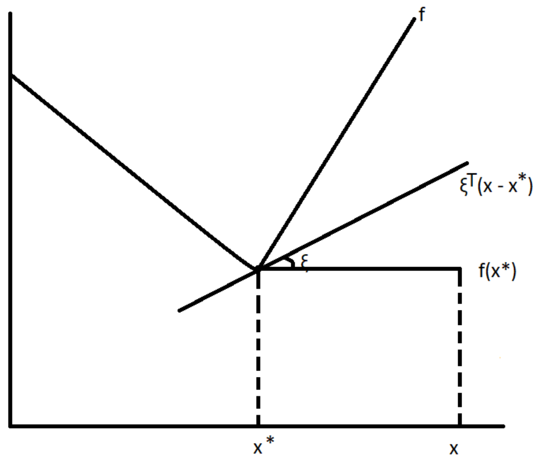


Figure 10: Subgradient

From this definition it directly follows that the subdifferential of f at $\mathbf{x}^*$ is a convex set. The function $f(\mathbf{x}^*)+\boldsymbol{\xi}^T(\mathbf{x}-\mathbf{x}^*)$ corresponds to a supporting hyperplane of the epigraph of f. The subgradient vector $\boldsymbol{\xi}$ corresponds to the slope of the supporting hyperplane.

Theorem 3.17. Let $S \subseteq R^n$ be a nonempty convex set and let $f : S \to R$ be convex. Then for $\mathbf{x}^* \in \text{int}(S)$ there exists a vector $\boldsymbol{\xi}$ such that the hyperplane

$$H = \{(\mathbf{x},\ y) \mid y = f(\mathbf{x}^*) + \boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*)\}$$

supports epi$f$ at $[\mathbf{x}^*,\ f(\mathbf{x}^*)]$. In particular,

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*),\ \forall\ \mathbf{x} \in S,$$

that is, $\boldsymbol{\xi}$ is a subgradient of f at $\mathbf{x}^*$.

Proof. By the convexity of f, epi$f$ is a nonempty convex set. The point $[\mathbf{x}^*, f(\mathbf{x}^*)]$ belongs to the boundary of epi$f$, so by Theorem 3.15 there exists a nonzero vector $(\boldsymbol{\xi}_0, \mu)$, $\boldsymbol{\xi}_0 \in R^n$, $\mu \in R$ such that

$$\boldsymbol{\xi}_0^T(\mathbf{x} - \mathbf{x}^*) + \mu[y - f(\mathbf{x}^*)] \leq 0,\ \forall\ (\mathbf{x},\ y) \in \text{epi}f.$$

Scalar μ is not positive, because otherwise the above inequality can be contradicted by choosing y sufficiently large. Let us show that $\mu < 0$. By contradiction, suppose that $\mu = 0$. Then $\boldsymbol{\xi}_0^T(\mathbf{x} - \mathbf{x}^*) \leq 0, \ \forall \ \mathbf{x} \in S$. Since $\mathbf{x}^* \in \mathrm{int}(S)$, there exists $\lambda > 0$ such that $\mathbf{x}^* + \lambda\boldsymbol{\xi}_0 \in S$, and therefore $\lambda\boldsymbol{\xi}_0^T\boldsymbol{\xi}_0 \leq 0$. This implies that $\boldsymbol{\xi}_0 = \mathbf{0}$ and $(\boldsymbol{\xi}_0, \ \mu) = (\mathbf{0}, \ 0)$, contradicting the fact that $(\boldsymbol{\xi}_0, \ \mu)$ is a nonzero vector. Therefore $\mu < 0$. Denoting $\boldsymbol{\xi}_0/|\mu|$ by $\boldsymbol{\xi}$ and dividing the above inequality by $|\mu|$ we get

$$\boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*) - y + f(\mathbf{x}^*) \leq 0, \ \forall \ (\mathbf{x}, \ y) \in \mathrm{epi}f.$$

In particular, the hyperplane $H = \{(\mathbf{x}, \ y) : y = f(\mathbf{x}^*) + \boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*)\}$ supports epi$f$ at $[\mathbf{x}^*, \ f(\mathbf{x}^*)]$. By letting $y = f(\mathbf{x})$, we get $f(\mathbf{x}) \geq f(\mathbf{x}^*) + \boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*), \ \forall \ \mathbf{x} \in S$. □

Theorem 3.18. [1] Let f be a differentiable function of the open convex set $S \subset R^n$. It is convex if and only if for every $\mathbf{x}^* \in S$, $\mathbf{x} \in S$ we have

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*).$$

It is strictly convex if and only if the above inequality is strict for $\mathbf{x} \neq \mathbf{x}^*$.

Proof. Let f be convex on S and let $\mathbf{x}^* \in S, \mathbf{x} \in S, \ \mathbf{x}^* \neq \mathbf{x}$. By the convexity of f

$$f(\mathbf{x}^* + \lambda(\mathbf{x} - \mathbf{x}^*)) = f(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}^*), \ \lambda \in (0, \ 1],$$

or

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \frac{1}{\lambda}[f(\mathbf{x}^* + \lambda(\mathbf{x} - \mathbf{x}^*)) - f(\mathbf{x}^*)], \ \lambda \neq 0.$$

By substituting the definition of differentiability to the above inequality we get

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) + \alpha(\mathbf{x}^*, \ \lambda(\mathbf{x} - \mathbf{x}^*)) \left|\left|\mathbf{x} - \mathbf{x}^*\right|\right|.$$

Since $\alpha(\mathbf{x}^*, \ \lambda(\mathbf{x} - \mathbf{x}^*))$ approaches zero as λ approaches zero we obtain the inequality of the theorem from the above inequality.

Conversely, let us assume that $\mathbf{x}_1 \in S, \ \mathbf{x}_2 \in S, \ 0 \leq \lambda \leq 1$. Then

$$f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + (1-\lambda)(\mathbf{x}_1 - \mathbf{x}_2)^T\nabla f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq f(\mathbf{x}_1)$$

and

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + \lambda(\mathbf{x}_2 - \mathbf{x}_1)^T \nabla f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq f(\mathbf{x}_2).$$

Multiplying the first inequality with λ and the second with $(1-\lambda)$ and adding up we get

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2),$$

and therefore the function is convex.

Let us prove the result for the strictly convex case. Let f be strictly convex and $\mathbf{x}^* \in S,\ \mathbf{x} \in S,\ \mathbf{x}^* \neq \mathbf{x}$. Since f is convex, the inequality of the theorem follows from the above results.

Then let us show that equality cannot hold. Suppose, to the contrary, that

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*).$$

Then, for $0 < \lambda < 1$,

$$f(\lambda \mathbf{x}^* + (1-\lambda)\mathbf{x}) < \lambda f(\mathbf{x}^*) + (1-\lambda) f(\mathbf{x})$$

$$= f(\mathbf{x}^*) + (1-\lambda)\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*).$$

Now let $\mathbf{x}^0 = \lambda \mathbf{x}^* + (1-\lambda)\mathbf{x}$. Then, since $\mathbf{x}^0 \in S$ and f is convex we get the inequality of the theorem by replacing $\mathbf{x}$ with $\mathbf{x}^0$; that is,

$$f(\lambda \mathbf{x}^* + (1-\lambda)\mathbf{x}) \geq f(\mathbf{x}^*) + (1-\lambda)\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*).$$

This causes a contradiction in which equality cannot hold. The proof of the converse statement for a strictly convex function is similar to the convex case. □

With this result we get the sufficient condition for the minimum of a convex function.

Theorem 3.19. [1] Let f be a differentiable (strictly) convex function on the open convex set S. If

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

at a point $\mathbf{x}^*$ of set S, then f attains its (unique) global minimum at $\mathbf{x}^*$.

Proof. It follows from the previous theorem that for every point $\mathbf{x} \in S$

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*).$$

The inequality is strict for a strictly convex function. If $\nabla f(\mathbf{x}^*) = 0$, then for every $\mathbf{x} \in S$

$$f(\mathbf{x}) \geq f(\mathbf{x}^*).$$

The inequality is strict for a strictly convex function. Therefore $\mathbf{x}^*$ is a (unique) global minimum of f. □

We get the same result by proving a more general theorem for nonsmooth functions.

Theorem 3.20. [2] Let $f : S \to R$ be a convex function, and let S be a nonempty convex set in R^n. Point $\mathbf{x}^* \in S$ is a global minimum of f if f has a subgradient $\boldsymbol{\xi}$ of subdifferential $\partial f(\mathbf{x}^*)$ such that

$$\boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*) \geq 0, \ \forall \, \mathbf{x} \in S.$$

Proof. Suppose that $\boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x} \in S$, where $\boldsymbol{\xi}$ is a subgradient of f at $\mathbf{x}^*$. By convexity of f, we get

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \boldsymbol{\xi}^T(\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}^*), \ \forall \, \mathbf{x} \in S.$$

Therefore $\mathbf{x}^*$ is a global minimum of f. □

It follows from the theorem that $\mathbf{x}^*$ is a global minimum of f if $\mathbf{0} \in \partial f(\mathbf{x}^*)$.

The next theorem shows that subgradients really are generalizations of the classical gradient.

Theorem 3.21. [18] If $f : R^n \to R$ is convex and differentiable at $\mathbf{x} \in R^n$, then

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}.$$

Proof. See, e.g. [18], Theorem 4.30. □

4 Nonsmooth and nonconvex functions

In this chapter we will discuss functions which are not continuously differentiable nor convex. They have points in their domain where their gradient is not continuous. In these points we will define Clarke's generalized directional derivative and subgradient. We will also handle generalized convexities for nonsmooth functions.

4.1 Generalized derivative

Let us define the generalized directional derivative and discuss its properties.

Definition 4.1. [19] Let $f : R^n \to R$ be locally Lipschitz-continuous at $\mathbf{x}^* \in R^n$. The *Clarke's generalized directional derivative* of f at $\mathbf{x}^*$ in direction $\mathbf{d} \in R^n$ is defined as follows:

$$f^o(\mathbf{x}^*; \mathbf{d}) = \limsup_{\mathbf{y} \to \mathbf{x}^*,\ t \to 0} \frac{f(\mathbf{y} + t\mathbf{d}) - f(\mathbf{y})}{t}, \; f^o(\mathbf{x}^*; \mathbf{d}) < \infty.$$

In the definition of the classical directional derivative the base point for taking differences is a fixed vector $\mathbf{x}^*$. In the generalized directional derivative they are taken from a variable vector $\mathbf{y}$ which approaches $\mathbf{x}^*$.

Example 4.2. Let us define a convex function $f(x) = |x|$. The function can be written as

$$f(x) = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}.$$

The derivative of f is

$$f^{'}(x) = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases}.$$

Function derivative $f^{'}$ is not continuous at $x = 0$. It is nonsmooth at this point. Let us calculate the generalized directional derivative for $x = 0$:

$$\begin{aligned} f^o(0; d) &= \lim_{y \to 0} \sup_{t \to 0} \frac{f(y + td) - f(y)}{t} \\ &= \lim_{y \to 0} \sup_{t \to 0} \frac{|y + td| - |y|}{t} \end{aligned}$$

$$= \limsup_{t \to 0} \frac{|td|}{t}$$

$$= \limsup_{t \to 0} |d| = |d| .$$

By substituting $d = 1$ or $d = -1$ we get $f^o(0; d) = 1$ for both directions. Note that for a convex function $f^o(\mathbf{x}; \mathbf{d}) = f^{'}(\mathbf{x}; \mathbf{d})$ [18, Theorem 3.8].

Example 4.3. Let us calculate the derivative and generalized directional derivative for the nonconvex function $f(x) = -|x|$:

$$f^{'}(x) = \begin{cases} -1, & x > 0 \\ 1, & x < 0 \end{cases} .$$

The generalized directional derivative for f at $x = 0$ is

$$f^o(0; d) = \limsup_{y \to 0;\ t \to 0} \frac{f(y + td) - f(y)}{t}$$

$$= \limsup_{y \to 0 \ t \to 0} \frac{-|y + td| + |y|}{t}$$

$$= \limsup_{t \to 0} \frac{|td|}{t}$$

$$= \limsup_{t \to 0} |d| = |d| .$$

Like for $f = |x|$, $f^o(0, d) = 1$ for $d = 1$ and $d = -1$. However, the directional derivative $f^{'}(0, d)$ for $f(x) = -|x|$ is

$$f^{'}(0; d) = \begin{cases} -1, & d = 1 \\ -1, & d = -1 \end{cases} .$$

Next we will present some properties of f^o.

Theorem 4.4. [5] Let f be a locally Lipschitz continuous function at point $\mathbf{x}^*$ with constant K. Then

1. Function $\mathbf{d} \mapsto f^o(\mathbf{x}^*; \mathbf{d})$ is positively homogenous and subadditive in R^n and it holds that

$$|f^o(\mathbf{x}^*; \mathbf{d})| \leq K \, ||\mathbf{d}|| \, ,$$

2. Function $f^o(\mathbf{x};\mathbf{d})$ is upper semicontinuous as a function of $(\mathbf{x};\mathbf{d}) \in R^{2n}$ and Lipschitz continuous with constant K as a function of $\mathbf{d}$ in R^n,

3. $f^o(\mathbf{x}^*;-\mathbf{d}) = (-f)^o(\mathbf{x}^*;\mathbf{d})$.

Proof. Let us begin by proving the inequality of the first part. By the Lipschitz condition we get

$$\begin{aligned} |f^o(\mathbf{x}^*;\mathbf{d})| &= \left| \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{f(\mathbf{y}+t\mathbf{d})-f(\mathbf{y})}{t} \right| \\ &\leq \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{|f(\mathbf{y}+t\mathbf{d})-f(\mathbf{y})|}{t} \\ &\leq \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{K\,||\mathbf{y}+t\mathbf{d}-\mathbf{y}||}{t}, \end{aligned}$$

when $\mathbf{y},\ \mathbf{y}+t\mathbf{d} \in B(\mathbf{x}^*,\epsilon)$ with some $\epsilon > 0$. Thus

$$|f^o(\mathbf{x}^*;\mathbf{d})| \leq \frac{Kt\,||\mathbf{d}||}{t} = K\,||\mathbf{d}||$$

Next we prove that the derivative is positively homogeneous. Let $\lambda > 0$. Then

$$\begin{aligned} f^o(\mathbf{x}^*;\lambda\mathbf{d}) &= \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{f(\mathbf{y}+t\lambda\mathbf{d})-f(\mathbf{y})}{t} \\ &= \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \lambda \left\{ \frac{f(\mathbf{y}+t\lambda\mathbf{d})-f(\mathbf{y})}{\lambda t} \right\} \\ &= \lambda \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \left\{ \frac{f(\mathbf{y}+t\lambda\mathbf{d})-f(\mathbf{y})}{\lambda t} \right\} = \lambda f^o(\mathbf{x}^*;\mathbf{d}). \end{aligned}$$

Now we shall prove the subadditivity. Let $\mathbf{d},\ \mathbf{p} \in R^n$ be arbitrary. Then

$$\begin{aligned} f^o(\mathbf{x}^*; \mathbf{d}+\mathbf{p}) &= \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{f(\mathbf{y}+t(\mathbf{d}+\mathbf{p})) - f(\mathbf{y})}{t} \\ &= \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{f(\mathbf{y}+t\mathbf{d}+t\mathbf{p}) - f(\mathbf{y}+t\mathbf{p}) + f(\mathbf{y}+t\mathbf{p}) - f(\mathbf{y})}{t} \\ &\leq \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{f((\mathbf{y}+t\mathbf{p})+t\mathbf{d}) - f(\mathbf{y}+t\mathbf{p})}{t} \\ &+ \lim_{\mathbf{y}\to\mathbf{x}^*} \sup_{t\to 0} \frac{f(\mathbf{y}+t\mathbf{p}) - f(\mathbf{y})}{t} \\ &= f^o(\mathbf{x}^*; \mathbf{d}) + f^o(\mathbf{x}^*; \mathbf{p}). \end{aligned}$$

Thus $\mathbf{d} \mapsto f^o(\mathbf{x}^*; \mathbf{d})$ is subadditive.

Let us move to the second part of the theorem. Let $\{\mathbf{x}_i\}$ and $\{\mathbf{d}_i\} \subset R^n$ be sequences such that $\mathbf{x}_i \to \mathbf{x}$ and $\mathbf{d}_i \to \mathbf{d}$. By definition of upper limit, there exist sequences $\{\mathbf{y}_i\} \subset R^n$ and $\{t_i\} \subset R$ such that $t_i > 0$,

$$f^o(\mathbf{x}; \mathbf{d}_i) \leq [f(\mathbf{y}_i + t\mathbf{d}_i) - f(\mathbf{y}_i)]/t_i + 1/i$$

and

$$||\mathbf{y}_1 - \mathbf{x}_i|| + t_i < 1/i,\ \forall\ i \in N.$$

Now we have

$$\begin{aligned} f^o(\mathbf{x}_i; \mathbf{d}_i) - \frac{1}{i} &= \lim_{\mathbf{y}\to\mathbf{x}_i} \sup_{t\to 0} \frac{f(\mathbf{y}+t\mathbf{d}_i) - f(\mathbf{y})}{t} - \frac{1}{i} \\ &\leq \frac{f(\mathbf{y}_i + t_i\mathbf{d}_i) - f(\mathbf{y}_i)}{t_i} \\ &= \frac{f(\mathbf{y}_i + t_i\mathbf{d}_i) - f(\mathbf{y}_i)}{t_i} + \frac{f(\mathbf{y}_i + t_i\mathbf{d}_i) - f(\mathbf{y}_i + t_i\mathbf{d})}{t_i} \end{aligned}$$

and by the Lipschitz condition

$$\frac{|f(\mathbf{y}_i + t_i\mathbf{d}_i) - f(\mathbf{y}_i + t_i\mathbf{d})|}{t_i} \leq \frac{K\,||\mathbf{d}_i - \mathbf{d}||}{t_i} = K\,||\mathbf{d}_i - \mathbf{d}|| \to 0,$$

as $i \to \infty$ provided $\mathbf{y}_i + t_i\mathbf{d}_i,\ \mathbf{y}_i + t_i\mathbf{d} \in B(\mathbf{x}; \epsilon),\ \epsilon > 0$. As $i \to \infty$, we obtain

$$\lim_{i\to\infty} \sup f^o(\mathbf{x}_i; \mathbf{d}_i) \leq \lim_{i\to\infty} \sup \frac{f(\mathbf{y}_i + t_i\mathbf{d}) - f(\mathbf{y}_i)}{t_i} \leq f^o(\mathbf{x}; \mathbf{d}),$$

which establishes the upper semicontinuity.

