Lecture notes on Lexicographical derivative[3][1][2]

Tunan Wang

March 7, 2025

Contents

1 Notations 1

2 Motivation 2

3 Starting from directional derivatives 2

4 Lexicographically smooth functions 3

5 Lexicographic differentiation in Euclidean space 5

6 Necessary optimality condition with L−derivative 7

1 Notations

In this notes, we use $\mathbf{E}$ or $\mathbf{E}_i$, where $i = 1, 2, \ldots$, to denote finite-dimensional real vector spaces. For example, consider vector functions $f : \mathbf{E}_1 \to \mathbf{E}_2$, then its derivative $f'(\mathbf{x})$ at some point $\mathbf{x} \in \mathbf{E}_1$ is a *linear operator* from $\mathbf{E}_1$ to $\mathbf{E}_2$ defined by

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + f'(\mathbf{x}) \cdot \mathbf{h} + o_x(\mathbf{h}),$$

where $o_x(\mathbf{h}) \in \mathbf{E}_2$ is a small order term with

$$\lim_{\|h\| \to 0} \frac{\|o_x(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

For a convex fucntion $f : \mathbf{E} \to \mathbb{R}$, we denote by $\partial f(\mathbf{x})$ its subdifferential at point $\mathbf{x}$. Standard inner product of two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{E}$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$. Write out explicitly,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_{(i)} y_{(i)}, \quad n = \dim E.$$

For function $f : \mathbf{E}_1 \to \mathbf{E}_2$, we write $f \in \mathcal{D}(\mathbf{Q}, \mathbf{E}_2)$ that is locally Lipschitz continuous and differentiable in all directions at any point $\mathbf{x} \in \mathbf{Q} \in \mathbf{E}_1$.

Let $\mathbf{U}$ be a sequence of directions in $\mathbf{E}_1$. For any $k \in \mathbb{N}$, we denote the span of the first k directions of $\mathbf{U}$ by $\mathbf{L}_k(\mathbf{U})$. Let $\mathbf{L}_0(\mathbf{U}) := \{0\}$.

An important class of functions:

Definition 1.1: Piecewise differentiable($\mathcal{PC}^1$) function

A function $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is piecewise differentiable ($\mathcal{PC}^1$) at $x \in X$ if there exists a neighborhood $N_x \subset X$ of x and a finite collection of continuous differentiable functions

$$\mathcal{F}_f(x) = \{f_{(1)}, \ldots, f_{(k)}\}$$

of continuous functions mapping N_x into $\mathbb{R}^m$ such that f is continuous on N_x and for all $y \in N_x$, $f(y) \in \{f_{(i)}(y) : i \in \{1, \ldots, k\}\}$. If f is $\mathcal{PC}^1$ at all points of X, then f is called $\mathcal{PC}^1$ on X.

Remark.

> A composition of $\mathcal{PC}^1$ functions is $\mathcal{PC}^1$. A $\mathcal{PC}^1$ function is directionally differentiable.

2 Motivation

Consider writting a convex function in a nonconvex form,

Example.

> $f(x) = x = \max\{x, 0\} - \max\{-x, 0\}$, for $x \in \mathbb{R}$.

This function can also be written in the form of $\text{ReLU}(x) - \text{ReLU}(-x)$, which can be shown in some nerual network objective functions. When we want to have the derivative of this function at $x = 0$, i.e. 1, any subdifferential calculus fails to do that correctly (since it is a summation of a convex and a concave function). And for the current famous structure used (for example in *Torch*, *Tensorflow*, etc.) to calculate derivatives, i.e. automatic differentiation, also fails to calculate these kind of derivative correctly.

What's more, at each iteration of convex optimization methods, we do not need to have the whole subdifferential set of the objective function. Since most elements in the set can be considered as having the same power when using as some searching directions, and even in the convex case, subdifferential set is an unstable set. An arbitrarily small perturbation of the argument transforms it in a singleton (consider nonsmoothness at a knik point). Therefore, the majority of optimization schemes rely only on a possibility to get a single subgradient at current test point.

Therefore, the motivation is quite clear, we only want to consider one element in the subdifferential set and recover the nice calculation rule in smooth case.

