A Note on *Foundations of Optimization*

Xue Sen

March 22, 2022

1 Preparation

1.1 Differentiablility

Definition 1 *Differentiable(Fréchet Differentiable)*

A function of several real variables $\mathbf{f}:\mathbf{R}^m\to\mathbf{R}^n$ *is said to be differentiable at a point* $\mathbf{x}_0$ *if there exists a linear map* $\mathbf{J}:\mathbf{R}^m\to\mathbf{R}^n$ *such that*

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|\mathbf{f}(\mathbf{x_0}+\mathbf{h})-\mathbf{f}(\mathbf{x_0})-\mathbf{J}(\mathbf{h})\|_{\mathbf{R}^n}}{\|\mathbf{h}\|_{\mathbf{R}^m}}=0.$$

Lemma 1 *If all the partial derivatives of a function exist in a neighborhood of a point* x_0 *and are continuous at the point* x_0*, then the function is differentiable at that point* x_0*.*

However, the existence of the partial derivatives (or even of all the directional derivatives) does not in general guarantee that a function is differentiable at a point.

Definition 2 *Continuously Differentiable:*

A differentiable function f *is continuously differentiable if and only if* f *is of differentiability class* C^1*. That is, if the first order derivative of* f *(and possibly higher) is continuous.*

Real Function

Let $I\subset\mathbb{R}$ *be an open interval.*

Then f *is continuously differentiable on* I *if and only if* f *is differentiable on* I *and its derivative is continuous on* I*.*

Real-Valued Function

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

Let $f:U\to\mathbb{R}$ *be a real-valued function.*

Then f *is continuously differentiable in the open set* U *if and only if:*

(1): f *is differentiable in* U*.*

(2): the partial derivatives of f *are continuous in* U*.*

Vector-Valued Function

Let $U\subset\mathbb{R}^n$ *be an open set.*

Let $f:U\to\mathbb{R}^m$ *be a vector-valued function.*

Then f *is continuously differentiable in* U *if and only if* f *is differentiable in* U *and its partial derivatives are continuous in* U*.*

Definition 3 *Gateaux differentiable*

The function F *is called Gateaux differentiable at* $x\in U$ *of*

$$\lim_{t\to 0}\frac{F(x+td)-F(x)}{t}$$

is linear in d, that is, there exists a linear map $T:\mathbb{R}^n\to\mathbb{R}^m$*, say* $T(x)=Ax$*, where* A *is an* $m\times n$ *matrix, such that*

$$\lim_{t\to 0}\frac{F(x+td)-F(x)}{t}=Ad.$$

For the relationship among the definition, we have:

(iv) Frechet Differentiable $\Rightarrow$ (iii) Gateaux Differentiable $\Rightarrow$ (ii) All Directional Derivatives Exist $\Rightarrow$ (i) Partial Derivatives Exist

1.2 Algebra

Definition 4 *Topological Space*

A topological space is an ordered pair (X, τ), where X is a set and τ is a collection of subsets of X, satisfying the following axioms:

1. The empty set and X itself belong to τ.

2. Any arbitrary (finite or infinite) union of members of τ belongs to τ.

3. The intersection of any finite number of members of τ belongs to τ.

Definition 5 *Isomorphism*

The word isomorphism is derived from the Ancient Greek:isos "equal", and morphe "form" or "shape".

An isomorphism is a structure-preserving mapping between two structures of the same type that can be reversed by an inverse mapping.

Example: $A \to \alpha$, $b \to \beta$, then $a \times b \to \alpha \otimes \beta$.

An automorphism is an isomorphism from a structure to itself.

Definition 6 *Homomorphism*

The word homomorphism comes from the Ancient Greek language: homos meaning "same" and morphe meaning "form" or "shape".

A homomorphism is a map between two algebraic structures of the same type.

A map $f : A \to B$ between two sets A, B equipped with the same structure. $\otimes$ is an operation of the structure.

$$f(x \otimes y) = f(x) \otimes f(y)$$

Definition 7 *Homeomorphism*

The word homeomorphism comes from the Greek words homoios "similar" or "same" and morphe "shape" or "form".

A function $f : X \to Y$ between two topological spaces is a homeomorphism if it has the following properties:

1. f is a bijection (one-to-one and onto),

2. f is continuous,

3. the inverse function f^{-1} is continuous (f is an open mapping).