Let us show the Lipschitz continuity of $f^\circ(\mathbf{x};\mathbf{d})$. Let $\mathbf{v},\ \mathbf{w} \in R^n$. If $\mathbf{y}+t\mathbf{v},\ \mathbf{y}+t\mathbf{w} \in B(\mathbf{x}^*, \epsilon)$, we obtain

$$f(\mathbf{y}+t\mathbf{v}) - f(\mathbf{y}+t\mathbf{w}) \le Kt\,||\mathbf{v}-\mathbf{w}||\,.$$

It follows that

$$\begin{aligned}&\lim_{\mathbf{y}\to\mathbf{x}^*}\sup_{t\to 0}\frac{f(\mathbf{y}+t\mathbf{v})-f(\mathbf{y})}{t}\\ \le&\lim_{\mathbf{y}\to\mathbf{x}^*}\sup_{t\to 0}\frac{f(\mathbf{y}+t\mathbf{w})-f(\mathbf{y})}{t}K\,||\mathbf{v}-\mathbf{w}||\end{aligned}$$

and

$$f^o(\mathbf{x}^*;\mathbf{v}) - f^o(\mathbf{x}^*;\mathbf{w}) \le K\,||\mathbf{v}-\mathbf{w}||\,.$$

By switching $\mathbf{v}$ and $\mathbf{w}$, we get

$$f^o(\mathbf{x}^*;\mathbf{w}) - f^o(\mathbf{x}^*;\mathbf{v}) \le K\,||\mathbf{v}-\mathbf{w}||\,.$$

From this we get

$$|f^o(\mathbf{x}^*;\mathbf{v}) - f^o(\mathbf{x}^*;\mathbf{w})| \le K\,||\mathbf{v}-\mathbf{w}||\,.$$

Let us prove the third part of the theorem by making the following calculation:

$$\begin{aligned}f^o(\mathbf{x}^*;-\mathbf{d}) &= \lim_{\mathbf{y}\to\mathbf{x}^*}\sup_{t\to 0}\frac{f(\mathbf{y}-t\mathbf{d})-f(\mathbf{y})}{t}\\ &= \lim_{\mathbf{u}\to\mathbf{x}^*}\sup_{t\to 0}\frac{(-f)(\mathbf{u}+t\mathbf{d})-(-f)(\mathbf{u})}{t},\end{aligned}$$

by substituting $\mathbf{u}=\mathbf{y}-t\mathbf{d}$. We obtain the result

$$f^o(\mathbf{x}^*;-\mathbf{d}) = (-f)^o(\mathbf{x}^*;\mathbf{d}).$$

□

4.2 Generalized subgradient

Next we will define the generalized subgradient with the generalized directional derivative and discuss its properties.

Definition 4.5. [19] Let $f : R^n \to R$ be a locally Lipschitz continuous function at a point $\mathbf{x}^* \in R^n$. The *Clarke subdifferential* of f at $\mathbf{x}^*$ is the set $\partial f(\mathbf{x}^*)$ of vectors $\boldsymbol{\xi} \in R^n$ such that

$$\partial f(\mathbf{x}^*) = \{\boldsymbol{\xi} \mid f^o(\mathbf{x}^*; \mathbf{d}) \geq \boldsymbol{\xi}^T \mathbf{d},\ \forall\ \mathbf{d} \in R^n\}.$$

Each $\boldsymbol{\xi} \in \partial f(\mathbf{x}^*)$ is called a *subgradient* of f at $\mathbf{x}^*$. For a convex function the above definition equals the subdifferential of a convex function (Definition 3.16).

Theorem 4.6. If f is continuously differentiable in $\mathbf{x}^*$, it follows that

$$\partial f(\mathbf{x}^*) = \{\nabla f(\mathbf{x}^*)\}.$$

Proof. Look [5], Theorem 3.1.7. □

Theorem 3.19 for smooth convex functions follows from Theorems 3.20 and 4.6.

Theorem 4.7. [5] Let f be locally Lipschitz continuous at $\mathbf{x}$. If f obtains its local minimum value at $\mathbf{x}$, we have

$$\mathbf{0} \in \partial f(\mathbf{x}).$$

Proof. Let us assume that f obtains its local minimum at $\mathbf{x}$. Then there exists a positive real number ϵ such that $f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x}) \geq 0,\ \forall\ 0 < t < \epsilon,\ \mathbf{d} \in R^n$. We get the result

$$f^o(\mathbf{x}; \mathbf{d}) = \limsup_{\substack{\mathbf{y} \to \mathbf{x} \\ t \to 0}} \frac{f(\mathbf{y} + t\mathbf{d}) - f(\mathbf{y})}{t} \geq \limsup_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} \geq 0.$$

It follows that

$$f^o(\mathbf{x}; \mathbf{d}) \geq 0 = \mathbf{0}^T \mathbf{d},\ \forall\ \mathbf{d} \in R^n,$$

so by the definition of subgradient $\mathbf{0}$ belongs to the subdifferential of f at $\mathbf{x}$. □

Theorem 4.8. [5] Let f be locally Lipschitz continuous at $\mathbf{x}$ with a constant K. Then

1. $\partial f(\mathbf{x})$ is a nonempty, convex and compact set such that $||\boldsymbol{\xi}|| \leq K$, $\forall\, \boldsymbol{\xi} \in \partial f(\mathbf{x})$.
2. $f^o(\mathbf{x}; \mathbf{d}) = \max\{\boldsymbol{\xi}^T \mathbf{d} \mid \boldsymbol{\xi} \in \partial f(\mathbf{x})\}$, $\forall\, \mathbf{d} \in R^n$.
3. The mapping $\partial f(\cdot) : R^n \to P(R^n)$, where $P(R^n)$ is the powerset containing every subset of R^n, is upper semicontinuous.

Proof. According to Theorem 4.4 $f^o(\mathbf{x}; \mathbf{d})$ is positively homogeneous and subadditive. Then, by Hanh-Banach Theorem ([5], Theorem 1.2.1), there exists a vector $\boldsymbol{\xi} \in R^n$ such that $\boldsymbol{\xi}^T \mathbf{d} \leq f^o(\mathbf{x}, \mathbf{d})$, $\forall\, \mathbf{d} \in R^n$. Then, by the definition of the subdifferential, $\partial f(\mathbf{x})$ is nonempty.

Let us prove the convexity of the subdifferential by choosing $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ and $\lambda \in [0,\ 1]$. We get

$$[\lambda \boldsymbol{\xi} + (1-\lambda)\boldsymbol{\xi}']^T \mathbf{d} \leq \lambda f^o(\mathbf{x}; \mathbf{d}) + (1-\lambda) f^o(\mathbf{x}; \mathbf{d}) = f^o(\mathbf{x}; \mathbf{d}),$$

where $\lambda \boldsymbol{\xi} + (1-\lambda)\boldsymbol{\xi}' \in \partial f(\mathbf{x})$. Therefore the subdifferential is convex.

Let us prove the compactness by showing that $\partial f(\mathbf{x})$ is closed and bounded. By Theorem 4.4

$$||\boldsymbol{\xi}||^2 = |\boldsymbol{\xi}^T \boldsymbol{\xi}| \leq |f^o(\mathbf{x}; \boldsymbol{\xi})| \leq K\, ||\boldsymbol{\xi}|| \,.$$

It follows from the randomness of $\boldsymbol{\xi}$ that

$$||\boldsymbol{\xi}|| \leq K, \ \forall\, \boldsymbol{\xi} \in \partial f(\mathbf{x}).$$

Therefore the subdifferential is bounded.

Then let us choose a sequence $\{\boldsymbol{\xi}_i\} \in \partial f(\mathbf{x})$ which approaches $\boldsymbol{\xi}$. We get

$$\boldsymbol{\xi}^T \mathbf{d} = \lim_{i\to\infty} \boldsymbol{\xi}_i^T \mathbf{d} = \lim_{i\to\infty} (\boldsymbol{\xi}_i^T \mathbf{d}) \leq \lim_{i\to\infty} f^o(\mathbf{x}; \mathbf{d}) = f^o(\mathbf{x}; \mathbf{d}).$$

Thus $\boldsymbol{\xi} \in \partial f(\mathbf{x})$, so the subdifferential is closed.

Now let us prove the second part of the theorem. By the definition of the subdifferential we get

$$f^o(\mathbf{x};\mathbf{d}) \geq \max\{\boldsymbol{\xi}^T\mathbf{d} \mid \boldsymbol{\xi} \in \partial f(\mathbf{x})\}.$$

Let us assume that there exists a vector $\mathbf{d}_1 \in R^n$, such that

$$f^o(\mathbf{x};\mathbf{d}_1) > \max\{\boldsymbol{\xi}^T\mathbf{d}_1 \mid \boldsymbol{\xi} \in \partial f(\mathbf{x})\}.$$

Then, by the Hanh-Banach Theorem, there exists a vector $\boldsymbol{\xi}_1 \in R^n$, such that $f^o(\mathbf{x};\mathbf{d}) \geq \boldsymbol{\xi}_1^T\mathbf{d}$, $\forall\ \mathbf{d} \in R^n$ and $f^o(\mathbf{x};\mathbf{d}_1) = \boldsymbol{\xi}_1^T\mathbf{d}_1$. Therefore $\boldsymbol{\xi}_1 \in \partial f(\mathbf{x})$ and we get

$$f^o(\mathbf{x};\mathbf{d}_1) > \max\{\boldsymbol{\xi}^T\mathbf{d}_1 \mid \boldsymbol{\xi} \in \partial f(\mathbf{x})\} \geq \boldsymbol{\xi}_1^T\mathbf{d}_1 = f^o(\mathbf{x};\mathbf{d}_1).$$

From this contradiction we come to the result

$$f^o(\mathbf{x};\mathbf{d}) = \max\{\boldsymbol{\xi}^T\mathbf{d}_1 \mid \boldsymbol{\xi} \in \partial f(\mathbf{x})\},\ \forall \mathbf{d} \in R^n.$$

The second part of the theorem is proved.

Let us prove the third part of the theorem. Let us choose a sequence $\{\mathbf{y}_i\} \in R^n$, which approaches $\mathbf{x}$ and a sequence $\{\boldsymbol{\xi}_i\} \in \partial f(\mathbf{y}_i)$, which approaches $\boldsymbol{\xi}$. Then for every $\mathbf{d} \in R^n$ we get

$$\boldsymbol{\xi}^T\mathbf{d} = \lim_{i\to\infty} \boldsymbol{\xi}_i^T\mathbf{d} = \lim_{i\to\infty}(\boldsymbol{\xi}_i^T\mathbf{d}) \leq \lim_{i\to\infty} \sup f^o(\mathbf{y}_i;\mathbf{d}).$$

By the second part of Theorem 4.4 $f^o(\mathbf{x};\dot{})$ is upper semicontinuous. From this we get

$$\boldsymbol{\xi}^T\mathbf{d} \leq f^o(\mathbf{x};\mathbf{d}).$$

Therefore $\partial f(\cdot)$ is upper semicontinuous. □

We shall later be considering vector-valued functions. For this we will need to define the generalized Jacobian matrix.

Definition 4.9. [18] Let $F : R^n \to R^m$ be a vector-valued function $F(\mathbf{x}) = (F_1(\mathbf{x}), ..., F_m(\mathbf{x}))^T$. We denote by Ω_F the set in R^n where F fails to be differentiable and by $\nabla F(\mathbf{x})$ for $\mathbf{x} \notin \Omega_F$ the usual $m \times n$ Jacobian matrix. Let F be locally Lipschitz continuous at $\mathbf{x}$. Then the *generalized Jacobian matrix* of F at $\mathbf{x}$ is the set

$$\partial F(\mathbf{x}) := \text{conv}\{A \in R^{m\times n} \mid \exists(\mathbf{x}_i) \subset R^n \backslash \Omega_F \text{ such that} (\mathbf{x}_i) \to \mathbf{x},\ \nabla F(\mathbf{x}_i) \to A\}.$$

Some basic properties of $\partial F(\mathbf{x})$ will now be listed.

Theorem 4.10. [18] Let F_i for $i = 1, ..., m$ be locally Lipschitz continuous at $\mathbf{x}$ with constant K_i. Then

1. $F(\mathbf{x}) = (F_1(\mathbf{x}), ..., F_m(\mathbf{x}))^T$ is locally Lipschitz continuous at $\mathbf{x}$ with constant $K = ||(K_1, ..., K_m)^T||_2$,
2. $\partial F(\mathbf{x})$ is a nonempty, convex and compact subset of $R^{m\times n}$,
3. the mapping $\partial F(\cdot) : R^n \to P(R^n)$ is upper semicontinuous.

Proof. See [18]. □

4.3 Other derivatives

In this section we define B- and F-differentiability, which will be used in Section 5.

Definition 4.11. [21] A function $F : R^n \to R^m$ is *Bouligand-differentiable* (B-differentiable) at a point $\mathbf{x}^* \in R^n$ if it is locally Lipschitz-continuous and directionally differentiable at $\mathbf{x}^*$. If f is B-differentiable at $\mathbf{x}^*$, we call the directional derivative $F'(\mathbf{x}^*, \mathbf{h})$ the *B-derivative* of F at $\mathbf{x}^*$ along $\mathbf{h}$.

The B-derivative is *strong* if the error function

$$e(\mathbf{y}) = F(\mathbf{y}) - F(\mathbf{x}^*) - F'(\mathbf{x}^*; \mathbf{y} - \mathbf{x})$$

satisfies

$$\lim_{\mathbf{y}^1 \neq \mathbf{y}^2,\ (\mathbf{y}^1, \mathbf{y}^2) \to (\mathbf{x}^*, \mathbf{x}^*)} \frac{e(\mathbf{y}^1) - e(\mathbf{y}^2)}{||\mathbf{y}^1 - \mathbf{y}^2||} = 0 \tag{11}$$

Definition 4.12. [21] A function $F : R^n \to R^m$ is *Fréchet-differentiable* (F-differentiable) at $\mathbf{x}^* \in R^n$, if it is directionally differentiable at $\mathbf{x}^*$ and

$$F(\mathbf{x}^* + \mathbf{h}) = F(\mathbf{x}^*) + F^{'}(\mathbf{x}^*; \mathbf{h}) + o(||\mathbf{h}||),$$

where $F^{'}(\mathbf{x}^*; \mathbf{h})$ is called the *F-derivative* of F at point $\mathbf{x}^*$, if it is linear and continuous. So, the existence of the F-derivative implies the existence of directional derivatives in all directions. Function F is *strongly F-differentiable* if

$$\lim_{\mathbf{y}_1 \neq \mathbf{y}_2,\ (\mathbf{y}_1, \mathbf{y}_2) \to (\mathbf{x}^*, \mathbf{x}^*)} \frac{F(\mathbf{y}_1) - F(\mathbf{y}_2) - F^{'}(\mathbf{x})(\mathbf{y}_1 - \mathbf{y}_2)}{||\mathbf{y}_1 - \mathbf{y}_2||} = 0.$$

4.4 Generalized pseudoconvexity

In this section we consider pseudoconvexity. The section is based on source [4]. First let us define pseudoconvexity for smooth functions.

Definition 4.13. A continuously differentiable function $f : R^n \to R$ is *pseudoconvex*, if for all $\mathbf{x},\ \mathbf{y} \in R^n$

$$f(\mathbf{y}) < f(\mathbf{x}) \Rightarrow \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) < 0.$$

It follows from the definition that if the gradient ascent of a pseudoconvex function at $\mathbf{x}$ to the direction $\mathbf{y} - \mathbf{x}$ is non-negative, the function is non-decreasing in this direction. An example of a pseudoconvex function is $f(x) = x + x^3$.

The result of a global minimum for a convex function (Theorem 3.19) can be weakened: a pseudoconvex function attains its global minimum at $\mathbf{x}^*$ if and only if $\nabla f(\mathbf{x}^*) = 0$. The concept of pseudoconvexity can be extended with the generalized directional derivative.

Definition 4.14. A function $f : R^n \to R$ is *f^o-pseudoconvex*, if it is locally Lipschitz continuous and for all $\mathbf{x},\ \mathbf{y} \in R^n$

$$f(\mathbf{y}) < f(\mathbf{x}) \Rightarrow f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) < 0.$$

The following result shows that f^o-pseudoconvexity is a natural extension of pseudoconvexity.

Theorem 4.15. If f is smooth, it is f^o-pseudoconvex if and only if it is pseudoconvex.

Proof. The theorem follows directly from the second part of Theorem 4.9, because for a smooth function

$$f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) = f^{'}(\mathbf{x}; \mathbf{y} - \mathbf{x}) = \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$$

. □

Lemma 4.16. A locally Lipschitz continuous function is f^o-pseudoconvex if and only if

$$f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) \geq 0 \Rightarrow f(\mathbf{y}) \geq f(\mathbf{x}).$$

The proof follows directly from the definition of f^o-pseudoconvexity.

Next we show that the sufficient optimality condition of pseudoconvex functions extends to f^o-pseudoconvex functions.

Theorem 4.17. A f^o-pseudoconvex function reaches its global minimum at $\mathbf{x}^*$ if and only if

$$\mathbf{0} \in \partial f(\mathbf{x}^*).$$

Proof. The necessary condition follows from Theorem 4.7. Let $\mathbf{0} \in \partial f(\mathbf{x}^*)$ and let $\mathbf{y} \in R^n$. Then, by the definition of the Clarke subdifferential, we get

$$f^o(\mathbf{x}^*; \mathbf{y} - \mathbf{x}^*) \geq \mathbf{0}^T(\mathbf{y} - \mathbf{x}^*) = 0.$$

Then, by the previous lemma,

$$f(\mathbf{y}) \geq f(\mathbf{x}^*).$$

□

Let us show the generality of f^o-pseudoconvexity with the following example.

Example 4.18. Let $f(x) = \min\{|x|, x^2\}$. Function f is clearly locally Lipschitz continuous but not convex or pseudoconvex. However, when $x < y$, the generalized directional derivative of f is

$$f^o(x, y - x) = \begin{cases} -1, & x \in (-\infty,\ -1] \\ 2x, & x \in (-1,\ 1] \\ 1, & x \in (1,\ \infty) \end{cases}.$$

It follows from the symmetry of f and Lemma 4.16 that f is f^o-pseudoconvex. Moreover, $\partial f(0) = \{0\}$, so f attains its global minimum at this point.

Definition 4.19. f^o is *pseudomonotone*, if for every $\mathbf{x}$ and $\mathbf{y} \in R^n$ we have

$$f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) \geq 0 \Rightarrow f^o(\mathbf{y}; \mathbf{x} - \mathbf{y}) \leq 0,$$

or equivalently if

$$f^o(\mathbf{y}; \mathbf{x} - \mathbf{y}) > 0 \Rightarrow f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) < 0.$$

In addition, it is *strictly pseudomonotone*, if

$$f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) \geq 0 \Rightarrow f^o(\mathbf{y}; \mathbf{x} - \mathbf{y}) < 0.$$

Let us write without the proof the Mean-Value Theorem, which is needed to prove the next theorem.

Mean-Value Theorem 4.20. [5] Let $\mathbf{x} \neq \mathbf{y} \in R^n$ and let f be Lipschitz continuous on an open set $U \subseteq R^n$ so that $[\mathbf{x},\ \mathbf{y}]$ is a subset of U. Then there exists a point $\mathbf{u}$ between $\mathbf{x}$ and $\mathbf{y}$ such that

$$f(\mathbf{y}) - f(\mathbf{x}) \in \partial f(\mathbf{u})^T(\mathbf{y} - \mathbf{x}).$$

Thus we have here a set composed of scalar products of subgradients with the vector $\mathbf{y} - \mathbf{x}$.