3 Starting from directional derivatives

Definition 3.1: Homogeneous function

A function $f : \mathbf{E}_1 \to \mathbf{E}_2$ is called homogeneous if for all $\alpha \in \mathbb{R}$ and all $\mathbf{x} \in \mathbf{E}_1$,

$$f(\alpha \mathbf{x}) = \alpha f(\mathbf{x}).$$

Remark.

> Directional derivative is a homogeneous function. Can be proved by the definition.

Definition 3.2: Function homogenization

Let $f \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_2)$ and $\mathbf{x} \in \mathbf{E}_1$. The function $H_{\mathbf{x}}[f] : \mathbf{E}_1 \to \mathbf{E}_2$ defined by

$$H_{\mathbf{x}}[f](\mathbf{h}) = f'(\mathbf{x}; \mathbf{h}), \quad \mathbf{h} \in \mathbf{E}_1$$

is called the homogenization of f at x.

Proposition 3.3

Let $f, g \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_2)$ and $\mathbf{x} \in \mathbf{E}_1$. Then for any $\alpha, \beta \in \mathbb{R}$,

$$H_{\mathbf{x}}[\alpha f + \beta g] = \alpha H_{\mathbf{x}}[f] + \beta H_{\mathbf{x}}[g].$$

Note that, here we need to justify that the $\alpha f + \beta g$ belongs to $\mathcal{D}(\mathbf{E}_1, \mathbf{E}_2)$. Just use the definition of locally Lipschitz continuous and differentiable. Then, the linear property can be proved by the definition of directional derivative.

Proposition 3.4

If $f \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_2)$ and $F \in \mathcal{D}(\mathbf{E}_2, \mathbf{E}_3)$, then $F \circ f \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_3)$. Moreover,

$$H_{\mathbf{x}}[F \circ f](\cdot) = H_{f(\mathbf{x})}[F](H_{\mathbf{x}}[f](\cdot)).$$

This is the chain rule for directional derivative.

Proof. First show that $F \circ f \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_3)$, this is the same as above (Proposition 3.3).

Then, we need to show that $(F \circ f)'(\mathbf{x}, \mathbf{h}) = F'(f(\mathbf{x}), f'(\mathbf{x}, \mathbf{h}))$.

Sketch of proof:

Use linear expansion of directional derivative, then use the formulation to write $F(f(\mathbf{x} + \alpha\mathbf{h})) - F(f(\mathbf{x}))$. Then, take the limit w.r.t. α of

$$\frac{F(f(\mathbf{x} + \alpha\mathbf{h})) - F(f(\mathbf{x}))}{\alpha}.$$

□

Proposition 3.5

Let $f \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_2)$ be a homogeneous function. Then for any $\mathbf{x} \in \mathbf{E}_1$,

$$H_{\mathbf{x}}[H_{\mathbf{x}}[f]] = H_{\mathbf{x}}[f].$$

Moreover,

$$H_{\tau\mathbf{x}}[f] = H_{\mathbf{x}}[f] \quad \text{for any } \tau > 0,$$

$$H_{\mathbf{x}}[f](\alpha) = \alpha f(\mathbf{x}) \quad \text{for any } \alpha \in \mathbb{R}.$$

Finally, for any $\mathbf{x}, \mathbf{y} \in \mathbf{E}_1$ and any $\alpha \in \mathbb{R}$,

$$H_{\mathbf{x}}[f](\mathbf{y} + \alpha\mathbf{x}) = H_{\mathbf{x}}[f](\mathbf{y}) + \alpha f(\mathbf{x}).$$

Proof. First part, use the definition of homogeneous function and use the definition of $H_{\mathbf{x}}$.

The following part of the proof is just using the definitions. □

4 Lexicographically smooth functions

The functions we considered are non-smooth, but we still want some kind of subdifferntials, Lexicographically Smooth (Lexi-smooth) class is a relaxed definition of "smooth". We say that the Lexi-smooth functions to be the functions existing a sequence of directional derivatives at some points in the domain.