Definition 8 *Manifold*

A manifold is a topological space that locally homeomorphic to an open subset of an Euclidean space near each point(finite-dimensional).

Definition 9 *Diffeomorphism*

A diffeomorphism is a map between manifolds which is differentiable and has a differentiable inverse.

Theorem 1 *Implicit function theorem*

Theorem 2 *Lyusternik*

2 Unconstrained Opt

Theorem 3 *Weierstrass*

Theorem 4 *First-Order Necessary Condition for a local minima*

Theorem 5 *Second-Order Necessary Condition for a local minima*

Theorem 6 *Second-Order Sufficient Condition for a local minima*

Theorem 7 *Second-Order Sufficient Condition for a global minima*

3 Variational Principles

Definition 10 *Semi-Continuous*

4 Convex Analysis

Definition 11 *Affine Combination*

Definition 12 *Affine Set*

Definition 13 *Affine Hull*

Definition 14 *Affine Independent*

Definition 15 *Affine Map*

Definition 16 *Convex Set*

Definition 17 *Convex Combination*

Definition 18 *Cone(Convex Cone)*

Definition 19 *Positive Combination*

Definition 20 *Convex Conical Hull*

Definition 21 *Convex Function*

Theorem 8 *Differentiable and Convex*
Let C be a convex set in $\mathbb{R}^n$, and let f be a twice Fréchet differentiable function on an open set containing C.
(1) Convex $\Longleftrightarrow Hf(x)$ is positive semidefinite $\forall\ x \in C$.
(2) $Hf(x)$ is positive definite $\forall\ x \in C \longrightarrow$ strictly convex.

Theorem 9 *First-Order Necessary Condition for a local minima*
Add assumption that U is convex.

Theorem 10 *First-Order Sufficient Condition for a global minima*
Add assumption that f is convex.

Excercise:
The Theorem (9) and (10) can be used to solve opt problem when its constraints provide a convex set of x.
For example:

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & Ax = b \end{array}$$

and

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & Ax \leq a \\ & Bx = b \end{array}$$

5 Structure of Convex Sets and Functions

Definition 22 *Interior*

Definition 23 *Relative Interior*

Definition 24 *Algebraic interior*

Definition 25 *Relative Algebraic interior*

Definition 26 *Algebraic Closure*

Definition 27 *Minkowski Gauge Function*

6 Nonlinear Programming and KKT

For a nonlinear programming problem (P)

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \ldots, r, \quad (P) \\ & h_j(x) = 0, \quad j = 1, \ldots, m, \end{array}$$

Definition 28 *Active Constraints*

For a point $x \in$ feasible set $\mathcal{F}(P)$, The set of active constraints is

$$I(x) := \{i : g_i(x) = 0\}$$

Definition 29 *Tangent Direction(Tangent Cone)*

A vector $d \in \mathbb{R}^n$ is called a tangent direction of a nonempty set $M \subseteq \mathbb{R}^n$ at the point $x \in M$ if there exist a sequence $x_n \in M$ converging to x and a nonnegative sequence α_n such that

$$\lim_{n \to \infty} \alpha_n (x_n - x) = d$$

.

Definition 30 *Set of Feasible Directions($\mathcal{FD}(x^*)$)*

Let x^* be a feasible point, the set of all tangent directions of $\mathcal{F}(P)$ at x^* is the Set of Feasible Directions($\mathcal{FD}(x^*)$).

Linearized Set of Feasible Directions ($\mathcal{LFD}(x^*)$)

$$\begin{aligned} \mathcal{LFD}(x^*) := \{d : & \langle \nabla g_i(x^*), d \rangle \leq 0, \quad i \in I(x^*) \\ & \langle \nabla h_j(x^*), d \rangle = 0, \quad j = 1, \ldots, m, \}, \end{aligned}$$

Definition 31 *Set of Strict Descent Directions($\mathcal{SD}(f; x^*)$)*

A vector $d \in \mathbb{R}^n$ is called a strict descent direction for f at x^* if there exists a sequence of points $x_n \to x^*$ in $\mathbb{R}^n$ with tangent direction d such that $f(x_n) < f(x^*)$ for all n.