Now we will show a connection between f^o-pseudomonotonity and f^o-pseudoconvexity.

Theorem 4.21. Let f be locally Lipschitz continuous so that f^o is pseudomonotone. Then f is f^o-pseudoconvex.

Proof. Let us, on the contrary, assume that f is not f^o-pseudoconvex. Then there exist $\mathbf{x}$ and $\mathbf{y} \in R^n$, for which $f(\mathbf{y}) < f(\mathbf{x})$ and

$$f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) \geq 0. \tag{12}$$

By the Mean-Value Theorem there exists $\lambda^* \in (0,\ 1)$ such that $\mathbf{x}^* = \mathbf{x} + \lambda^*(\mathbf{y} - \mathbf{x})$ and

$$f(\mathbf{x}) - f(\mathbf{y}) \in \partial f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{y}).$$

By the definition of the Clarke subdifferential there exists a subgradient $\boldsymbol{\xi}^*$ such that

$$0 < f(\mathbf{x}) - f(\mathbf{y}) = \boldsymbol{\xi}^{*T}(\mathbf{x} - \mathbf{y}) \leq f^o(\mathbf{x}^*; \mathbf{x} - \mathbf{y}). \tag{13}$$

However, by (12) and the positive homogeneity of $\mathbf{d} \mapsto f^o(\mathbf{x}; \mathbf{d})$ we get

$$0 \geq f^o(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*) = \lambda^* f^o(\mathbf{x}^*; \mathbf{x} - \mathbf{y}) > 0,$$

which leads to a contradiction. Therefore f is f^o-pseudoconvex. □

The converse result for the theorem is also true. To prove this we will need a few lemmas.

Lemma 4.22. Let f be a f^o-pseudoconvex function, $\mathbf{x},\ \mathbf{y} \in R^n$ and $\lambda^* \in (0,\ 1)$. Denote $\mathbf{x}^* = \lambda^*\mathbf{x} + (1 - \lambda^*)\mathbf{y}$. Then $f(\mathbf{x}^*) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}$.

Proof. On the contrary assume that $f(\mathbf{x}^*) > \max\{f(\mathbf{x}), f(\mathbf{y})\}$. Since f is f^o-pseudoconvex and $\mathbf{d} \mapsto f^o(\mathbf{x}; \mathbf{d})$ is positively homogeneous, we get

$$0 > f^o(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*) = f^o(\mathbf{x}^*; (1 - \lambda^*)(\mathbf{x} - \mathbf{y})) = (1 - \lambda^*) f^o(\mathbf{x}^*; \mathbf{x} - \mathbf{y})$$

and thus

$$f^o(\mathbf{x}^*; \mathbf{x} - \mathbf{y}) < 0.$$

Correspondingly, we obtain

$$0 > f^o(\mathbf{x}^*; \mathbf{y} - \mathbf{x}^*) = f^o(\mathbf{x}^*; \lambda^*(\mathbf{y} - \mathbf{x})) = \lambda^* f^o(\mathbf{x}^*; \mathbf{y} - \mathbf{x})$$

and thus

$$f^o(\mathbf{x}^*; \mathbf{y}-\mathbf{x}) < 0.$$

Since $\mathbf{d} \mapsto f^o(\mathbf{x}; \mathbf{d})$ is subadditive, we get

$$0 > f^o(\mathbf{x}^*; \mathbf{x}-\mathbf{y}) + f^o(\mathbf{x}^*; \mathbf{y}-\mathbf{x}) \geq f^o(\mathbf{x}^*; (\mathbf{x}-\mathbf{y}) + (\mathbf{y}-\mathbf{x})) = f^o(\mathbf{x}^*; \mathbf{0}) = 0,$$

which is impossible. Therefore $f(\mathbf{x}^*) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}$. □

Let us present the next lemma without the proof.

Lemma 4.23. [4] Let f be a locally Lipschitz continuous function. Let us choose $\epsilon > 0$ and a non-zero $\mathbf{d} \in R^n$. Let Ω_f be a set of points in which f is not differentiable. Then

$$f^o(\mathbf{x}; \mathbf{d}) - \epsilon \leq \limsup\{\nabla f(\mathbf{y})^T \mathbf{d} \mid \mathbf{y} \to \mathbf{x},\ \mathbf{y} \notin \Omega_f\}.$$

Lemma 4.24. Let f be a f^o-pseudoconvex function. Then there exist no points $\mathbf{x},\ \mathbf{y} \in R^n$ such that

1. $f(\mathbf{x}) = f(\mathbf{y})$
2. $f(\mathbf{x}; \mathbf{y}-\mathbf{x}) > 0$.

Proof. On the contrary, let us assume that there exist points $\mathbf{x},\ \mathbf{y} \in R^n$ and $\delta > 0$ such that $f^o(\mathbf{x}; \mathbf{y}-\mathbf{x}) = \delta$ and $f(\mathbf{x}) = f(\mathbf{y})$. Since f is locally Lipschitz continuous, there exist $\epsilon,\ K > 0$ such that K is the the Lipschitz constant in the ball $B(\mathbf{x}; \epsilon)$. Since $f^o(\mathbf{x}; \mathbf{y}-\mathbf{x}) = \delta$, by the previous lemma there exists a sequence $\{\mathbf{z}^i\}$ of points where f is differentiable and $I \in N$ such that $\mathbf{z}^i \to \mathbf{x}$ and

$$f'(\mathbf{z}^i; \mathbf{y}-\mathbf{x}) = \nabla f(\mathbf{z}^i)^T(\mathbf{y}-\mathbf{x}) > \frac{\delta}{2}, \tag{14}$$

when $i \geq I$. Let

$$\widehat{\epsilon} = \min\left\{\epsilon, \frac{\delta}{2K}\right\}$$

and $\mathbf{z} \in B(\mathbf{x}; \widehat{\epsilon}) \cap \{(\mathbf{z}^i) | i \geq I\}$. According to Theorem 4.4 $f'(\mathbf{z}; \cdot)$ is Lipschitz continuous with the constant K. Therefore

$$|f'(\mathbf{z}; \mathbf{y}-\mathbf{x}) - f'(\mathbf{z}; \mathbf{y}-\mathbf{z})| \leq K\|\mathbf{y}-\mathbf{x}-(\mathbf{y}-\mathbf{z})\|$$

$$= K\|\mathbf{z} - \mathbf{x}\| < K\frac{\delta}{2K} = \frac{\delta}{2}. \tag{15}$$

Thus, $f'(\mathbf{z}; \mathbf{y} - \mathbf{z}) > 0$ according to (14) and (15). Since $f'(\mathbf{z}; \mathbf{y} - \mathbf{z}) > 0$, there exists $\mu \in (0,\ 1)$ such that

$$f(\mu\mathbf{z} + (1 - \mu)\mathbf{y}) > f(\mathbf{z}). \tag{16}$$

Since $f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) = \delta$, part 1 of Theorem 4.4 implies that there exists $\bar{\epsilon} > 0$ such that $f^o(\mathbf{x}; \mathbf{d}) > 0$, when $\mathbf{d} \in B(\mathbf{y} - \mathbf{x}; \bar{\epsilon})$. Let $\bar{\mathbf{z}} \in B(\mathbf{y}; \bar{\epsilon})$. Since

$$\|\bar{\mathbf{z}} - \mathbf{x} - (\mathbf{y} - \mathbf{x})\| = \|\bar{\mathbf{z}} - \mathbf{y}\| < \bar{\epsilon},$$

it follows that $\bar{\mathbf{z}} - \mathbf{x} \in B(\mathbf{y} - \mathbf{x}; \bar{\epsilon})$. Thus, $f^o(\mathbf{x}; \bar{\mathbf{z}} - \mathbf{x}) > 0$ and the f^o-pseudoconvexity of f implies that $f(\bar{\mathbf{z}}) \geq f(\mathbf{x}) = f(\mathbf{y})$. Thus $\mathbf{y}$ is a local minimum for f and by Theorem 4.7 we have $\mathbf{0} \in \partial f(\mathbf{y})$. Therefore, by Theorem 4.17 $\mathbf{y}$ is also a global minimum. Thus we have $f(\mathbf{y}) \leq f(\mathbf{z})$ and (16) implies that

$$f(\mu\mathbf{z} + (1 - \mu)\mathbf{y}) > \max\{f(\mathbf{z}), f(\mathbf{y})\},$$

which is impossible by Lemma 4.22.

□

Theorem 4.25. The generalized directional derivative of a f^o-pseudoconvex function f is pseudomonotone.

Proof. On the contrary, let us assume that for a f^o-pseudoconvex function there exist $\mathbf{x},\ \mathbf{y} \in R^n$, such that $f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) \geq 0$ and $f^o(\mathbf{y}, \mathbf{x} - \mathbf{y}) > 0$. Then, by the f^o-pseudoconvexity we get $f(\mathbf{x}) \leq f(\mathbf{y})$ and $f(\mathbf{y}) \leq f(\mathbf{x})$, or $f(\mathbf{x}) = f(\mathbf{y})$. Thus we have $f^o(\mathbf{y}; \mathbf{x} - \mathbf{y}) > 0$ and $f(\mathbf{x}) = f(\mathbf{y})$, which is impossible by lemma 4.24 □

4.5 Generalized quasiconvexity

In this section we discuss quasiconvexity. The section is based on source [4]. Let us first present the most common definition of quasiconvexity.

Definition 4.26. Function $f : R^n \to R$ is *quasiconvex*, if for all $\mathbf{x},\ \mathbf{y} \in R^n$ and $\lambda \in [0,1]$

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}.$$

An example of a quasiconvex function is x^3. Also, an increasing function is quasiconvex, because for an increasing function $f(\mathbf{x}) \leq f(\mathbf{z}) \leq f(\mathbf{y}),\ \forall\ \mathbf{x} < \mathbf{z} < \mathbf{y}$. Therefore for an increasing function $f(\mathbf{z}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}$. The same applies for decreasing functions.

When two random points are chosen from the function graph, the graph cannot rise above the point where the function value is higher between these points (see figure 11 for a non-quasiconvex function).

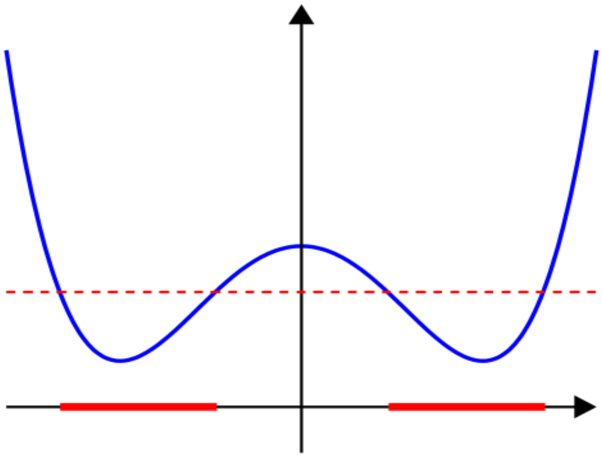


Figure 11: A non-quasiconvex function

A quasiconvex function does not need to be continuous. Lemma 4.22 implies that a f^o-pseudoconvex function is also quasiconvex. A quasiconvex function can also be defined geometrically with its level set. It was shown in Theorem 3.5 that the level set of a convex function is convex. Now we will show a similar result for quasiconvex functions.

Theorem 4.27. A function f is quasiconvex if and only if the level set S_α is convex for all $\alpha \in R$.

Proof. Let f be quasiconvex, $\mathbf{x},\ \mathbf{y} \in S_\alpha,\ \lambda \in [0,1]$ and $\alpha \in R$. Then

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\} \leq \max\{\alpha, \alpha\} = \alpha.$$

Therefore, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S_\alpha$.

On the other hand, let S_α be a convex set for all $\alpha \in R$. By choosing $\beta = \max\{f(\mathbf{x}), f(\mathbf{y})\}$, we have $\mathbf{x},\ \mathbf{y} \in S_\beta$. The convexity of S_β implies that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S_\beta$ for all $\lambda \in [0,1]$. It follows that

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \le \beta = \max\{f(\mathbf{x}), f(\mathbf{y})\}.$$

□

Next we will present the generalization of a quasiconvex function.

Definition 4.28. Function $f : R^n \to R$ is *f^o-quasiconvex*, if it is locally Lipschitz continuous and for every $\mathbf{x},\ \mathbf{y} \in R^n$

$$f(\mathbf{y}) \le f(\mathbf{x}) \Rightarrow f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) \le 0,$$

or equivalently

$$f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) > 0 \Rightarrow f(\mathbf{y}) > f(\mathbf{x}).$$

A quasiconvex function does not need to be continuous. However, a locally Lipschitz continuous and quasiconvex function can be expressed as follows.

Definition 4.29. Function $f : R^n \to R$ is *l-quasiconvex*, if it is locally Lipschitz continuous and for every $\mathbf{x},\ \mathbf{y} \in R^n$

$$f(\mathbf{y}) < f(\mathbf{x}) \Rightarrow f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) \le 0.$$

It follows from the definition of a f^o-quasiconvex function that a f^o-quasiconvex function is l-quasiconvex.

Theorem 4.30. A locally Lipschitz continuous and quasiconvex function $f : R^n \to R$ is l-quasiconvex.

Proof. Let f be locally Lipschitz continuous and quasiconvex. Let $\mathbf{x},\ \mathbf{z} \in R^n$ be such that $f(\mathbf{z}) < f(\mathbf{x})$. Since local Lipschitz continuity implies continuity, there exists $\epsilon > 0$ such that $f(\mathbf{z} + \mathbf{d}) < f(\mathbf{x} + \mathbf{d}), \forall\ \mathbf{d} \in B(\mathbf{0}, \epsilon)$. For the generalized directional derivative $f^o(\mathbf{x}; \mathbf{z} - \mathbf{x})$ we have

$$f^o(\mathbf{x}; \mathbf{z} - \mathbf{x}) = \limsup_{\substack{\mathbf{y} \to \mathbf{x} \\ t \to 0}} \frac{f(\mathbf{y} + t(\mathbf{z} - \mathbf{x})) - f(\mathbf{y})}{t}$$

$$= \limsup_{\mathbf{y}\to\mathbf{x}\; t\to 0} \frac{f(\mathbf{y}+t(\mathbf{z}-\mathbf{x}+\mathbf{y}-\mathbf{y}))-f(\mathbf{y})}{t}$$
$$= \limsup_{\mathbf{y}\to\mathbf{x}\; t\to 0} \frac{f((1-t)\mathbf{y}+t(\mathbf{z}+\mathbf{y}-\mathbf{x}))-f(\mathbf{y})}{t}.$$

When $t \in (0,1)$ and $\mathbf{y}-\mathbf{x} \in B(\mathbf{0},\epsilon)$, the quasiconvexity of f implies

$$\frac{f((1-t)\mathbf{y}+t(\mathbf{z}+\mathbf{y}-\mathbf{x}))-f(\mathbf{y})}{t}$$
$$\leq \frac{\max\{f(\mathbf{y}), f(\mathbf{z}+\mathbf{y}-\mathbf{x})\}-f(\mathbf{y}-\mathbf{x}+\mathbf{x})}{t}$$
$$= \frac{\max\{0, f(\mathbf{z}+\mathbf{y}-\mathbf{x})-f(\mathbf{x}+\mathbf{y}-\mathbf{x})\}}{t}.$$

When $t \to 0$ and $\mathbf{y} \to \mathbf{x}$, we get $f^o(\mathbf{x};\mathbf{z}-\mathbf{x}) \leq 0$. Thus, f is l-quasiconvex. □

Next we will show that an l-quasiconvex function is quasiconvex. First we need the following lemma.

Lemma 4.31. Let $\mathbf{x},\ \mathbf{y} \in R^n$. Let f be locally Lipschitz continuous in $[\mathbf{x},\mathbf{y}]$ so that $f(\mathbf{x}) < f(\mathbf{y})$. Then there exists $\mathbf{x}^* = \lambda\mathbf{x} + (1-\lambda)\mathbf{y},\ \lambda \in (0,1)$, for which $f(\mathbf{x}^*) > f(\mathbf{x})$ and $f^o(\mathbf{x}^*;\mathbf{y}-\mathbf{x}) > 0$.

Proof. Consider the nonempty set $A = S_{f(\mathbf{x})} \cap [\mathbf{x},\mathbf{y}]$. Since level sets of a continuous function are closed sets and $[\mathbf{x},\mathbf{y}]$ is compact, the set A is a compact set. Since the function $g(\mathbf{w}) = ||\mathbf{w}-\mathbf{y}||$ is continuous, it has a minimum on the set A. Let $\mathbf{z}$ be this minimum point. Then $\mathbf{z}$ is the nearest point to $\mathbf{y}$ on the set A and the continuity of f implies $f(\mathbf{z}) = f(\mathbf{x})$. Moreover, $f(\mathbf{x}) < f(\mathbf{y})$ implies that $\mathbf{z} \neq \mathbf{y}$. By the Mean-Value Theorem 4.20 there exists a point $\mathbf{z}^* \in (\mathbf{z},\mathbf{y})$ and a subgradient $\boldsymbol{\xi} \in \partial f(\mathbf{z}^*)$ for which

$$f(\mathbf{y}) - f(\mathbf{z}) = \boldsymbol{\xi}^T(\mathbf{y}-\mathbf{z}).$$

Then $f(\mathbf{z}) < f(\mathbf{y})$ implies

$$0 < f(\mathbf{y}) - f(\mathbf{z}) = \boldsymbol{\xi}^T(\mathbf{y}-\mathbf{z}) \leq f^o(\mathbf{z}^*;\mathbf{y}-\mathbf{z}) \leq f^o(\mathbf{z}^*;\mathbf{y}-\mathbf{x}).$$

The last inequation follows from the positive homogeneity (Theorem 4.4) of the generalized directional derivative and the inequality $||\mathbf{y}-\mathbf{z}|| \leq ||\mathbf{y}-\mathbf{x}||$, because $f^o(\mathbf{z}^*;\mathbf{y}-\mathbf{x}) = \lambda f(\mathbf{z}^*;\mathbf{y}-\mathbf{z})$, where $\lambda > 1$. The choice of $\mathbf{z}$ implies $f(\mathbf{z}) < f(\mathbf{z}^*)$, because $\mathbf{z}^* \in (\mathbf{z},\mathbf{y})$. By setting $\mathbf{x}^* = \mathbf{z}^*$, we get $f(\mathbf{x}) = f(\mathbf{z}) < f(\mathbf{x}^*)$. □

Theorem 4.32. If function $f : R^n \to R$ is l-quasiconvex, it is quasiconvex.