Definition 4.1: Homogenization sequence

Let $f \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_2)$, $\mathbf{x} \in \mathbf{E}_1$, and $\mathbf{U} = \{\mathbf{u}_k\}_{k=1}^{\infty}$, a sequence of vectors, called directions, in $\mathbf{E}_1$. The sequence of the recursively defined functions

$$f^{(0)}_{\mathbf{x},\mathbf{U}} = H_{\mathbf{x}}[f],$$

$$f^{(k)}_{\mathbf{x},\mathbf{U}} = H_{\mathbf{u}_k}[f^{(k-1)}_{\mathbf{x},\mathbf{U}}], \quad k \in \mathbb{N},$$

is called the homogenization sequence of f generated by $\mathbf{x}$ and $\mathbf{U}$, if it exists.

Definition 4.2: Lexicographically smooth functions

A function $f \in \mathcal{D}(\mathbf{E}_1, \mathbf{E}_2)$ is lexicographically smooth on $\mathbf{E}_1$, or Lexi-smooth on $\mathbf{E}_1$ for short, if its homogenization sequence exists for any $\mathbf{x} \in \mathbf{E}_1$ and any sequence $\mathbf{U}$ of directions in $\mathbf{E}_1$. If f is Lexi-smooth on $\mathbf{E}_1$, we write $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$.

Remark.

> The class of Lexi-smooth functions contains component-wise convex functions, differentiable functions, and compositions of Lexi-smooth functions. In the Euclidean spaces, i.e. $\mathbf{E}_1 = \mathbb{R}^n$ and $\mathbf{E}_2 = \mathbb{R}^m$, the class of Lexi-smooth functions also contains $\mathcal{PC}_1$ functions.

Proposition 4.3

1. Let $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2), \mathbf{x} \in \mathbf{E}_1$, and $\mathbf{U} = \{\mathbf{u}_k\}_{k=1}^{\infty}$, a sequence of directions in $\mathbf{E}_1$. Then $f^{(k)}_{\mathbf{x},\mathbf{U}}$ is Lipschitz continuous for each $k \in \{0, 1, 2, \dots\}$.
2. Let $f, g \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$. Then, for any $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$.

Proof. 1. Definition

2. Prove this by induction. The first iteration, there is

$$(\alpha f + \beta g)^{(0)}_{\mathbf{x},\mathbf{U}} = H_{\mathbf{x}}[\alpha f + \beta g] = \alpha H_{\mathbf{x}}[f] + \beta H_{\mathbf{x}}[g]$$

.

□

Theorem 4.4: Lexi-smooth is closed under composition

If $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$ and $F \in \mathcal{L}(\mathbf{E}_2, \mathbf{E}_3)$, then $F \circ f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_3)$.
Moreover, for any $\mathbf{x} \in \mathbf{E}_1$ and any sequence $\mathbf{U} = \{\mathbf{u}_k\}_{k=1}^{\infty}$ of directions in $\mathbf{E}_1$,

$$(F \circ f)^{(k)}_{\mathbf{x},\mathbf{U}}(\cdot) = F^{(k)}_{f(\mathbf{x}),\mathbf{V}}\left(f^{(k)}_{\mathbf{x},\mathbf{U}}(\cdot)\right), \quad k \in \{0, 1, 2, \dots\},$$

where the sequence $\mathbf{V} = \{\mathbf{v}_k\}_{k=1}^{\infty}$ of directions in $\mathbf{E}_2$ is given by

$$\mathbf{v}_k = f^{(k)}_{\mathbf{x},\mathbf{U}}(\mathbf{u}_k), \quad k \in \mathbb{N}.$$

Proof. By induction. □

Proposition 4.5

Let $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$, $\mathbf{x} \in \mathbf{E}_1$, and $\mathbf{U} = \{\mathbf{u}_k\}_{k=1}^{\infty}$, a sequence of directions in $\mathbf{E}_1$. Then the members of the homogenization sequence of f generated by $\mathbf{x}$ and $\mathbf{U}$ satisfy the following:

$$f^{(k)}_{\mathbf{x},\mathbf{U}}(\tau\mathbf{h}) = \tau f^{(k)}_{\mathbf{x},\mathbf{U}}(\mathbf{h}), \quad \text{where } \mathbf{h} \in \mathbf{E}_1 \text{ and } \tau \geq 0,$$