Linearized Set of Strict Descent Directions($\mathcal{SD}(f; x^*)$)

$$\mathcal{LSD}(f; x^*) := \{d : \langle \nabla f(x), d \rangle < 0\}.$$

Lemma 2 *If $x^* \in \mathcal{F}(P)$ is a local minimum of (P), then*

$$\mathcal{FD}(x^*) \cap \mathcal{SD}(f; x^*) = \emptyset$$

Theorem 11 *Fritz John*

If a point x^* is a local minimizer of (P), then there exist multipliers $(\lambda, \mu) := (\lambda_0, \lambda_1, \ldots, \lambda_r, \mu_1, \ldots, \mu_m)$, not all zero, $(\lambda_0, \lambda_1, \ldots, \lambda_r) \geq 0$, such that

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0 \tag{1}$$

$$\lambda_i \geq 0, g_i(x^*) \leq 0, \lambda_i g_i(x^*) = 0, \quad i = 1, \ldots, r \tag{2}$$

Proof: Consider Lyusternik's theorem for linear independent scenario.

Theorem 12 *KKT*

If a point x^* is a local minimizer of (P), and the vectors

$$\{\nabla g_i(x^*), i \in I(x^*), \quad \nabla h_j(x^*), j = 1, \ldots, m\}$$

are linearly independent, then,

$$\nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0 \tag{3}$$

$$\lambda_i \geq 0, \quad g_i(x^*) \leq 0, \quad \lambda_i g_i(x^*) = 0, \quad i = 1, \ldots, r \tag{4}$$

$$h_j(x^*) = 0, \quad j = 1, \ldots, m \tag{5}$$

Theorem 13 *First-Order Sufficient Optimality Conditions*

Let x^* be a feasible solution to the optimization problem (P) satisfying (1) and (2), If the totality of the vectors

$$span\{\lambda_0 \nabla f(x^*), \lambda_i \nabla g_i(x^*), \nabla h_j(x^*) | i \in i \in I(x^*), j \in \{1, \ldots, m\}\} = \mathbb{R}^n \tag{6}$$

then x^* is a local minimizer of (P).

Theorem 14 *Convex and KKT*

Let x^* be a feasible point of problem (P) if the active constraints $\{g_i\}_{i \in I(x^*)}$ are concave functions in a convex neighborhood of x^* and the equality constraints $\{h_j\}_1^m$ are affine functions on $\mathbb{R}^n$.

$$x \text{ is a local minimizer} \longrightarrow \text{ KKT conditions hold at } x$$

Theorem 15 *Variations of KKT**

Theorem 16 *Second-Order Necessary Conditions*

Let x^* be a local minimizer of (P) satisfying the KKT conditions with multipliers λ^*, μ^*. If the active gradient vectors,

$$\nabla g_i\left(x^*\right), i \in I\left(x^*\right), \nabla h_j\left(x^*\right), j = 1, \ldots, m$$

are linearly independent, then for any direction d satisfies

$$\left\langle d, \nabla g_i\left(x^*\right)\right\rangle = 0, i \in I\left(x^*\right), \quad \left\langle d, \nabla h_j\left(x^*\right)\right\rangle = 0, j = 1, \ldots, m,$$

then $\left\langle \nabla_x^2 L\left(x^*, \lambda^*, \mu^*\right) d, d\right\rangle \geq 0$

Theorem 17 *Second-Order Sufficient Conditions*

Theorem 9.21. Let x^* be a feasible point for (P) that satisfies the KKT conditions with multipliers λ^*, μ^*. If

$$\left\langle \nabla_x^2 L\left(x^*, \lambda^*, \mu^*\right) d, d\right\rangle > 0$$

for all $d \neq 0$ satisfying the conditions

$$\begin{aligned}
&\left\langle d, \nabla g_i\left(x^*\right)\right\rangle \leq 0, i \in I\left(x^*\right), \\
&\left\langle d, \nabla g_i\left(x^*\right)\right\rangle = 0, i \in I\left(x^*\right) \text{ and } \lambda_i^* > 0, \\
&\left\langle d, \nabla h_j\left(x^*\right)\right\rangle = 0, j = 1, \ldots, m,
\end{aligned}$$

then x^* is a strict local minimizer of (P).