Proof. On the contrary assume that an l-quasiconvex function f is not quasiconvex. Then there exist $\mathbf{x}, \mathbf{y} \in R^n$ and $\overline{\lambda} \in (0, 1)$ such that $f(\overline{\mathbf{x}}) > \max\{f(\mathbf{x}), f(\mathbf{y})\}$, where $\overline{\mathbf{x}} = \overline{\lambda}\mathbf{x} + (1 - \overline{\lambda})\mathbf{y}$. Without a loss of generality we may assume that $f(\mathbf{x}) \geq f(\mathbf{y})$. By the previous lemma there exists $\widetilde{\mathbf{x}} \in (\mathbf{x}, \overline{\mathbf{x}})$, for which

$$f(\widetilde{\mathbf{x}}) > f(\mathbf{x}) \text{ and } f^o(\widetilde{\mathbf{x}}; \overline{\mathbf{x}} - \mathbf{x}) > 0.$$

Denote $\widetilde{\mathbf{x}} = \widetilde{\lambda}\mathbf{x} + (1 - \widetilde{\lambda})\mathbf{y}$, where $\widetilde{\lambda} \in (\overline{\lambda}, 1)$. From the definitions of points $\overline{\mathbf{x}}$ and $\widetilde{\mathbf{x}}$ we get

$$\overline{\mathbf{x}} - \mathbf{x} = (1 - \overline{\lambda})(\mathbf{y} - \mathbf{x}) \text{ and } \mathbf{y} - \widetilde{\mathbf{x}} = \widetilde{\lambda}(\mathbf{y} - \mathbf{x}).$$

Thus

$$\overline{\mathbf{x}} - \mathbf{x} = \frac{1 - \overline{\lambda}}{\widetilde{\lambda}}(\mathbf{y} - \widetilde{\mathbf{x}})$$

and

$$0 < f^o(\widetilde{\mathbf{x}}; \overline{\mathbf{x}} - \mathbf{x}) = \frac{1 - \overline{\lambda}}{\widetilde{\lambda}} f^o(\widetilde{\mathbf{x}}; \mathbf{y} - \widetilde{\mathbf{x}}).$$

Therefore $0 < f^o(\widetilde{\mathbf{x}}; \mathbf{y} - \widetilde{\mathbf{x}})$ and $f(\widetilde{\mathbf{x}}) > f(\mathbf{x}) \geq f(\mathbf{y})$, which contradicts the l-quasiconvexity of f. Hence f is quasiconvex. □

It follows from the two previous theorems that a locally Lipschitz continuous function is quasiconvex if and only if it is l-quasiconvex. Moreover, the l-quasiconvexity implies that a f^o-quasiconvex function is quasiconvex.

Quasimonotonity can be defined similarly to pseudomonotonity.

Definition 4.33. The generalized directional derivative f^o is called *quasimonotone*, if for all $\mathbf{x},\ \mathbf{y} \in R^n$

$$f^o(\mathbf{x}, \mathbf{y} - \mathbf{x}) > 0 \Rightarrow f^o(\mathbf{y}, \mathbf{x} - \mathbf{y}) \leq 0,$$

or equivalently

$$\min\{f^o(\mathbf{x}, \mathbf{y} - \mathbf{x}), f^o(\mathbf{y}, \mathbf{x} - \mathbf{y})\} \leq 0.$$

Strict quasimonotonity can be defined similarly to strict pseudomonotonity and it is equivalent to pseudomonotonity.

Theorem 4.34. If f^o is quasimonotone, then f is quasiconvex.

Proof. Let us, on the contrary, assume that f is not quasiconvex. Then there exist $\mathbf{x},\ \mathbf{y} \in R^n$ and $\overline{\lambda} \in (0,1)$ such that

$$f(\overline{\mathbf{x}}) > f(\mathbf{x}) \geq f(\mathbf{y}),$$

where $\overline{\mathbf{x}} = \mathbf{x} + \overline{\lambda}(\mathbf{y} - \mathbf{x})$. Then by the Mean-Value Theorem 4.20 there exist $\widehat{\mathbf{x}},\ \widetilde{\mathbf{x}} \in R^n$ such that

$$f(\overline{\mathbf{x}}) - f(\mathbf{y}) \in \partial f(\widehat{\mathbf{x}})^T(\overline{\mathbf{x}} - \mathbf{y})$$

and

$$f(\overline{\mathbf{x}}) - f(\mathbf{x}) \in \partial f(\widetilde{\mathbf{x}})^T(\overline{\mathbf{x}} - \mathbf{x}),$$

where

$$\widehat{\mathbf{x}} = \mathbf{x} + \widehat{\lambda}(\mathbf{y} - \mathbf{x}),\ \widetilde{\mathbf{x}} = \mathbf{x} + \widetilde{\lambda}(\mathbf{y} - \mathbf{x}),\ 0 < \widetilde{\lambda} < \overline{\lambda} < \widehat{\lambda} < 1.$$

Therefore, due to the definition of the Clarke subdifferential, there exist $\widehat{\boldsymbol{\xi}} \in \partial f(\widehat{\mathbf{x}})$ and $\widetilde{\boldsymbol{\xi}} \in \partial f(\widetilde{\mathbf{x}})$ such that

$$0 < f(\overline{\mathbf{x}}) - f(\mathbf{y}) = \widehat{\boldsymbol{\xi}}^T(\overline{\mathbf{x}} - \mathbf{y}) \leq f^o(\widehat{\mathbf{x}}; \overline{\mathbf{x}} - \mathbf{y}) = (1 - \overline{\lambda}) f^o(\widehat{\mathbf{x}}; \mathbf{x} - \mathbf{y})$$

$$0 < f(\overline{\mathbf{x}}) - f(\mathbf{x}) = \widetilde{\boldsymbol{\xi}}^T(\overline{\mathbf{x}} - \mathbf{x}) \leq f^o(\widetilde{\mathbf{x}}; \overline{\mathbf{x}} - \mathbf{x}) = \overline{\lambda} f^o(\widetilde{\mathbf{x}}; \mathbf{y} - \mathbf{x})$$

by the positive homogeneity of $\mathbf{d} \mapsto f^o(\mathbf{x}; \mathbf{d})$. It follows that

$$f^o(\widehat{\mathbf{x}}; \widetilde{\mathbf{x}} - \widehat{\mathbf{x}}) = (\widehat{\lambda} - \widetilde{\lambda}) f^o(\widehat{\mathbf{x}}; \mathbf{x} - \mathbf{y}) > 0$$

and

$$f^o(\widetilde{\mathbf{x}}; \widehat{\mathbf{x}} - \widetilde{\mathbf{x}}) = (\widehat{\lambda} - \widetilde{\lambda}) f^o(\widetilde{\mathbf{x}}; \mathbf{y} - \mathbf{x}) > 0,$$

which contradicts the quasimonotonicity. Thus, f is quasiconvex. □

Theorem 4.35. If function $f : R^n \to R$ is locally Lipschitz continuous and quasiconvex then the generalized directional derivative f^o is quasimonotone.

Proof. On the contrary, assume that f^o is not quasimonotone. Then there exist $\mathbf{x},\ \mathbf{y} \in R^n$ such that $f^o(\mathbf{x};\mathbf{y}-\mathbf{x}) > 0$ and $f^o(\mathbf{y};\mathbf{x}-\mathbf{y}) > 0$. Let

$$\delta = \min\{f^o(\mathbf{x};\mathbf{y}-\mathbf{x}),\ f^o(\mathbf{y};\mathbf{x}-\mathbf{y})\}.$$

Let $\epsilon_1 > 0$ be such that the local Lipschitz condition holds in the ball $B(\mathbf{x};\epsilon_1)$ with Lipschitz constant K_1. Correspondingly, let $\epsilon_2 > 0$ be such that the local Lipschitz condition holds in the ball $B(\mathbf{y};\epsilon_2)$ with the Lipschitz constant K_2. Let $K = \max\{K_1, K_2\}$ and $\epsilon = \min\{\frac{\delta}{4K}, \epsilon_1, \epsilon_2\}$. According to Lemma 4.23 there exists a sequence $\{\mathbf{z}_1^i\}$, such that f is differentiable, $\lim_{i\to\infty} \mathbf{z}_1^i = \mathbf{x}$ and an index $I \in N$ such that

$$f^{'}(\mathbf{z}_1^i;\mathbf{y}-\mathbf{x}) = \nabla f(\mathbf{z}_1^i)^T(\mathbf{y}-\mathbf{x}) \geq \frac{\delta}{2},$$

when $i \geq I$. Similarly, there exists a sequence $\{\mathbf{z}_2^j\}$, such that f is differentiable, $\lim_{j\to\infty} \mathbf{z}_2^j = \mathbf{y}$ and an index $J \in N$ such that

$$f^{'}(\mathbf{z}_2^j;\mathbf{x}-\mathbf{y}) = \nabla f(\mathbf{z}_2^j)^T(\mathbf{x}-\mathbf{y}) \geq \frac{\delta}{2},$$

when $j \geq J$. Let $\mathbf{z}_1 \in B(\mathbf{x};\epsilon) \cap \{\{\mathbf{z}_1^i\} \mid i \geq I\}$ and $\mathbf{z}_2 \in B(\mathbf{y};\epsilon) \cap \{\{\mathbf{z}_2^j\} \mid j \geq J\}$. Due to symmetry we may assume that $f(\mathbf{z}_1) \geq f(\mathbf{z}_2)$ without a loss of generality. According to part 1 of Theorem 4.4

$$|f^{'}(\mathbf{z}_1;\mathbf{z}_2-\mathbf{z}_1) - f^{'}(\mathbf{z}_1;\mathbf{y}-\mathbf{x})| \leq K\,||\mathbf{z}_2-\mathbf{z}_1-(\mathbf{y}-\mathbf{x})||$$

$$\leq K\,||\mathbf{x}-\mathbf{z}_1|| + K\,||\mathbf{z}_2-\mathbf{y}|| < 2K\frac{\delta}{4K} = \frac{\delta}{2}.$$

Since $f^{'}(\mathbf{z}_1;\mathbf{y}-\mathbf{x}) > \frac{\delta}{2}$ also $f^{'}(\mathbf{z}_1;\mathbf{z}_2-\mathbf{z}_1) > 0$. Thus, there exists $\lambda \in (0,1)$ such that

$$f(\mathbf{z}_1 + \lambda(\mathbf{z}_2-\mathbf{z}_1)) > f(\mathbf{z}_1) \geq f(\mathbf{z}_2),$$

which contradicts the quasiconvexity. □

It follows from the previous theorems that a function f is l-quasiconvex if and only if the generalized directional derivative f^o is quasimonotone. Additionally, if f is f^o-quasiconvex, f^o is quasimonotone.

The l-quasiconvexity of an f^o-quasiconvex function and Theorem 4.32 imply that it is quasiconvex. Next we will define subdifferential reqularity and show that for a subdifferentially regular function quasiconvexity and f^o-quasiconvexity are equivalent.

Definition 4.36. A function $f : R^n \to R$ is *subdifferentially regular* at a point $\mathbf{x} \in R^n$, if f is locally Lipschitz continuous at $\mathbf{x}$ and if for every directional vector $\mathbf{d} \in R^n$ the classical directional derivative exists and $f^{'}(\mathbf{x};\mathbf{d}) = f^o(\mathbf{x};\mathbf{d})$. Every convex or differentiable function is subdifferentially regular.

Theorem 4.37. If f is both quasiconvex and subdifferentially regular, then it is f^o-quasiconvex.

Proof. Due to the subdifferential regularity f is locally Lipschitz continuous. Suppose that $f(\mathbf{y}) \leq f(\mathbf{x})$. Then the subdifferential regularity and quasiconvexity imply that

$$f^o(\mathbf{x};\mathbf{y}-\mathbf{x}) = f^{'}(\mathbf{x};\mathbf{y}-\mathbf{x}) = \lim_{t\to 0}\frac{f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))-f(\mathbf{x})}{t}$$

$$= \lim_{t\to 0}\frac{f(t\mathbf{y}+(1-t)\mathbf{x})-f(\mathbf{x})}{t} \leq \lim_{t\to 0}\frac{f(\mathbf{x})-f(\mathbf{x})}{t} = 0.$$

Therefore, f is f^o-quasiconvex. □

It follows from Theorems 4.30, 4.32 and 4.37 that a subdifferentially regular l-quasiconvex function is f^o-quasiconvex. Additionally, a subdifferentially regular function f with a quasimonotone f^o is f^o-quasiconvex.

The following example shows the importance of subifferential regularity in the previous theorem.

Example 4.38. Let us define $f : R \to R$ (see Figure 12) such that

$$f(x) = \begin{cases} |x|, & x \in (-\infty,\ 1) \\ 1, & x \in [1,\ 2] \\ x-1, & x \in (2,\ \infty) \end{cases}.$$

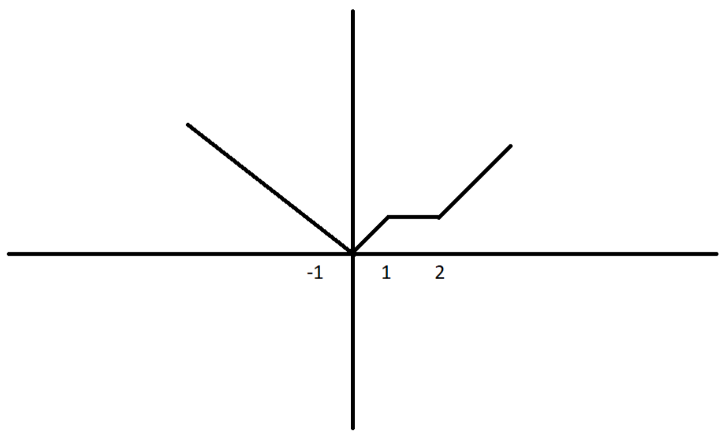


Figure 12: A non-subdifferentially regular function

Function f is clearly locally Lipschitz continuous and quasiconvex. However, by taking $x = 1$ and $y = 2$ we have $f^o(x; y - x) = f^o(1; 1) = 1 > 0$, but $f(y) = f(2) = 1 \leq 1 = f(1) = f(x)$, and thus f is not f^o-quasiconvex. Note that f is not subdifferentially regular, since $f^{'}(1; 1) = 0 \neq 1 = f^o(1; 1)$. Furthermore, f is not f^o-pseudoconvex, since $0 \in \partial f(1) = [0, 1]$ although $x = 1$ is not a global minimum.

Definition 4.39. A function $f : R^n \to R$ is said to satisfy *nonconstancy property*, or NC-property, if there exists no line segment $[\mathbf{a}, \mathbf{b}]$ along which f is constant.

Definition 4.40. A function $f : R^n \to R$ is said to satisfy *generalized nonconstancy property*, or GNC-property, if there exist no $\mathbf{x}$ and $\epsilon > 0$ such that $f(\mathbf{y}) = f(\mathbf{x})\ \forall\ \mathbf{y} \in B(\mathbf{x}; \epsilon)$.

If $f : R^n \to R$ satisfies NC-property it satisfies GNC-property.

Example 4.41. Let us present a function which is subdifferentially regular but does not satisfy NC-property. Define function g_1 as

$$g_1(x) = \begin{cases} (x+1)^2, & x \leq -1 \\ 0, & -1 \leq x \leq 1 \\ (x-1)^2, & x \geq 1 \end{cases}.$$

On the other hand, the function

$$g_2(x) = \begin{cases} 2x, & x \le 0 \\ \frac{1}{2}x, & x \ge 0 \end{cases}.$$

possesses the NC-property but is not subdifferentially regular since $g_2^o(0;1) = 2 \neq \frac{1}{2} = g_2'(0;1)$.

Theorem 4.42. Let f be an l-quasiconvex function that possesses the GNC-property. Then it is f^o-quasiconvex.

Proof. If f is not f^o-quasiconvex then there exist $\mathbf{x}$, $\mathbf{y} \in R^n$ such that $f^o(\mathbf{x}; \mathbf{y} - \mathbf{y}) > 0$ but $f(\mathbf{x}) = f(\mathbf{y})$.

Suppose that $f(\mathbf{x})$ is the global minimum value of f. By the similar deductions used in Lemma 4.24 there exist $\mathbf{z} \in R^n$ and $0 < \mu < 1$ such that

$$f(\mu\mathbf{z} + (1-\mu)\mathbf{y}) > f(\mathbf{z}). \tag{17}$$

Since $f(\mathbf{x}) = f(\mathbf{y})$ is the global minimum value we also have $f(\mathbf{z}) \ge f(\mathbf{y})$. This and (17) contradict with the definition of quasiconvexity.

Suppose then that $f(\mathbf{x})$ is not the global minimum value of f. By continuity of f^o as a function of $\mathbf{d}$ there exists $\epsilon > 0$ such that $f^o(\mathbf{x}; \mathbf{d}) > 0$, $\forall\, \mathbf{d} \in B(\mathbf{y} - \mathbf{x}; \epsilon)$. Then, by the l-quasiconvexity of f there exists $\delta > 0$ such that $f(\overline{\mathbf{y}}) \ge f(\mathbf{x}) = f(\mathbf{y})$, $\forall\, \overline{\mathbf{y}} \in B(\mathbf{y}; \delta)$. Since there exists $\mathbf{z}$ such that $f(\mathbf{z}) < f(\mathbf{x})$, the continuity of f implies there exists $\gamma > 0$ such that $f(\overline{\mathbf{z}}) < f(\mathbf{x})$, $\forall\, \overline{\mathbf{z}} \in B(\mathbf{z}; \gamma)$. Furthermore, there exist $\mathbf{w} \in R^n$ and $r > 0$ such that

$$B(\mathbf{w}; r) \subset B(\mathbf{y}; \delta) \cap \operatorname{conv}\{\{\mathbf{y}\}, B(\mathbf{z}; \gamma)\}. \tag{18}$$

Since $B(\mathbf{w}; r) \subset B(\mathbf{y}; \delta)$, we have $f(\overline{\mathbf{w}}) \ge f(\mathbf{y})$, $\forall\, \overline{\mathbf{w}} \in B(\mathbf{w}; r)$. By inclusion (12), for any $\overline{\mathbf{w}} \in B(\mathbf{w}; r)$ there exist $\overline{\mathbf{z}} \in B(\mathbf{z}; \gamma)$ and $0 < \lambda < 1$ such that $\overline{\mathbf{w}} = \lambda\mathbf{y} + (1-\lambda)\overline{\mathbf{z}}$. By the quasiconvexity of f we have $f(\overline{\mathbf{w}}) \le \max\{f(\mathbf{y}), f(\overline{\mathbf{z}})\} = f(\mathbf{y})$. Hence, $f(\overline{\mathbf{w}}) = f(\mathbf{y})$, $\forall\, \overline{\mathbf{w}} \in B(\mathbf{w}; r)$ contradicting the generalized NC-property assumption. □

It follows that if f^o is quasimonotone and f possesses the GNC-property, then f is f^o-quasiconvex.

Next we show that the relationship between pseudo- and quasiconvexity is true also for f^0-pseudo- and f^0-quasiconvexity.

Theorem 4.43. An f^o-pseudoconvex function is f^o-quasiconvex.

Proof. On the contrary, assume that an f^o-pseudoconvex function f is not f^o-quasiconvex. Then there exist points $\mathbf{x},\ \mathbf{y} \in R^n$ such that $f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) > 0$ and $f(\mathbf{x}) = f(\mathbf{y})$. According to Lemma 4.24 this is impossible for an f^o-pseudoconvex function. Therefore f is f^o-quasiconvex. □

The following example shows that the result in the theorem cannot be reversed.