$$f^{(k)}_{\mathbf{x},\mathbf{U}}(\mathbf{h} + \alpha\mathbf{d}) = f^{(k)}_{\mathbf{x},\mathbf{U}}(\mathbf{h}) + \alpha f^{(k)}_{\mathbf{x},\mathbf{U}}(\mathbf{d}), \quad \text{where } \mathbf{h} \in \mathbf{E}_1,\ \mathbf{d} \in \mathbf{L}_{\text{k}}(\mathbf{U}), \text{ and } \alpha \in \mathbb{R},$$

$$f^{(k)}_{\mathbf{x},\mathbf{U}}(\sum_{i=1}^{k} \alpha_i \mathbf{u}_i) = \sum_{i=1}^{k} \alpha_i f^{(k)}_{\mathbf{x},\mathbf{U}}(\mathbf{u}_i), \quad \text{where } \sum_{i=1}^{k} \alpha_i \mathbf{u}_i \in \mathbf{L}_{\text{k}}(\mathbf{U}), \text{ and } \alpha_i \in \mathbb{R},\ \text{for } i \in \{1, 2, \dots\},$$

$$f^{(k)}_{\mathbf{x},\mathbf{U}}(\mathbf{d}) = f^{(k-1)}_{\mathbf{x},\mathbf{U}}(\mathbf{d}), \quad \text{where } \mathbf{d} \in \mathbf{L}_{\text{k-1}}(\mathbf{U}).$$

Proof. Induction. □

5 Lexicographic differentiation in Euclidean space

Theorem 5.1: Lexi-smooth is closed under finite dimensional composition

If $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$ and $F \in \mathcal{L}(\mathbf{E}_2, \mathbf{E}_3)$, then $F \circ f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_3)$.
Moreover, for any $\mathbf{x} \in \mathbf{E}_1$ and any sequence $\mathbf{U} = \{\mathbf{u}_k\}_{k=1}^m$ of directions in $\mathbf{E}_1$,

$$(F \circ f)^{(m)}_{\mathbf{x},\mathbf{U}}(\cdot) = F^{(m)}_{f(\mathbf{x}),\mathbf{V}}\left(f^{(m)}_{\mathbf{x},\mathbf{U}}(\cdot)\right), \quad k \in \{0, 1, 2, \dots\},$$

where the sequence $\mathbf{V} = \{\mathbf{v}_k\}_{k=1}^m$ of directions in $\mathbf{E}_2$ is given by

$$\mathbf{v}_k = f^{(m)}_{\mathbf{x},\mathbf{U}}(\mathbf{u}_k), \quad k \in \mathbb{N}.$$

Theorem 5.2: Degree of nondifferentiability

Let $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2), \mathbf{x} \in \mathbf{E}_1$, and $\mathbf{U} = \mathbf{u}_{k}{}_{k=1}^{m}$, an ordered set of directions in $\mathbf{E}_1$ that span $\mathbf{E}_1$. Then there exists $k_0 \in \{0, 1, \dots, m\}$ such that the function $f^{(k)}_{\mathbf{x},\mathbf{U}}$ is linear for any $k \geq k_0$. For any $k \geq k_0$,

$$f^{(k)}_{\mathbf{x},\mathbf{U}} = f^{(k_0)}_{\mathbf{x},\mathbf{U}}$$

where $f^{(k)}_{\mathbf{x},\mathbf{U}}$ is restricted to $\mathbf{L}_k(\mathbf{U})$, $f^{(k_0)}_{\mathbf{x},\mathbf{U}}$ is restricted to $\mathbf{L}_{k_0}(\mathbf{U})$, and $\mathbf{L}_k(\mathbf{U}) = \mathbf{L}_{k_0}(\mathbf{U}) = \mathbf{E}_1$.