Example 4.44. Define $f : R \to R$ such that $f(x) = x^3$. Clearly f is quasiconvex and as a smooth function also subdifferentially regular. Thus, by Theorem 4.37 it is f^o-quasiconvex. However, by taking $x = 0$ and $y = -1$ we have $f^o(x; y - x) = f^o(0; -1) = 0$, but $f(y) = f(-1) = -1 < 0 = f(0) = f(x)$ and thus, by Lemma 4.16, f is not f^o-pseudoconvex.

If a quasiconvex function f is continuously differentiable and the condition

$$\nabla f(\mathbf{x}) = \mathbf{0} \Longleftrightarrow \mathbf{x} \text{ is a global minimum}$$

holds, f is pseudoconvex. A similar result can be shown for generalized convexities.

Lemma 4.45. Let f be l-quasiconvex and $\mathbf{x},\ \mathbf{y} \in R^n$. If $f(\mathbf{y}) < f(\mathbf{x})$ and $\mathbf{0} \notin \partial f(\mathbf{x})$, then $f^o(\mathbf{x}, \mathbf{y} - \mathbf{x}) < 0$.

Proof. Suppose that $f(\mathbf{y}) < f(\mathbf{x})$ and $\mathbf{0} \notin \partial f(\mathbf{x})$. By continuity of f there exists $r > 0$ such that $f(\mathbf{z}) < f(\mathbf{x}), \forall\ \mathbf{z} \in \text{cl}(B(\mathbf{y}; r))$.

Let $\boldsymbol{\xi} \in \partial f(\mathbf{x})$ be arbitrary. Since $\mathbf{0} \notin \partial f(\mathbf{x})$, we may define $\widehat{\mathbf{y}} = y + \frac{\boldsymbol{\xi}}{||\boldsymbol{\xi}||} r$. The l-quasiconvexity of f and the inequality $f(\widehat{\mathbf{y}}) < f(\mathbf{x})$ imply that

$$f^o(\mathbf{x}; \widehat{\mathbf{y}} - \mathbf{x}) \leq 0. \tag{19}$$

By the basic properties of the Clarke generalized directional derivative, the above inequality implies $\boldsymbol{\xi}^T(\widehat{\mathbf{y}} - \mathbf{x}) \leq 0$. Thus,

$$\boldsymbol{\xi}^T(\mathbf{y} - \mathbf{x}) = \boldsymbol{\xi}^T(\widehat{\mathbf{y}} - \frac{\boldsymbol{\xi}}{||\boldsymbol{\xi}||} r - \mathbf{x}) = -r\,||\boldsymbol{\xi}|| + \boldsymbol{\xi}^T(\widehat{\mathbf{y}} - \mathbf{x}) \leq -r\,||\boldsymbol{\xi}||\,.$$

Since $\mathbf{0} \notin \partial f(\mathbf{x})$, we have $-r\,||\boldsymbol{\xi}|| < 0$. Thus

$$f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) = \max_{\boldsymbol{\xi} \in \partial f(\mathbf{x})} \boldsymbol{\xi}^T(\mathbf{y} - \mathbf{x}) \leq \max_{\boldsymbol{\xi} \in \partial f(\mathbf{x})} -r\,||\boldsymbol{\xi}|| < 0,$$

proving the lemma. □

Theorem 4.46. If f is l-quasiconvex and $\mathbf{0} \in \partial f(\mathbf{x})$ implies $\mathbf{x} \in R^n$ is a global minimum of f, then f is f^o-pseudoconvex.

Proof. Let $\mathbf{x},\ \mathbf{y} \in R^n$ be such that $f(\mathbf{x}) > f(\mathbf{y})$. By assumption $\mathbf{0} \notin \partial f(\mathbf{x})$. Then, Lemma 4.45 implies $f^o(\mathbf{x}; \mathbf{y} - \mathbf{x}) < 0$. □

Next we will show a result concerning the subdifferential of composite functions, which we will need later.

Theorem 4.47. [18] Let $f : R^n \to R$ be such that $f = g \circ H$, where $H : R^n \to R^m$ is locally Lipschitz continuous at $\mathbf{x}$ and $g : R^m \to R$ is locally Lipschitz continuous at $g(\mathbf{x}) \in R$. Then f is locally Lipschitz continuous at $\mathbf{x}$ and

$$\partial f(\mathbf{x}) \subseteq \operatorname{conv}\{\partial H(\mathbf{x})^T \partial g(H(\mathbf{x}))\}.$$

Proof. See [18]. □

5 Semismooth and well-behaved functions

In this chapter we discuss different kinds of semismooth functions. We will also define well-behaved functions and show their relation with generalized convexities.

5.1 Well-behaved functions

In this section we will define a well-behaved generalized directional derivative function and discuss its relation with different convexities.

Definition 5.1. [13] A generalized directional derivative f^o of function f is *well-behaved* if $f^o(\mathbf{x};\mathbf{d}) > 0$ implies that there exists $t \to 0$ such that $f(\mathbf{x}+t\mathbf{d}) > f(\mathbf{x})$.

If f is smooth, then the directional derivative is well-behaved. Also, the generalized directional derivative for a subifferentiably regular function is always well-behaved. Next we will show that for a f^o- quasiconvex function the generalized directional derivative is well-behaved.

Theorem 5.2. The generalized directional derivative for a f^o-quasiconvex function is well-behaved.

Proof. By the definition of a f^o-quasiconvex function, if $f^o(\mathbf{x};\mathbf{y}-\mathbf{x}) > 0$, then $f(\mathbf{y}) > f(\mathbf{x}),\ \forall\ \mathbf{x},\mathbf{y} \in R^n$. We can substitute $\mathbf{y} = \mathbf{x}+t\mathbf{d}, t > 0$, which gives us the definition of a well-behaved generalized directional derivative. □

Since an f^o-pseudoconvex function is f^o-quasiconvex, this result also holds for f^o-pseudoconvex functions.

Next we will show an example of a function whose generalized directional derivative is well-behaved, but that is not subdifferentially regular.

Example 5.3. Let us define function f as

$$f(x) = \begin{cases} |x|, & x \in (-\infty,\ 1) \\ \frac{1}{2}x, & x \in [1,\ 2] \\ x - \frac{1}{2}, & x \in (2,\ \infty) \end{cases} .$$

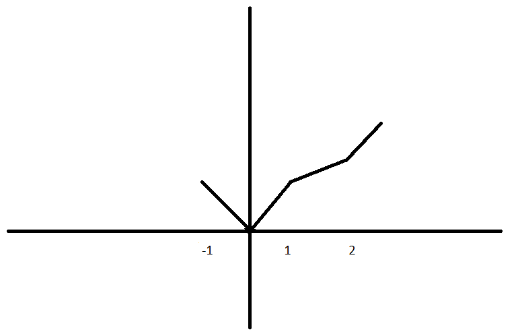


Figure 13: A function whose generalized directional derivative is well-behaved

For $x = 1$ and $x = 2$, $f^o(x; 1) = 1$ and f is increasing in the direction $d = 1$ at both points. However, the function is not subdifferentially regular, because $f'(1; 1) = 1/2 < 1 = f^o(1; 1)$, although it is subdifferentially regular at $x = 2$, because $f^o(2; 1) = f'(2; 1) = 1$. The value of f' increases in this cornerpoint. The generalized directional derivative of f is well-behaved, because f' does not change from positive to non-positive in this direction for any nonsmooth points. Compare with the function from Example 4.38, where the function is not well-behaved at $x = 1$.

As stated previously, for a subdifferentially regular function f^o is well-behaved. According to Theorem 4.37, a subdifferentially regular quasiconvex function is f^o-quasiconvex. The next theorem will show that this condition can be relaxed.

Theorem 5.4. Let $f : R^n \to R$ be quasiconvex and locally Lipschitz continuous and let f^o be well-behaved. Then f is f^o-quasiconvex.

Proof. On the contrary, let us assume that f is not f^o-quasiconvex. Then there exist $\mathbf{x}$ and $\mathbf{y} \in R^n$ such that $f^o(\mathbf{x}, \mathbf{y} - \mathbf{x}) > 0$ and $f(\mathbf{x}) \geq f(\mathbf{y})$. Since f^o is well-behaved, there exists $t \to 0$ such that $f(\mathbf{x} + t\mathbf{d}) > f(\mathbf{x})$, where $\mathbf{d} = \mathbf{y} - \mathbf{x}$. Then $f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) = f((1-t)\mathbf{x} + t\mathbf{y}) > f(\mathbf{x}) = \max\{f(\mathbf{x}), f(\mathbf{y})\}$. Therefore, f is not quasiconvex. $\square$

Since the generalized directional derivative of a f^o-quasiconvex function is always well-behaved, it follows that a quasiconvex function is f^o-quasiconvex if and only if its generalized directional derivative is well-behaved. Since an l-quasiconvex function is quasiconvex, this result applies for l-quasiconvex functions as well. It also follows that if a quasiconvex function has the GNC-property, its generalized directional derivative is well-behaved.

5.2 Semismooth functions

Next we will define semismoothness and examine its relations with generalized convexities.

Definition 5.5. [6,9,10] Let $F : R^n \to R^m$ be locally Lipschitz continuous at $\mathbf{x}$. It is *semismooth* at $\mathbf{x}$ if the limit

$$\lim_{A \in \partial F(\mathbf{x}+t\mathbf{d}'),\ \mathbf{d}' \to \mathbf{d},\ t \to 0} \{A\mathbf{d}'\} \tag{20}$$

exists for all $\mathbf{d} \in R^n$. It is *weakly semismooth* at $\mathbf{x}$ if the limit

$$\lim_{A \in \partial f(\mathbf{x}+t\mathbf{d}),\ t \to 0} \{A\mathbf{d}\} \tag{21}$$

exists for all $\mathbf{d} \in R^n$.

Clearly a semismooth function is also weakly semismooth. An equivalent definition for weak semismoothness is as follows [9]:

Function $F : R^n \to R^m$ is weakly semismooth at $\mathbf{x}$ if

1. F is locally Lipschitz continuous at $\mathbf{x}$ and
2. for each $\mathbf{d} \in R^n$ and for any sequences $\{t_k\} \subset R_+$, $\{\boldsymbol{\theta}_k\} \subset R^n$ and $\{A_k\} \subset R^{m\times n}$ such that $\{t_k\} \to 0$, $\{\boldsymbol{\theta}_k/t_k\} \to \mathbf{0} \in R^n$ and $A_k \in \partial F(\mathbf{x} + t_k\mathbf{d} + \boldsymbol{\theta}_k)$ the sequence $\{A_k\mathbf{d}\}$ has exactly one accumulation point.

Function $F : R^n \to R^m$ is *strongly semismooth* at $\overline{\mathbf{x}}$ if it is B-differentiable at $\overline{\mathbf{x}}$ and the following limit holds:

$$\lim_{\overline{\mathbf{x}} \neq \mathbf{x} \to \overline{\mathbf{x}}} \frac{||F'(\mathbf{x}; \mathbf{x} - \overline{\mathbf{x}}) - F'(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}})||}{||\mathbf{x} - \overline{\mathbf{x}}||} = 0.$$

If the above requirement is strengthened to

$$\limsup_{\overline{\mathbf{x}} \neq \mathbf{x} \to \overline{\mathbf{x}}} \frac{||F'(\mathbf{x}; \mathbf{x} - \overline{\mathbf{x}}) - F'(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}})||}{||\mathbf{x} - \overline{\mathbf{x}}||^2} < \infty,$$

we say that F is *strictly semismooth* at $\overline{\mathbf{x}}$.

Next we will show a connection between continuously differentiable and convex functions and semismooth functions, but first we will show a necessary theorem.

Theorem 5.6. [10] Let $G : R^n \to R^m$ be continuously differentiable in a neighborhood Ω of $\overline{\mathbf{x}}$. Then a nondecreasing function $\delta : (0, \infty) \to [0, \infty)$ with

$$\lim_{t \to 0} \delta(t) = 0$$

exists such that

$$||G(\mathbf{x}) + \partial G(\mathbf{x})(\mathbf{z} - \mathbf{x}) - G(\mathbf{z})|| \leq ||\mathbf{x} - \mathbf{z}|| \, \delta(||\mathbf{x} - \mathbf{z}||)$$

for all $\mathbf{z}$ and $\mathbf{x} \in \Omega$. Here $\partial G(\mathbf{x})$ is the generalized Jacobian matrix of G in $\mathbf{x}$ (see Definition 4.9). Furthermore, if $\partial G(\cdot)$ is Lipschitz continuous in a neighborhood of $\overline{\mathbf{x}}$, then a subneighborhood $\Omega' \subseteq \Omega$ of $\overline{\mathbf{x}}$ and a positive constant L' exist so that

$$||G(\mathbf{x}) + \partial G(\mathbf{x})(\mathbf{z} - \mathbf{x}) - G(\mathbf{z})|| \leq L' \, ||\mathbf{x} - \mathbf{z}||^2$$

for all $\mathbf{z}$ and $\mathbf{x}$ in Ω'.

Proof. By the Mean-Value Theorem 4.20 for real-valued functions, we can write, for a given $\mathbf{z} \in \Omega$ and any $\mathbf{x} \in \Omega$:

$$G(\mathbf{x}) = G(\mathbf{z}) + \sum_{i=1}^{m} \alpha_i(\mathbf{x}) \partial G(\mathbf{y}^i(\mathbf{x}))(\mathbf{x} - \mathbf{z}),$$

where $\alpha_i(\mathbf{x})$ and $\mathbf{y}^i(\mathbf{x})$ are such that

$$\alpha_i(\mathbf{x}) \geq 0 \; \forall \; i, \; \sum_{i=1}^{m} \alpha_i(\mathbf{x}) = 1, \; \mathbf{y}^i(\mathbf{x}) \in [\mathbf{x}, \mathbf{z}] \; \forall \; i. \qquad (22)$$

But then we can write

$$||G(\mathbf{x}) + \partial G(\mathbf{x})(\mathbf{z} - \mathbf{x}) - G(\mathbf{z})|| =$$

$$\left|\left|\left[\sum_{i=1}^{m} \alpha_i(\mathbf{x}) G(\mathbf{y}^i(\mathbf{x})) - \partial G(\mathbf{x})\right](\mathbf{x} - \mathbf{z})\right|\right|$$

$$\leq \sum_{i=1}^{m} \alpha_i(\mathbf{x}) \left|\left|\partial G(\mathbf{y}^i(\mathbf{x})) - \partial G(\mathbf{x})\right|\right| ||\mathbf{x} - \mathbf{z}||,$$

where in the last inequality we have used the first two relations in (22). Furthermore, by using the last relation in (22) and the continuity of ∂G we see that, for every i,

$$\lim_{\mathbf{x}\to\mathbf{z}} \left|\left|\partial G(\mathbf{y}^i(\mathbf{x})) - \partial G(\mathbf{x})\right|\right| = 0.$$

It is then easy to see that for the first assertion of the proposition it suffices to take

$$\delta(t) = \sup_{\mathbf{x},\mathbf{y}\in\Omega, ||\mathbf{x}-\mathbf{y}||\leq t} ||\partial G(\mathbf{y}) - \partial G(\mathbf{x})||.$$

By the continuity of ∂G and the boundedness of Ω, $\delta(t)$ is clearly finite and goes the zero when t tends to zero. If ∂G is locally Lipschitz continuous at $\overline{\mathbf{x}}$, with Lipschitz constant L, we can write, by possibly restricting $\mathbf{z}$ and $\mathbf{x}$ to a suitable subneighborhood $\Omega' \subseteq \Omega$,

$$\left|\left|\partial G(\mathbf{y}^i(\mathbf{x})) - \partial G(\mathbf{x})\right|\right| \leq L \left|\left|\mathbf{y}^i(\mathbf{x}) - \mathbf{x}\right|\right| \leq L\, ||\mathbf{x} - \mathbf{z}||,$$

which yields

$$||G(\mathbf{x}) + \partial G(\mathbf{x})(\mathbf{z} - \mathbf{x}) - G(\mathbf{z})|| \leq L\, ||\mathbf{x} - \mathbf{z}||^2$$

as we have asserted.

□

Theorem 5.7. [10] Let $f : \Omega \subseteq R^n \to R$, with Ω open, and a point $\overline{\mathbf{x}}$ belonging to Ω be given.

1. If f is continuously differentiable in a neighborhood of $\overline{\mathbf{x}}$, then f is strongly semismooth at $\overline{\mathbf{x}}$.

2. If f is continuously differentiable with a Lipschitz continuous gradient in a neighborhood on $\overline{\mathbf{x}}$, then f is strictly semismooth at $\overline{\mathbf{x}}$.

3. If f is convex on a neighborhood of $\overline{\mathbf{x}}$, then f is strongly semismooth at $\overline{\mathbf{x}}$.

The proof for this theorem will be shown after Theorem 5.19.
Next we will show a result for the maximum function of weakly semismooth functions.

Theorem 5.8. Let $f_i : R^n \to R$ for $i = \{1, \ldots, m\}$ be weakly semismooth at $\mathbf{x} \in R^n$. Then $g(\mathbf{x}) = \max\{f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also weakly semismooth at $\mathbf{x}$.

Proof. Let us define set J as a subset of $I = \{1, \ldots, m\}$ which contains at least two elements of I. First let us assume that $\mathbf{x}$ is an element of R^n such that $g(\mathbf{y}) = f_i(\mathbf{y}),\ \forall \mathbf{y} \in B(\mathbf{x}, \delta),\ i \in I$. Then g is weakly semismooth at $\mathbf{x}$ due to the weak semismoothness of f_i. Then let us assume that $\mathbf{x}$ is an element of R^n such that $g(\mathbf{x}) = f_j(\mathbf{x}),\ \forall j \in J$. Let D_j be the set of directional vectors $\mathbf{d}_j$ for which $g(\mathbf{x}+t\mathbf{d}_j) = f_j(\mathbf{x}+t\mathbf{d}_j),\ t \to 0,\ j \in J$. Then $\lim_{\boldsymbol{\xi} \in \partial f(\mathbf{x}+t\mathbf{d}_j), t \to 0} \boldsymbol{\xi}^T \mathbf{d}_j$ exists for all $\mathbf{d}_j \in D_j$ by the weak semismoothness of f_j. This applies for every $j \in J$. Therefore g is weakly semismooth in $\mathbf{x}$.

□

Next we will show a relation between weak semismoothness and the directional derivative, but first we present the following generalization of the Mean-Value Theorem 4.20.

Theorem 5.9. [6] Let $F : R^n \to R^m$ be Lipschitz continuous on an open set S in R^n and $\mathbf{x}, \mathbf{y}$ be two points in S. Then

$$F(\mathbf{y}) - F(\mathbf{x}) \in \text{conv}(\partial F([\mathbf{x}, \mathbf{y}])(\mathbf{y} - \mathbf{x})). \tag{23}$$

Theorem 5.10. [6] Let $F : R^n \to R^m$ be a weakly semismooth function. Then the directional derivative

$$F^{'}(\mathbf{x}; \mathbf{d}) = \lim_{t \to 0} \frac{F(\mathbf{x} + t\mathbf{d}) - F(\mathbf{x})}{t}$$

exists for all $\mathbf{d} \in R^n$ and

$$F^{'}(\mathbf{x};\mathbf{d}) = \lim_{A\in\partial F(\mathbf{x}+t\mathbf{d}),\ t\to 0}\{A\mathbf{d}\}.$$

Proof. The difference quotient $\frac{F(\mathbf{x}+t\mathbf{d})-F(\mathbf{x})}{t}$ is bounded due to the local Lipschitz continuity of weakly semismooth functions. So, there exists a sequence $t_i \to 0$ and some $\mathbf{l} \in R^m$ such that

$$\frac{F(\mathbf{x}+t_i\mathbf{d})-F(\mathbf{x})}{t_i} \to \mathbf{l}.$$

It suffices to show that $\mathbf{l}$ equals the limit in (21). By Theorem 5.9

$$\frac{F(\mathbf{x}+t_i\mathbf{d})-F(\mathbf{x})}{t_i} \in \operatorname{conv}(\partial F([\mathbf{x},\mathbf{x}+t_i\mathbf{d}])\mathbf{d}).$$

By Carathéodory's Theorem 2.21 there exist numbers $t_i^{(k)} \in [0,t_i]$, coefficients of a convex combination $\lambda_i^{(k)}$ and matrices $A_i^{(k)} \in \partial F(\mathbf{x}_i + t_i^{(k)}\mathbf{d})$ for $k = 1,2,...,m+1$, such that

$$\frac{F(\mathbf{x}+t_i\mathbf{d})-F(\mathbf{x})}{t_i} = \sum_{k=1}^{m+1}\lambda_i^{(k)}A_i^{(k)}\mathbf{d}, \qquad \sum_{k=1}^{m+1}\lambda_i^{(k)} = 1.$$

By passing to a subsequence, if necessary, we can assume that $\lambda_i^{(k)} \to \lambda^{(k)}$ as $i \to \infty$. Clearly, $\lambda^{(k)} \in [0,1]$ and $\sum_{k=1}^{m+1}\lambda^{(k)} = 1$. Then

$$\mathbf{l} = \lim_{i\to\infty}\{\sum_{k=1}^{m+1}\lambda_i^{(k)}A_i^{(k)}\mathbf{d}\} = \sum_{k=1}^{m+1}\lim_{i\to\infty}\lambda_i^{(k)}\lim_{i\to\infty}\{A_i^{(k)}\mathbf{d}\}$$

$$\sum_{k=1}^{m+1}\lambda^{(k)}\lim_{A\in\partial F(\mathbf{x}+t\mathbf{d}),\ t\to 0}\{A\mathbf{d}\} = \lim_{A\in\partial f(\mathbf{x}+t\mathbf{d}),\ t\to 0}\{A\mathbf{d}\}$$

as required. □

The following example shows that semismoothness does not imply f^o-pseudoconvexity.

Example 5.11. Let us consider the function from Example 4.38. The derivative of the function is $f^{'}(x) = 0,\ x \in [1,2]$, which is also the subgradient ξ of f on this interval. When $x = 1$ and $d = 1$, $\lim_{\xi\in\partial f(x+td),\ t\to 0}\xi d = 0$. On the other hand, the derivative is $f'(x) = 1,\ x \in [0,1]$. When $x = 1$ and $d = -1$, $\lim_{\xi \in \partial f(x+td),\ t \to 0} \xi d = -1$. The function has limits to ξd for both directions, so it is weakly semismooth. However, it is not f^o-pseudoconvex.

Next we will show that f^o-pseudoconvexity does not imply semismoothness.

Example 5.12. Let us present a function

$$f(x) = \begin{cases} x\sin(\ln(x)) + x\cos(\ln(x)) + 5x, & x > 0 \\ x^2 + x, & x \le 0 \end{cases}.$$

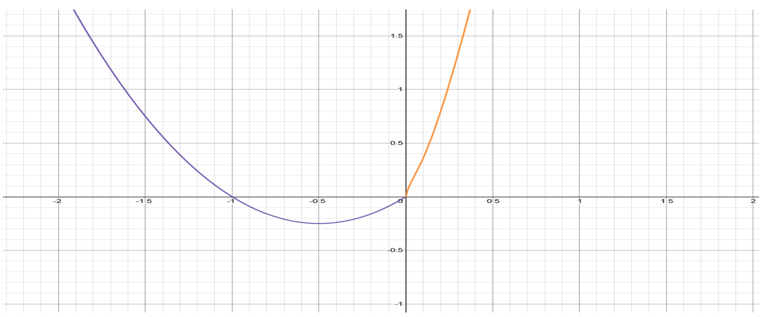


Figure 14: A non-weakly semismooth function

The derivative of f is $f'(x) = 2\cos(\ln(x)) + 5,\ x > 0$. When x approaches 0, $\ln(x)$ approaches $-\infty$, which causes $2\cos(\ln(x))$ to jump at the interval $[-2, 2]$ and $f'(x)$ at the interval $[3, 7]$. The limit $\lim_{\xi \in \partial f(x+td),\ t \to 0} \xi d$ does not approach a solid value, when $x = 0$ and $d = 1$, so f is not weakly semismooth. It is, however, locally Lipschitz continuous and f^o-pseudoconvex. It also has a generalized directional derivative at $x = 0$, $f^o(0; 1) = 7$.

Because the generalized directional derivative of the function in example 5.12 is well-behaved, but the function is not weakly semismooth, a well-behaved generalized directional derivative does not imply weak semismoothness. On the other hand, since the function in Example 4.38 is semismooth, but its generalized directional derivative is not well-behaved, semismoothness does not imply a well-behaved generalized directional derivative either.

Next we will show that the result from Theorem 5.10 can be reversed.

Theorem 5.13. If a locally Lipschitz continuous function $F : R^n \to R^m$ is directionally differentiable at $\mathbf{x}$, it is weakly semismooth at $\mathbf{x}$.

Proof. The result follows from the proof of Theorem 5.10, because if $F'(\mathbf{x}, \mathbf{d})$ exists for every $\mathbf{d} \in R^n$ in $\mathbf{x}$, then it can be written as $\lim_{A \in \partial F(\mathbf{x}+t\mathbf{d}),\ t \to 0} \{A\mathbf{d}\}$. □

It follows that B-differentiability is equivalent with weak semismoothness, since a function is weakly semismooth if and only if it is locally Lipschitz continuous and directionally differentiable. It also follows that a strongly semismooth function is also weakly semismooth, since it is required to be B-differentiable. Moreover, the results in Theorem 5.7 also hold for weakly semismooth functions.

Next we show an example of function that is weakly semismooth but not semismooth.

Example 5.14. Let us consider the function in the following picture.

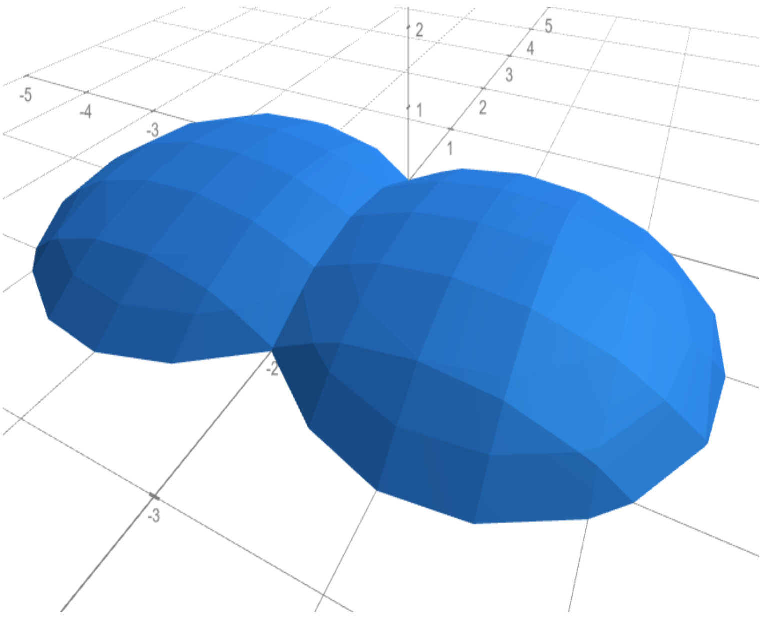


Figure 15: Apple function

The function is

$$\begin{cases} (z+1)^2 = 3.25 - (x-1.1)^2 - (y+1)^2, & x \geq 0, z \geq 0 \\ (z+1)^2 = 3.25 - (x+1.1)^2 - (y+1)^2, & x \leq 0, z \geq 0 \end{cases}$$

inside the apple-shaped area

$$\begin{cases} (x-1.1)^2 + (y+1)^2 \leq 2.25, & x \geq 0 \\ (x+1.1)^2 + (y+1)^2 \leq 2.25, & x \leq 0 \end{cases}$$

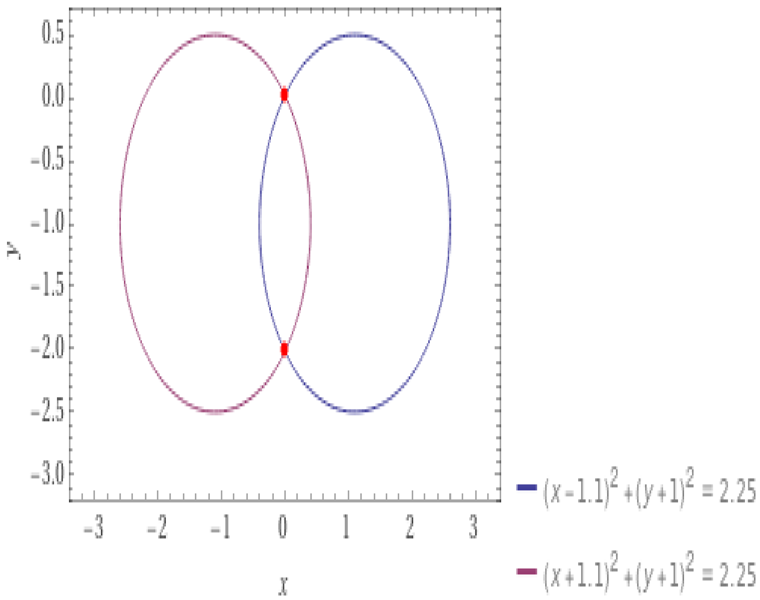


Figure 16: Apple function

and $f(x, y) = 0$ outside. The function is locally Lipschitz continuous and directionally differentiable, and therefore weakly semismooth. The generalized directional derivative in the origo is 0, when $d = (0, 1)$, and not 0, otherwise. Therefore, when the directional derivative d approaches $(0, 1)$ in the origo, the limit at (20) does not exist. Therefore, the function is not semismooth.

Next we will show that subdifferential regularity and NC-property cannot be substituted with semismoothness.

Example 5.15. The function in Example 4.38 is quasiconvex and locally Lipschitz continuous, so it is also l-quasiconvex. The function is not subdifferentially regular nor does satisfy NC-property, but it is semismooth. It was shown in the example that the function is not f^o-quasiconvex. Therefore subdifferential regularity and NC-property cannot be substituted with semismoothness in Theorem 4.37 and Theorem 4.42.

In Theorem 5.19 we will prove a result concerning the semismoothness of composite functions. For this we will need the following results.

Theorem 5.16. [10] Let D and D' be open sets in R^n and R^m respectively. Let $\Phi : D \rightarrow R^m$ and $\Psi : D^{'} \rightarrow R^p$ be B-differentiable at $\mathbf{x} \in D$ and $\Phi(\mathbf{x}) \in D^{'}$ respectively. Suppose that $\Phi(D) \subseteq D^{'}$. The following statements hold:

1. The composite map $\Gamma = \Psi \circ \Phi : D \rightarrow R^p$ is B-differentiable at $\mathbf{x}$, moreover
$$\Gamma^{'}(\mathbf{x}; \mathbf{d}) = \Psi^{'}(\Phi(\mathbf{x}); \Phi^{'}(\mathbf{x}; \mathbf{d})), \ \forall \mathbf{d} \in R^n.$$
2. If Ψ is strongly F-differentiable at $\Phi(\mathbf{x})$ and Φ has a strong B-derivative at $\mathbf{x}$, then Γ has a strong B-derivative at $\mathbf{x}$.

Proof. We only prove the second part, since the following proof is applicable to the first part with a minor modification. It suffices to show that (11) holds:
$$\lim_{\mathbf{y}^1 \neq \mathbf{y}^2,\ (\mathbf{y}^1, \mathbf{y}^2) \rightarrow (\mathbf{x}, \mathbf{x})} \frac{\mathbf{e}_\Gamma(\mathbf{y}^1) - \mathbf{e}_\Gamma(\mathbf{y}^2)}{||\mathbf{y}^1 - \mathbf{y}^2||} = 0 \tag{24}$$
where
$$\mathbf{e}_\Gamma(\mathbf{y}) \equiv \Psi(\Phi(\mathbf{y})) - \Psi(\Phi(\mathbf{x})) - \Psi^{'}(\Phi(\mathbf{x}); \Phi^{'}(\mathbf{x}; \mathbf{y} - \mathbf{x})).$$
Since Ψ is F-differentiable at $\mathbf{v} = \Phi(\mathbf{x})$, thus $\Psi^{'}(\mathbf{v}; \cdot)$ is linear in the second argument, we have, for $i = 1, 2$,
$$\mathbf{e}_\Gamma(\mathbf{y}^i) = \mathbf{e}_\Psi(\Phi(\mathbf{y}^i)) + \partial\Psi(\Phi(\mathbf{x}))\mathbf{e}_\Phi(\mathbf{y}^i)$$
where
$$\mathbf{e}_\Psi(\mathbf{u}) = \Psi(\mathbf{u}) - \Psi(\mathbf{v}) - \partial\Psi(\mathbf{v})(\mathbf{u} - \mathbf{v}), \ \forall \mathbf{u} \in R^m$$
and
$$\mathbf{e}_\Phi(\mathbf{y}) = \Phi(\mathbf{y}) - \Phi(\mathbf{x}) - \partial\Psi(\mathbf{x})(\mathbf{y} - \mathbf{x}), \ \forall \mathbf{y} \in R^n$$

Since Ψ has a strong F-derivative at v, we have

$$\lim_{\mathbf{u}^1 \neq \mathbf{u}^2,\ (\mathbf{u}^1,\mathbf{u}^2)\to(\mathbf{v},\mathbf{v})} \frac{\mathbf{e}_\psi(\mathbf{u}^1) - \mathbf{e}_\Psi(\mathbf{u}^2)}{||\mathbf{u}^1 - \mathbf{u}^2||} = 0;$$

hence

$$\lim_{\mathbf{y}^1 \neq \mathbf{y}^2,\ (\mathbf{y}^1,\mathbf{y}^2)\to(\mathbf{x},\mathbf{x})} \frac{\mathbf{e}_\Psi(\Phi(\mathbf{y}^1)) - \mathbf{e}_\Psi(\Phi(\mathbf{y}^2))}{||\mathbf{y}^1 - \mathbf{y}^2||} -$$
$$\lim_{\mathbf{y}^1 \neq \mathbf{y}^2,\ (\mathbf{y}^1,\mathbf{y}^2)\to(\mathbf{x},\mathbf{x})} \frac{\mathbf{e}_\Psi(\Phi(\mathbf{y}^1)) - \mathbf{e}_\Psi(\Phi(\mathbf{y}^2))}{||\Phi(\mathbf{y}^1) - \Phi(\mathbf{y}^2)||} \cdot \frac{||\Phi(\mathbf{y}^1) - \Phi(\mathbf{y}^2)||}{||\mathbf{y}^1 - \mathbf{y}^2||} = 0,$$

where the last equality holds because Φ is locally Lipschitz continuous at $\mathbf{x}$, because it is B-differentiable at $\mathbf{x}$. Similarly, since Φ has a strong B-derivative at $\mathbf{x}$, we have

$$\lim_{\mathbf{y}^1 \neq \mathbf{y}^2,\ (\mathbf{y}^1,\mathbf{y}^2)\to(\mathbf{x},\mathbf{x})} \frac{\mathbf{e}_\Phi(\mathbf{y}^1) - \mathbf{e}_\Phi(\mathbf{y}^2)}{||\mathbf{y}^1 - \mathbf{y}^2||} = 0;$$

Combining the last two expressions gives us the desired result (24).

□

It is important to note that in the second part of the theorem, the order of composition is important; more precisely, if Ψ has a strong B-derivative at $\Phi(\mathbf{x})$ and Φ has a strong F-derivative at $\mathbf{x}$ the composite map $\Psi \circ \Phi$ does not necessarily have a strong B-derivative at $\mathbf{x}$ [10]. It also follows from the theorem that the composite of two weakly semismooth functions is weakly semismooth, since weak semismoothness and B-differentiability are equivalent.

Theorem 5.17. [10] Let a function $G : \Omega \subseteq R^n \to R^m$, with Ω open, be B-differentiable at a point $\mathbf{x}$ in Ω. For every vector $\mathbf{d} \in R^n$, there exists $H \in \partial G(\mathbf{x})$ such that $G^{'}(\mathbf{x}; \mathbf{d}) = H\mathbf{d}$.

Proof. Let $\{\tau_k\}$ be an arbitrary sequence of positive scalars converging to zero. For every k, we can write

$$G(\mathbf{x} + \tau_k \mathbf{d}) - G(\mathbf{x}) = \sum_{i=1}^{m} \tau_k \alpha_{i,k} H_i^k \mathbf{d} \tag{25}$$

for some scalars $\alpha_{i,k}$ satisfying

$$\sum_{i=1}^{m} \alpha_{i,k} = 1, \ \ \alpha_{i,k} \geq 0$$

and some matrices $H_i^k \in \partial G(\mathbf{x} + \tau_{i,k}^{'}\mathbf{d})$, where $\tau_{i,k}^{'} \in (0, \tau_k)$. By [10, Proposition 7.1.4], the sequence $\{H_i^k\}$ is bounded for every $i = 1, ..., m$. Without loss of generality, we may assume that each sequence $\{H_i^k\}$ converges to a limiting matrix $\{H_i^\infty\}$, which must belong to $\partial G(\cdot)$, by the closedness of the Clarke generalized Jacobian [Theorem 4.10, part 2] and by [10, Proposition 7.1.4]. We may further assume that each sequence $\{\alpha_{i,k}\}$ of scalars, for $i = 1, ..., m$, converges to a nonnegative scalar $\{\alpha_{i,\infty}\}$. Clearly, we have

$$\sum_{i=1}^{m} \alpha_{i,\infty} = 1.$$

Thus, dividing (25) by τ_k and letting $k \to \infty$, we deduce

$$G^{'}(\mathbf{x}; \mathbf{d}) = \sum_{i=1}^{m} \alpha_{i,\infty} H_i^\infty$$

belongs to $\partial G(\mathbf{x})$, by the convexity of the generalized Jacobian [Theorem 4.10, part 2].