Proof. Just show that $\mathbf{L}_k(\mathbf{U}) = \mathbf{L}_{k_0}(\mathbf{U}) = \mathbf{E}_1$ □

Definition 5.3: L-derivative

Let $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$, $\mathbf{x} \in \mathbf{E}_1$, and $\mathbf{U} = \{\mathbf{u}_k\}_{k=1}^m$, an ordered set of directions in $\mathbf{E}_1$ that span $\mathbf{E}_1$. For any k greater than or equal to the degree of nondifferentiability of f at $\mathbf{x}$ along $\mathbf{U}$, the function $f^{(k)}_{\mathbf{x},\mathbf{U}}$ restricted to $\mathbf{L}_{\mathrm{k}}(\mathbf{U})$ is called the lexicographic derivative (or $\mathbf{L}$-derivative for short) of f at $\mathbf{x}$ along $\mathbf{U}$ and is denoted by $\mathbf{J}_{\mathbf{L}} f(\mathbf{x}; \mathbf{U})$.

Definition 5.4: $\mathcal{L}$ subdifferential

The lexicographic subdifferential (or $\mathcal{L}$ subdifferential for short) of $f \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$ at $\mathbf{x} \in \mathbf{E}_1$ is defined to be

$$\partial_{\mathbf{L}} f(\mathbf{x}) = \{\mathbf{J}_{\mathbf{L}} f(\mathbf{x}; \mathbf{U}) : \mathbf{U} \text{ spans } \mathbf{E}_1\}$$

Proof. Combine (Theorem 4.4) and (Proposition 4.5) □

Claim: Relationships to other subdifferentials

Let X be an open subset of $\mathbb{R}^n$.

(i) If $f : X \to \mathbb{R}$ is a Lexi-smooth function, then $\partial_L f(x) \subseteq \partial f(x)$ for any $x \in X$.

(ii) If $f : X \to \mathbb{R}^m$ is a PC^1 function, then $\partial_L f(x) \subseteq \partial_B f(x) \subseteq \partial f(x)$ for any $x \in X$.

(iii) If $f : X \to \mathbb{R}^m$ is a continuous function, then $\partial_L f(x) = \partial_B f(x) = \partial f(x) = \{\mathbf{J} f(x)\}$ for any $x \in X$.

(iv) If $f : X \to \mathbb{R}^m$ is a Lexi-smooth function, then for any $\mathbf{x} \in X$ and any $\mathbf{d} \in \mathbb{R}^n$,

$$\{\mathbf{A}\mathbf{d} : \mathbf{A} \in \partial_L f(\mathbf{x})\} \subseteq \{\mathbf{A}\mathbf{d} : \mathbf{A} \in \partial f(\mathbf{x})\}.$$

where $\partial_B f(x)$ is the Bouligand subdifferentials, i.e. some limiting derivative around x. Note that the Clarke subdifferential is the convex hull of it.

Example.

L-derivative of $f(x) = |x|$ at $x = 0$

The direction vector is $\mathbf{U} = \{1, -1\}$. Then, the L-derivative at $x = 0$ can be calculated as:

$$f^{(0)}_{0,\mathbf{U}}(d) = f'(0; d) = |d| \tag{1}$$
$$f^{(1)}_{0,\mathbf{U}}(d) = [f^{(0)}_{0,\mathbf{U}}]'(1; d) = d \tag{2}$$
$$f^{(2)}_{0,\mathbf{U}}(d) = [f^{(0)}_{0,\mathbf{U}}]'(-1; d) = -d \tag{3}$$

Then, the $\mathbf{J_L} f(0; \mathbf{U}) = -1$.

Example.

L-derivative of $f(x_1, x_2) = \max(0, \min(x_1, x_2))$ at $x_1 = 0$ and $x_2 = 0$?

Suppose that the directions are $\mathbf{U} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$, the direction vector $\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$. Then, the L-derivative of f at $(0\ 0)^T$ is calculated with

$$f^{(0)}_{(0\ 0)^T,\mathbf{U}}(\mathbf{d}) = f'\left((0\ 0)^T; \mathbf{d}\right) = \max(0, \min(\mathbf{d})) \tag{4}$$
$$f^{(1)}_{(0\ 0)^T,\mathbf{U}}(\mathbf{d}) = [f^{(0)}_{(0\ 0)^T,\mathbf{U}}]'((1,0)^T; \mathbf{d}) = \max(0, d_2) \tag{5}$$
$$f^{(2)}_{(0\ 0)^T,\mathbf{U}}(\mathbf{d}) = [f^{(1)}_{(0\ 0)^T,\mathbf{U}}]'((0,1)^T; \mathbf{d}) = d_2 \tag{6}$$

That is, $\mathbf{J_L} f((0\ 0)^T; \mathbf{U}) = \mathbf{J} f^{(2)}_{(0\ 0)^T,\mathbf{U}}(\mathbf{d}) = (0\ 1)$, where $\mathbf{J}$ is the Jacobian matrix.