□

Theorem 5.18. [10] Let $G : \Omega \subseteq R^n \to R^m$, with Ω open, be B-differentiable in $\overline{\mathbf{x}} \in \Omega$. The following three statements are equivalent:

1. G is strongly semismooth at $\overline{\mathbf{x}}$;
2. the following limit holds:
$$\lim_{\overline{\mathbf{x}} \neq \mathbf{x} \to \overline{\mathbf{x}}, H \in \partial G(\mathbf{x})} \frac{G^{'}(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}) - H(\mathbf{x} - \overline{\mathbf{x}})}{||\mathbf{x} - \overline{\mathbf{x}}||} = 0; \qquad (26)$$
3. the following limit holds:
$$\lim_{\overline{\mathbf{x}} \neq \mathbf{x} \to \overline{\mathbf{x}}, H \in \partial G(\mathbf{x})} \frac{G(\mathbf{x}) + H(\overline{\mathbf{x}} - \mathbf{x}) - G(\overline{\mathbf{x}})}{||\mathbf{x} - \overline{\mathbf{x}}||} = 0. \qquad (27)$$

In addition, if G is strictly semismooth at $\overline{\mathbf{x}}$, then

$$\lim \sup_{\overline{\mathbf{x}} \neq \mathbf{x} \to \overline{\mathbf{x}}} \frac{G(\mathbf{x}) - G(\overline{\mathbf{x}}) - G'(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}})}{||\mathbf{x} - \overline{\mathbf{x}}||^2} < \infty \tag{28}$$

$$\lim_{\overline{\mathbf{x}} \neq \mathbf{x} \to \overline{\mathbf{x}}, H \in \partial G(\mathbf{x})} \frac{G(\mathbf{x}) + H(\overline{\mathbf{x}} - \mathbf{x}) - G(\overline{\mathbf{x}})}{||\mathbf{x} - \overline{\mathbf{x}}||^2} < \infty \tag{29}$$

Proof. 1) $\Rightarrow$ 2). Suppose that G is strongly semismooth at $\overline{\mathbf{x}}$. Then the limit (26) is clearly equivalent to

$$\lim_{0 \neq d \to 0, H \in \partial G(\overline{\mathbf{x}} + \mathbf{d})} \frac{G'(\overline{\mathbf{x}}; \mathbf{d}) - H\mathbf{d}}{||\mathbf{d}||} = 0.$$

By Caratheodory's Theorem [Theorem 2.23] applied to the convex compact set $\partial G(\overline{\mathbf{x}} + \mathbf{d})$, it follows that for every $\mathbf{d} \in R^n$ and every element $H \in \partial G(\overline{\mathbf{x}} + \mathbf{d})$, there exist scalars α_i for $i = 1, ..., m+1$ and sequences of vectors $\{\mathbf{d}^{i,k}\}$ such that

$$\sum_{i=1}^{m+1} \alpha_i = 1, \ \alpha_i \geq 0$$

$$\lim_{k \to \infty} \mathbf{d}^{i,k} = \mathbf{d}, \ \forall i = 1, ..., m+1,$$

function G is F-differentiable at $\overline{\mathbf{x}} + \mathbf{d}^{i,k}$, and

$$H = \sum_{i=1}^{m+1} \alpha_i \lim_{k \to \infty} \partial G(\overline{\mathbf{x}} + \mathbf{d}^{i,k}).$$

Thus we have

$$\begin{aligned} & G'(\overline{\mathbf{x}}; \mathbf{d}) - H\mathbf{d} \\ &= \sum_{i=1}^{m+1} \alpha_i \lim_{k \to \infty} [(G'(\overline{\mathbf{x}}; \mathbf{d}^{i,k}) - G'(\overline{\mathbf{x}} + \mathbf{d}^{i,k}; \mathbf{d})] \\ &= \sum_{i=1}^{m+1} \alpha_i \lim_{k \to \infty} [(G'(\overline{\mathbf{x}}; \mathbf{d}^{i,k}) - G'(\overline{\mathbf{x}} + \mathbf{d}^{i,k}; \mathbf{d}^{i,k}) + \\ & \quad G'(\overline{\mathbf{x}} + \mathbf{d}^{i,k}; \mathbf{d}^{i,k}) - G'(\overline{\mathbf{x}} + \mathbf{d}^{i,k}; \mathbf{d})]. \end{aligned}$$

By the strong semismoothness of G at $\overline{\mathbf{x}}$, it follows that for every $\epsilon > 0$, there exists $\delta > 0$ such that for every vector $\mathbf{d}'$ satisfying $||\mathbf{d}'|| \leq \delta$,

$$||G'(\overline{\mathbf{x}}; \mathbf{d}') - G'(\overline{\mathbf{x}} + \mathbf{d}'; \mathbf{d}')|| < \epsilon ||\mathbf{d}'||.$$

Moreover, by the local Lipschitz continuity of G at $\overline{\mathbf{x}}$, it follows that there exists a neighborhood N of $\overline{\mathbf{x}}$ and a constant $K > 0$ such that for all vectors $\mathbf{y} \in N$ and $\mathbf{u}$ and $\mathbf{v} \in R^n$,

$$||G^{'}(\mathbf{y};\mathbf{u}) - G^{'}(\mathbf{y};\mathbf{v})|| \leq K\,||\mathbf{u}-\mathbf{v}||\,.$$

Thus

$$\lim_{\mathbf{y}\to\mathbf{x},\ ||\mathbf{u}-\mathbf{v}||\to 0}(G^{'}(\mathbf{y};\mathbf{u}) - G^{'}(\mathbf{y};\mathbf{v})) = 0.$$

Consequently, we deduce that for all $\mathbf{d}$ with $||\mathbf{d}||$ sufficiently small,

$$\left|\left|G^{'}(\overline{\mathbf{x}};\mathbf{d}) - H\mathbf{d}\right|\right| \leq \epsilon\,||\mathbf{d}||\,.$$

Hence (26) holds.

2) $\Leftrightarrow$ 3). By the B-differentiability of G at $\overline{\mathbf{x}}$, we have [10, Proposition 3.1.3]

$$\lim_{\overline{\mathbf{x}}\neq\mathbf{x}\to\overline{\mathbf{x}}} \frac{G(\mathbf{x}) - G(\overline{\mathbf{x}}) - G^{'}(\overline{\mathbf{x}};\mathbf{x}-\overline{\mathbf{x}})}{||\mathbf{x}||} = 0.$$

This limit clearly shows that (26) and (27) are equivalent.

2) $\Rightarrow$ 1). By Theorem 5.17, for every $\mathbf{x}$ sufficiently close to $\overline{\mathbf{x}}$, there exists $H \in \partial G(\mathbf{x})$ such that $G^{'}(\mathbf{x};\mathbf{x}-\overline{\mathbf{x}}) = H(\mathbf{x}-\overline{\mathbf{x}})$. Thus 1) follows from 2) readily.

Assume now that G is strictly semismooth at $\overline{\mathbf{x}}$. For any given vector $\mathbf{d}$, let $\Gamma(t) = G(\overline{\mathbf{x}} + t\mathbf{d})$. It is clear that Γ is locally Lipschitz continuous, and hence differentiable almost everywhere on $[0,1]$. Therefore, for all $\mathbf{d}$ with $||\mathbf{d}||$ sufficiently small, we can write

$$\begin{aligned} G(\overline{\mathbf{x}}+\mathbf{d}) - G(\overline{\mathbf{x}}) = \Gamma(1) - \Gamma(0) &= \int_0^1 \Gamma^{'}(t)dt \\ &= \int_0^1 G^{'}(\overline{\mathbf{x}}+t\mathbf{d};\mathbf{d})dt \\ &= \int_0^1 [G^{'}(\overline{\mathbf{x}};\mathbf{d}) + to(||\mathbf{d}||^2)]dt \\ &= G^{'}(\overline{\mathbf{x}};\mathbf{d}) + o(||\mathbf{d}||^2), \end{aligned}$$

where the fourth equality follows from the definition of strict semismoothness. By taking $\mathbf{d} = \mathbf{x} - \overline{\mathbf{x}}$, this chain of equalities establishes (28). Finally, using

(28) and following the above proof of the equivalence of statements 1), 2) and 3), we can easily establish (29).

□

An important way to obtain strictly semismooth functions is through composition. The next theorem makes this statement precise. This theorem implies in particular that the sum and difference of two strongly or strictly semismooth functions are strongly or strictly semismooth.

Theorem 5.19. [10] Let a function $F : \Omega_F \subseteq R^n \to R^m$, with Ω_F open, a point $\overline{\mathbf{x}}$ belonging to Ω_F, and a function $g : \Omega_g \subseteq R^m \to R$, with Ω_g being a neighborhood of $F(\overline{\mathbf{x}})$, be given. If F and g are strongly or strictly semismooth at $\overline{\mathbf{x}}$ and $F(\overline{\mathbf{x}})$ respectively, then the composite function $g \circ F$ is strongly or strictly semismooth at $\overline{\mathbf{x}}$.

Proof. We only consider the strongly semismooth case; the strictly semismooth case can be proven in a similar way. Since g is assumed to be a real-valued function, elements in $\partial(g \circ F)(\mathbf{x})$ are column vectors. By Theorem 5.16 we know that $g \circ F$ is B-differentiable and that

$$(g \circ F)'(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}) = g'(F(\overline{\mathbf{x}}); F(\mathbf{x} - \overline{\mathbf{x}})). \tag{30}$$

We need to show

$$\lim_{\overline{\mathbf{x}} \neq \mathbf{x} \to \overline{\mathbf{x}}, \boldsymbol{\xi} \in \partial(g \circ F)(\mathbf{x})} \frac{(g \circ F)'(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}) - \boldsymbol{\xi}^T(\mathbf{x} - \overline{\mathbf{x}})}{||\mathbf{x} - \overline{\mathbf{x}}||} = 0.$$

By Theorem 4.46 we can write, for every $\mathbf{x}$ sufficiently close to $\overline{\mathbf{x}}$:

$$\partial(g \circ F)(\mathbf{x}) \subseteq \text{conv } S(\mathbf{x}),$$

where $S(\mathbf{x}) = \partial F(\mathbf{x})^T \partial g(F(\mathbf{x}))$. Therefore we get

$$\begin{aligned} &\max_{\boldsymbol{\xi} \in \partial(g \circ F)(\mathbf{x})} \left|\left| \boldsymbol{\xi}^T(\mathbf{x} - \overline{\mathbf{x}}) - F'(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}) \right|\right| \\ \leq &\max_{\boldsymbol{\xi} \in \text{conv} S(\mathbf{x})} \left|\left| \boldsymbol{\xi}^T(\mathbf{x} - \overline{\mathbf{x}}) - F'(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}) \right|\right|, \end{aligned} \tag{31}$$

where we can write the maximum in the above formula because both the sets $\partial(g \circ F)(\mathbf{x})$ and $S(\mathbf{x})$ are nonempty and compact [Theorem 4.10, part 2]. Let us denote by $r : R^n \to [0, \infty)$ the function

$$r(\boldsymbol{\xi}) = \left|\left|\boldsymbol{\xi}^T(\mathbf{x} - \overline{\mathbf{x}}) - \partial(g \circ F)^{'}(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}})\right|\right|.$$

Obviously, r is a convex function on R^n; thus the maximum of $r(\boldsymbol{\xi})$ over conv$S(\mathbf{x})$ is attained at a point $\overline{\boldsymbol{\xi}}$ in $S(\mathbf{x})$. Let then $\overline{\boldsymbol{\xi}}$ be a point in $S(\mathbf{x})$ where r achieves the maximum. By the definition of $S(\mathbf{x})$ we can find a V in $\partial F(\mathbf{x})$ and a ς in $\partial g(F(\mathbf{x}))$ such that $\overline{\boldsymbol{\xi}} = V\varsigma$. Therefore, writing

$$\mathbf{d} = \mathbf{x} - \overline{\mathbf{x}} \text{ and } F_{\mathbf{d}} = f(\mathbf{x}) - F(\overline{\mathbf{x}}),$$

we have, for every $\mathbf{x}$ sufficiently close to $\overline{\mathbf{x}}$,

$$\max_{\boldsymbol{\xi} \in \partial(g \circ F)(\mathbf{x})} \left|\left|\boldsymbol{\xi}^T\mathbf{d} - (g \circ F)^{'}(\overline{\mathbf{x}}; \mathbf{d})\right|\right|$$
$$\leq \left|\left|\overline{\boldsymbol{\xi}}^T\mathbf{d} - (g \circ F)^{'}(\overline{\mathbf{x}}; \mathbf{d})\right|\right|, \text{ by (31)}$$
$$= \left|\left|\boldsymbol{\varsigma}^T V^T\mathbf{d} - g^{'}(F(\mathbf{x}); F^{'}(\overline{\mathbf{x}}; \mathbf{d}))\right|\right|, \text{ by (30)}$$
$$\leq \left|\left|\boldsymbol{\varsigma}^T V^T\mathbf{d} - g^{'}(F(\mathbf{x}); F_{\mathbf{d}})\right|\right| + o(||\mathbf{d}||), \text{ by (26) and the Lip.continuity of } g^{'}(F(\mathbf{x}); \cdot)$$
$$\leq \left|\left|\boldsymbol{\varsigma}^T V F_{\mathbf{d}} - g^{'}(F(\mathbf{x}); F_{\mathbf{d}})\right|\right| + o(||\mathbf{d}||), \text{ by (27)}$$
$$\leq \max_{\boldsymbol{\xi}\partial g(F(\mathbf{x}))} \left|\left|\boldsymbol{\xi}^T V F_{\mathbf{d}} - g^{'}(F(\mathbf{x}); F_{\mathbf{d}})\right|\right|$$
$$+ o(||\mathbf{d}||), \text{ because } \boldsymbol{\varsigma} \in \partial g(F(\mathbf{x}))$$
$$\leq o(||F_{\mathbf{d}}||) + o(||\mathbf{d}||), \text{ by the strong semismoothness of } g$$
$$= o(||\mathbf{d}||), \text{ by the Lip. continuity of F.}$$

This chain of inequalities obviously completes the proof.

□

By the definition of semismoothness, it is easy to check that a vector valued function is strongly or strictly semismooth if and only if each of its component functions are strongly or strictly semismooth. Thus Theorem 5.19

implies that the composition of two strongly semismooth vector functions is strongly semismooth.

Let us finally show the proof for Theorem 5.7.

Proof. If f is continuously differentiable in an open neighborhood on $\overline{\mathbf{x}}$, then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$ in the same neighborhood. With this observation, the strong semismoothness of f follows from Theorems 5.6 and 5.18; so does the strict semismoothness if ∇f is Lipschitz continuous near $\overline{\mathbf{x}}$.

To prove 3), we have to check that for every sequence $\{\mathbf{x}^k\}$ converging to $\overline{\mathbf{x}}$ and every sequence $\{\boldsymbol{\xi}^k\}$, with $\mathbf{x}^k \neq \overline{\mathbf{x}}$ and $\boldsymbol{\xi}^k \in \partial f(\mathbf{x}^k)$ for every k, we have

$$\lim_{k\to\infty} f^{'}(\overline{\mathbf{x}}; \mathbf{d}^k) = \lim_{k\to\infty} (\boldsymbol{\xi}^k)^T \mathbf{d}^k, \tag{32}$$

where

$$\mathbf{d}^k = \frac{\mathbf{x}^k - \overline{\mathbf{x}}}{||\mathbf{x}^k - \overline{\mathbf{x}}||}.$$

Without loss of generality, we may assume that

$$\lim_{k\to\infty} \mathbf{d}^k = \overline{\mathbf{d}} \text{ and } \lim_{k\to\infty} \boldsymbol{\xi}^k = \overline{\boldsymbol{\xi}} \in \partial f(\overline{\mathbf{x}}).$$

Since the left-hand limit and right-hand limit in (32) are equal to $f^{'}(\overline{\mathbf{x}}; \overline{\mathbf{d}})$ and $\overline{\boldsymbol{\xi}}^T \overline{\mathbf{d}}$, respectively, it remains to show that

$$\overline{\boldsymbol{\xi}}^T \overline{\mathbf{d}} = f^{'}(\overline{\mathbf{x}}; \overline{\mathbf{d}}).$$

Since f is convex and $\boldsymbol{\xi}^k$ is a subgradient (in the sense of classical convex analysis (Definition 3.16)) of f at $\mathbf{x}^k$, we have

$$f(\overline{\mathbf{x}}) - f(\mathbf{x}^k) \geq (\boldsymbol{\xi}^k)^T (\overline{\mathbf{x}} - \mathbf{x}^k);$$

dividing by $||\mathbf{x}^k - \overline{\mathbf{x}}||$ and letting $k \to \infty$, we deduce

$$\overline{\boldsymbol{\xi}}^T \overline{\mathbf{d}} \geq f^{'}(\overline{\mathbf{x}}; \overline{\mathbf{d}}).$$

But since $\overline{\boldsymbol{\xi}} \in \partial f(\overline{\mathbf{x}})$, we also have

$$f^{'}(\overline{\mathbf{x}}; \overline{\mathbf{d}}) \geq \overline{\boldsymbol{\xi}}^T \overline{\mathbf{d}}.$$

Thus equality holds.

□

5.3 Weakly upper semismooth functions

Now we will define the class of weakly upper semismooth functions and discuss its relation to semismooth functions.

Definition 5.20. [14] Function $f : R^n \to R$ is *weakly upper semismooth* at $\mathbf{x} \in R^n$ if it is Lipschitz continuous in $B(\mathbf{x}; \boldsymbol{\delta})$, $\boldsymbol{\delta} > 0$ and for each $\mathbf{d} \in R^n$ and for any sequences $\{t_k\} \subset R_+$ and $\boldsymbol{\xi}_k \subset R^n$ such that $\{t_k\} \to 0$ and $\boldsymbol{\xi}_k \in \partial f(\mathbf{x} + t_k\mathbf{d})$ it follows that

$$\liminf_{k\to\infty} \boldsymbol{\xi}_k^T \mathbf{d} \geq \limsup_{t\to 0} [f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})]/t.$$

Next we will show a relation between weakly semismooth functions and the directional derivative.