Note that,

$$\mathbf{J_L} f((0\ 0)^T; \mathbf{U}) \in \partial_B f(0,0) = \{(1,0), (0,1), (0,0)\}$$

.

Example.

L-derivative of $f(x) = \max\{x, 0\} - \max\{-x, 0\}$, at $x = 0$.

Suppose that $f_1(x) = \max\{x, 0\}$ and $f_2(x) = \max\{-x, 0\}$. Here $\mathbf{U} = \{1, -1\}$.

$$f^{(0)}_{1\,(0,\mathbf{U})}(d) = f_1'(0; d) = \max(d, 0) \tag{7}$$
$$f^{(1)}_{1\,(0,\mathbf{U})}(d) = [f^{(0)}_{1\,(0,\mathbf{U})}]'(1; d) = d \tag{8}$$
$$f^{(2)}_{1\,(0,\mathbf{U})}(d) = [f^{(1)}_{1\,(0,\mathbf{U})}]'(-1; d) = d \tag{9}$$

That is, $\mathbf{J_L} f_1(0; \mathbf{U}) = 1$.

$$f^{(0)}_{2\,(0,\mathbf{U})}(d) = f_1'(0; d) = \max(d, 0) \tag{10}$$
$$f^{(1)}_{2\,(0,\mathbf{U})}(d) = [f^{(0)}_{1\,(0,\mathbf{U})}]'(1; d) = 0 \tag{11}$$
$$f^{(2)}_{2\,(0,\mathbf{U})}(d) = [f^{(1)}_{1\,(0,\mathbf{U})}]'(-1; d) = 0 \tag{12}$$

That is, $\mathbf{J_L} f_1(0; \mathbf{U}) = 0$.

From (Proposition 4.3), we know that,

$$\mathbf{J_L} f(0; \mathbf{U}) = \mathbf{J_L} f_1(0; \mathbf{U}) - \mathbf{J_L} f_2(0; \mathbf{U}) = 1 \tag{13}$$

6 Necessary optimality condition with $\mathbb{L}-$derivative

Lemma 6.1: Necessary condition of directional derivative

If f from $\mathcal{D}(\mathbf{E}, \mathbb{R})$ attains an unconstrained local minimum at $\mathbf{x}^* \in \mathbf{E}$, then

$$f'(\mathbf{x}^*, \mathbf{h}) \geq 0, \mathbf{h} \in \mathbf{E}.$$

Theorem 6.2: Necessary optimality condition

If $x^*$is a local minimum of $f \in \mathcal{L}(\mathbf{E}, \mathbb{R})$, then for any basis $\mathbf{U}$ of $\mathbf{E}$ we have

$$\mathbf{U}^T \mathbf{J}_{\mathbf{L}} f(\mathbf{x}; \mathbf{U}) \succ 0$$

Proof. Refer to [3]'s Theorem 8. □

References

[1] Jaeho Choi. “Theory of Lexicographic Differentiation in the Banach Space Setting”. Electronic Theses and Dissertations. 3131. Master thesis. University of Maine, 2019. URL: https://digitalcommons.library.umaine.edu/etd/3131.

[2] Kamil A. Khan and Paul I. Barton. “Evaluating an element of the Clarke generalized Jacobian of a composite piecewise differentiable function”. In: *ACM Trans. Math. Softw.* 39.4 (July 2013). ISSN: 0098-3500. DOI: 10.1145/2491491.2491493. URL: https://doi.org/10.1145/2491491.2491493.

[3] Yu. Nesterov. “Lexicographic differentiation of nonsmooth functions”. In: *Math. Program.* 104.2–3 (Nov. 2005), pp. 669–700. ISSN: 0025-5610.