Theorem 5.21. [14] If f if weakly upper semismooth at $\mathbf{x}$, then for each $\mathbf{d} \in R^n$, $f^{'}(\mathbf{x}; \mathbf{d})$ exists and there exist sequences $\{\tau_k\} \subset R_+$ and $\{\boldsymbol{\xi}_k\} \subset R^n$ such that $\{\tau_k\} \to 0$, $\boldsymbol{\xi}_k \in \partial f(\mathbf{x} + \tau_k\mathbf{d})$ and

$$\lim_{k\to\infty} \boldsymbol{\xi}_k^T \mathbf{d} = f^{'}(\mathbf{x}; \mathbf{d}).$$

Proof. Suppose $\{\tau_k\} \to 0$ is a sequence such that

$$\lim_{k\to\infty} [f(\mathbf{x} + \tau_k\mathbf{d}) - f(\mathbf{x})]/\tau_k = \liminf_{t\to 0} [f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})]/t.$$

By the Mean-value theorem there exists $t_k \in (0, \tau_k)$ and $\boldsymbol{\xi}_k \in \partial f(\mathbf{x} + t_k\mathbf{d})$ such that

$$f(\mathbf{x} + \tau_k\mathbf{d}) - f(\mathbf{x}) = t_k \boldsymbol{\xi}_k^T \mathbf{d}.$$

Then, by the definition of weakly upper semismooth functions, since $\{t_k\} \to 0$, we have

$$\lim_{k\to\infty} [f(\mathbf{x} + \tau_k + \mathbf{d}) - f(\mathbf{x})]/\tau_k = \lim_{k\to\infty} \boldsymbol{\xi}_k^T \mathbf{d} \geq \limsup_{t\to 0} [f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})]/t.$$

So,

$$\liminf_{t\to 0} [f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})]/t = \lim_{k\to\infty} \boldsymbol{\xi}_k^T \mathbf{d} \geq \limsup_{t\to 0} [f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})]/t,$$

and the desired result follow immediately. □

It follows from the above thorems and definitions that if f is weakly semismooth, then f and $-f$ are weakly upper semismooth.

Let us show an example of a function that is weakly upper semismooth but not weakly semismooth.

Example 5.22. [14] Let us define function f

$$f(x) = x^2,\ x \leq 0 \text{ or } x \geq 1,$$

and for each integer $n = 1, 2, ...$

$$f(x) = \begin{cases} (1+\frac{1}{n})(x - \frac{1}{n+1}), & \frac{1}{n}[1 - (\frac{1}{n+1})^2] \leq x \leq \frac{1}{n} \\ \frac{1}{n}[1 - (\frac{1}{n+1})^2], & \frac{1}{n+1} \leq x \leq \frac{1}{n}[1 - (\frac{1}{n+1})^2]. \end{cases}$$

It can be verified that $f^{'}(0;1) = 0$ and $\partial f(0) = \text{conv}\{0,1\}$ is the set of possible accumulation points of $\{\xi_k\}$ where $\xi_k \in \partial f(x_k)$ and $\{x_k\} \to 0$.

Next we will define the class of upper semidifferentiable functions, which is closely related to the class of weakly upper semismooth functions.

Definition 5.23. [15] A locally Lipschitz continuous function f is *upper semidifferentiable*, if for all $\mathbf{d}$ and for all sequences

$$\{\theta_i\} \subset R_+,\ \{\boldsymbol{\xi}_i\} \subset R^n,\ \theta_i \to 0,\ \boldsymbol{\xi}_i \in \partial f(\mathbf{x} + \theta_i \mathbf{d}),$$

there exists $K \subset N$ such that

$$\lim_{k\to\infty,\ k\in K} \{[f(\mathbf{x} + \theta_k \mathbf{d}) - f(\mathbf{x})]/\theta_k - \boldsymbol{\xi}_k^T \mathbf{d}\} \leq 0.$$

Next we will analyze the relationship between the two classes of functions.

Theorem 5.24. [15] If f is weakly upper semismooth at $\mathbf{x}$, then for each $\mathbf{d} \in R^n$, $f^{'}(\mathbf{x};\mathbf{d})$ exists and there are sequences

$$\{t_k\} \subset R_+,\ \{\boldsymbol{\xi}_k\} \subset R^n,$$

such that

$$t_k \to 0,\ \boldsymbol{\xi}_k \in \partial f(\mathbf{x} + t_k \mathbf{d}),$$

and

$$\lim_{k\to\infty} \boldsymbol{\xi}_k^T \mathbf{d} = f^{'}(\mathbf{x};\mathbf{d}).$$

Proof. See [14]. □

Theorem 5.25. [15] Let $f(\mathbf{x})$ be a locally Lipschitz continuous function. Denote

$$\partial_{\epsilon\eta} f(\mathbf{x}) = \text{conv}\left\{ \bigcup_{||\mathbf{y}-\mathbf{x}||\leq\epsilon} \{\partial f(\mathbf{y}) + \eta\partial(|| - \mathbf{x}||)(\mathbf{y})\} \right\}.$$

For all $\epsilon \geq 0$, $\eta \geq 0$, we have that $\partial_{\epsilon\eta} f(\mathbf{x})$ is a nonempty, convex and compact set of R^n.

Proof. See [15]. □

Theorem 5.26. [15] If f is weakly upper semismooth at $\mathbf{x}$, then f is upper semidifferentiable at $\mathbf{x}$.

Proof. Let $\mathbf{d}$ be any vector of R^n. Let $\{\theta_i\} \subset R_+$, with $\theta_i \to 0$, and $\boldsymbol{\xi}_i \in \partial f(\mathbf{x} + \theta_i \mathbf{d})$. Since

$$\lim_{i\to\infty} (\mathbf{x} + \theta_i \mathbf{d}) = \mathbf{x},$$

the sequence $\{\mathbf{x} + \theta_i \mathbf{d}\}_{i\in N}$ belongs to some compact set of R^n, and, by Theorem 5.25, we can find a set $K \subseteq N$, with

$$\lim_{k\to\infty, k\in K} \boldsymbol{\xi}_k^T \mathbf{d} = \lim \inf_{k\to\infty} \boldsymbol{\xi}_k^T \mathbf{d} = \alpha \in R.$$

By Theorem 5.24, $f^{'}(\mathbf{x}; \mathbf{d})$ exists, hence,

$$\lim_{k\to\infty, k\in K} \{[f(\mathbf{x} + \theta_k \mathbf{d}) - f(\mathbf{x})]/\theta_k\} = f^{'}(\mathbf{x}; \mathbf{d}).$$

Since f is weakly upper semismooth at $\mathbf{x}$, we obtain

$$f^{'}(\mathbf{x}; \mathbf{d}) \leq \alpha,$$

hence,

$$\lim_{k\to\infty, k\in K} \{[f(\mathbf{x} + \theta_k \mathbf{d}) - f(\mathbf{x})]/\theta_k - \boldsymbol{\xi}_k^T \mathbf{d}\} \leq 0,$$

which proves that f is upper semidifferentiable at $\mathbf{x}$. □

The next example shows that the class of upper semidifferentiable functions contains strictly the class of weakly upper semismooth functions.

Example 5.27. [15] Let

$$f(t) = \begin{cases} t\sin[\log(\log(1/t)), & 0 < t \leq 1/2, \\ (1/2)\sin[\log(\log 2)], & 1/2 < t, \\ 0, & t \leq 0. \end{cases}$$

This function is locally Lipschitz continuous at R. It cannot be weakly upper semismooth at 0, because $f^{'}(0;1)$ does not exist. Nevertheless, this function is upper semidifferentiable at 0. Indeed, let $d = 1$ (the case $d = -1$ is obvious). If $0 < t < 1/2$, we have

$$\partial f(0+t1) = \{\nabla f(t)\} = \{\sin[\log(\log(1/t))] - \cos[\log(\log(1/t))] \times [1/\log(1/t)]\}.$$

Hence, we have

$$\lim_{k\to\infty}\{[f(0+t_k1) - f(0)]/t_k - < g_k, 1 >\}$$

$$= \lim_{t_k\to 0}\{\sin[\log(\log(1/t_k))]\}$$

$$-\{\sin[\log(\log(1/t_k))] - \cos[\log(\log(1/t_k))][1/\log(1/t_k)]\}$$

$$\lim_{t\to 0}\{\cos[\log(\log(1/t_k))][1/\log(1/t_k)]\} = 0.$$

The nonexistence of $f^{'}(0,1)$ is the key point of this example, as we shall see in the following theorem.

Theorem 5.28. [15] If $f'(\mathbf{x};\mathbf{d})$ exists for all $\mathbf{d} \in R^n$, and if f is upper semidifferentiable at $\mathbf{x}$, then f is weakly upper semismooth at $\mathbf{x}$.

Proof. Let

$$\mathbf{d} \in R^n,\ \{t_k\} \in R_+,\ \{\boldsymbol{\xi}_k\} \subset R^n,$$

such that

$$\boldsymbol{\xi}_k \in \partial f(\mathbf{x} + t_k\mathbf{d}) \text{ and } t_k \to 0.$$

Let $K^{'} \subseteq N$ be a set such that

$$\lim_{k\to\infty,\ k\in K^{'}} \boldsymbol{\xi}_k^T\mathbf{d} = \lim\inf_{k\to\infty} \boldsymbol{\xi}_k^T\mathbf{d} = \alpha \in R.$$

Since f is upper semidifferentiable at x, there is a set $K^{''} \subset K^{'}$ such that

$$\lim_{k\to\infty,\ k\in K^{''}} \{[f(\mathbf{x}+t_k\mathbf{d}) - f(\mathbf{x})]/t_k - \boldsymbol{\xi}_k^T\mathbf{d}\} \leq 0. \tag{33}$$

Both limits

$$\lim_{k\to\infty,\ k\in K^{''}} \{[f(\mathbf{x}+t_k\mathbf{d}) - f(\mathbf{x})]/t_k\} \text{ and } \lim_{k\to\infty,\ k\in K^{''}} \boldsymbol{\xi}_k^T\mathbf{d}$$

exist and are equal to $f^{'}(\mathbf{x};\mathbf{d})$ and α, respectively. So, by (33), we have

$$f^{'}(\mathbf{x};\mathbf{d}) \leq \alpha \text{ and } \lim\inf_{k\to\infty} \boldsymbol{\xi}_k^T\mathbf{d} \geq f^{'}(\mathbf{x};\mathbf{d});$$

hence, f is weakly upper semismooth at $\mathbf{x}$. □

6 Summary

This thesis starts with a discussion on convex sets and functions. First we define a convex set. Then we define a convex and strictly convex function. Convex functions and convex sets are linked with each other through the function epigraph. Quasi- and pseudoconvex functions are generalizations of a convex function. The minimum of a smooth convex function is found on a point where the function gradient is zero. The minimum of a nonsmooth function is found on a point where zero is one of the function subgradients.

The thesis continues with discussion on differentials of nonsmooth functions, which are not continuous. Clarke's generalized directional derivative is defined on their points of non-continuity, with which Clarke's subdifferential is defined. Then we discuss generalized pseudo- and quasiconvexity and l-quasiconvexity. We also define pseudo- and quasimonotonity for the generalized directional derivative.

The thesis ends with defining semismooth and weakly semismooth functions and through examples discussing the relations between semismoothness and the different kinds of convexities. Semismoothness and the directional derivative are related, because for a semismooth function at a point $\mathbf{x}$ the directional derivative exists for all directions at that point. We show that the maximum function of semismooth functions is semismooth. We also show that the composite function of semismooth functions is semismooth.

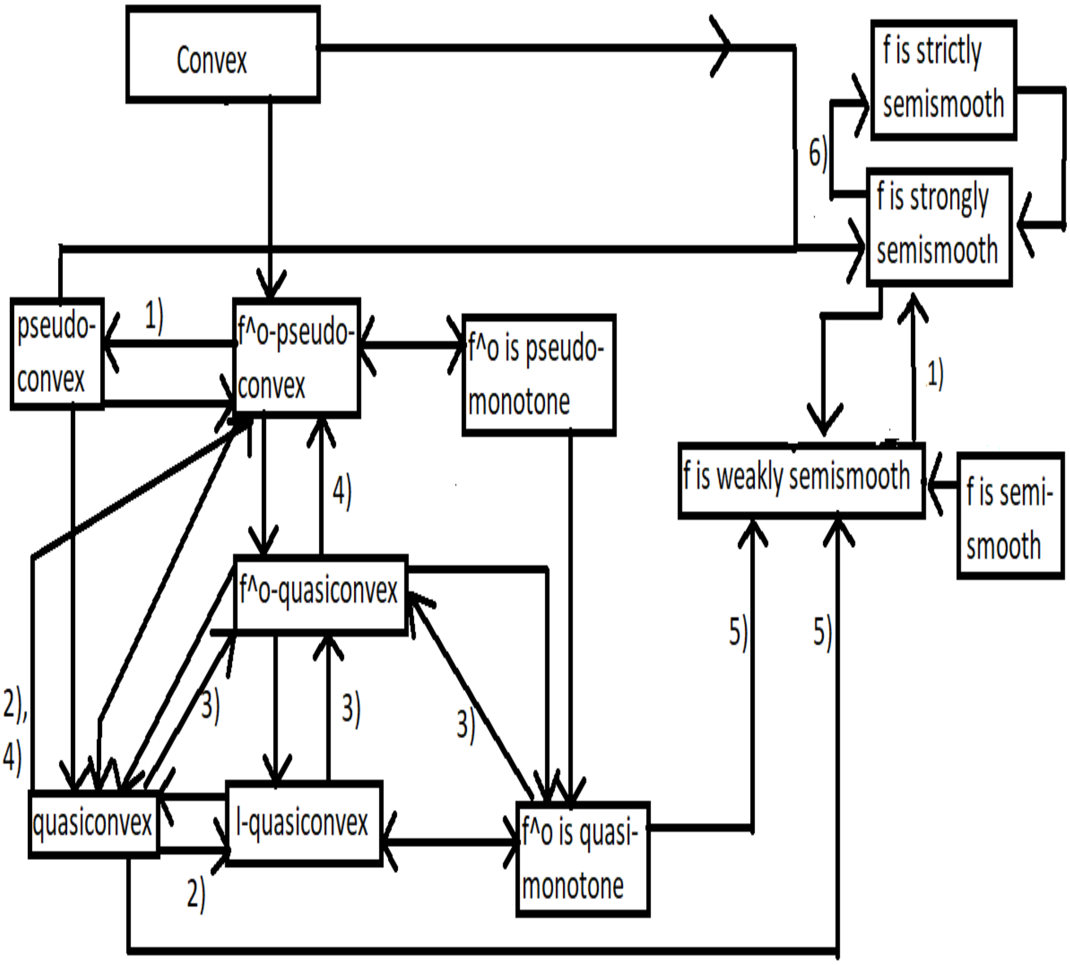
We define well-behaved generalized directional derivatives and show that a quasiconvex function is f^o-quasiconvex if and only if its generalized directional derivative is well-behaved.

Finally we define weakly upper semismooth and upper semidifferentiable functions and show their relation to weakly semismooth functions.

The purpose of this thesis was to figure out the relations between different convexities and other assumptions, like weak semismoothness and subdifferential regularity. If some assumptions followed from others, they would overlap and it would not be necessary to make them all. In [11, Theorem 3.18] it was assumed that one of the functions was subdifferentially reqular, f^o-quasiconvex and increasing. Since an increasing function is quasiconvex, it follows from subdifferential reqularity that the function is f^o-quasiconvex. Therefore the assumption of f^o-quasiconvexity is unnecessary. In addition, in [11, Theorem 3.17] it was assumed that the scaled improvement function $H(\mathbf{x},\mathbf{y}) = \max\{\mu_i(f_i(\mathbf{x})) - \mu_i(f_i(\mathbf{y})), \delta_l(g_l(\mathbf{x})), \ : \ i \in I,\ l \in L\}$ is weakly semismooth. It was also assumed that the functions μ_i and δ_l are subdifferentially regular and therefore weakly semismooth. If we assume that the functions f_i and g_l are weakly semismooth, the weak semismoothness of H follows from theorems 5.8 and 5.16.

The thesis introduced some completely new results, which shall be listed here: Theorem 5.2, Theorem 5.4, Theorem 5.8 and Theorem 5.13

Let us show a graph which summarizes the relations between different convexities and semismoothness.



1) Requires continuous differentiability. 2) Requires local Lipschitz continuity. 3) Requires that f^o is well-behaved 4) Requires that 0 ∈ δf(x) implies that x is a global minimum of f. 5) Requires that f is directionally derivative. 6) Requires that f is continuously differentiable with a Lipschitz continuous gradient.

Figure 17: Relations between different convexities and semismoothness

References

[1] Avriel, Diewert, Schaible, Zang: *Generalized concavity*, Plenum Press, New York, 1988.

[2] Bazaraa, Shetty: *Nonlinear Programming, Theory and Algorithms*, John Wiley & Sons, Inc, 1979.

[3] Papadopoulos: *Metric Spaces, Convexity and Nonpositive Curvature*, European Mathematical Society, IRMA, Strasbourg, France, 2005

[4] Mäkelä, Karmitsa, Eronen: *On Generalized Pseudo- and Quasiconvexities for Nonsmooth Functions. In "Current Research in Nonlinear Analysis. In Honor of Haim Brezis and Louis Nierenberg" (ed. Th. M. Rassias) Springer Optimization and Its Application*, 135, 2018.

[5] Mäkelä: *Nonsmooth Optimization: Theory and Algorithms with Applications to Optimal Control*, Jyväskylä, 1990.

[6] Outrata, Kocvara, Zowe: *Nonsmooth Approach to Optimization Problems with Equilibium Constraints*, Kluwer Academic Publishers, 1998.

[7] Bárány, Karasev: *Notes About the Carathéodory Number, Discrete & Computational Geometry 48*, July 2012.

[8] Robinson: *An Implicit-Function Theorem for B-differentiable Functions* ,International Institute for Applied Systems Analysis, Laxenburg, Austria, July 1988

[9] Mifflin: *Semismooth and Semiconvex Functions in Constrained Optimization*, International Institute for Applied Systems Analysis, Laxenburg, Austria, December 1976

[10] Facchinei, Pan: *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag New York, Inc., 2003

[11] Mäkelä, Montonen: *New multiobjective proximal bundle method with scaled improvement function, Numerical Nonsmooth Optimization,* p. 461-480, Bagirov, Gaudioso, Karmitsa, Mäkelä, Taheri (eds.), Springer, 2020

[12] `https://math.mit.edu/~jspeck/18.01_Fall14/Supplementary20notes/01c.pdf`, referred 2016-09-02

[13] Serrano, Schwartz, Gleixner: *On the Relation between the Extended Supporting Hyperplane Algorithm and Kelley's Cutting Plane Algorithm, Journal of Global Optimization,* May 2020

[14] Mifflin: *An Algorithm for Constrained Optimization with Semismooth Functions*, International Institute for Applied Systems Analysis, Laxenburg, Austria, February 1977

[15] Bihain: *Optimization of Upper Semidifferentiable functions. Journal of Optimization Theory and Algorithms 44, 4*, Plenum Publishing Corporation, New York-London, December 1984

[16] *"ADE (G1156) Spring 2006 Handout 3: Lipschitz condition and Lipschitz continuity", lecture notes of a Technical University of Eindhoven course*

[17] Ulbrich: *Semismooth Newton Methods for Operator Equations in Function Spaces*, *SIAM Journal on Optimization,* Rice University, Houston, May 2000

[18] Bagirov, Karmitsa, Mäkelä : *Introduction to Nonsmooth Optimization : Theory, Practice and Software*, Springer, 2014

[19] Clarke: *Optimization and Nonsmooth analysis*, University of Montreal, 1990

[20] Frigyik, Srivastava, Gupta : *An Introduction to Functional Derivatives*, University of Washington, January 2008

[21] Yu, Liu: *Four Kinds of Differentiable Maps, International Journal of Pureand Applied Mathematics, Volume 83*, Sichuan University of Science and Engineering, 2013