Study Notes for Basic Mathematics

Mike Mingcheng Wei

May 8, 2020

Contents

Chapter 1

Function properties

1.1 Convex Functions

In this section, we need to discuss different kinds of convexity and the relationship between them.

1.1.1 Convex set

Definition 1 *Let $S \subseteq R^n$, S is* ***convex set*** *if for each $x^1, x^2 \in S$, $\lambda x^1 + (1-\lambda)x^2 \in S$ for $\forall\ \lambda \in [0,1]$.*

Theorem 2 *S is convex if when two points are in the set, then the line segment joining them is also in the set.*

Theorem 3 *If S is convex, then $S = H(S)$.*

Theorem 4 *Let E and F be nonempty convex sets inR^n.*
1. *$E + F$ is convex set;*
2. *rE is convex set, for $\forall r \in R$;*
3. *$E \cap F$ is convex set;*
4. *$H(E)$, convex hull of E, is convex set.*

Definition 5 *Let $S \subseteq R^n$, the* ***convex hull*** *of S, $H(S)$ or $Conv(S)$, is the set of all convex combination of points in S.*

Theorem 6 *$H(S)$ is the smallest convex set containing S.*

1.1.2 Convex Function and Equivalence

Definition 7 *Let $S \subseteq R^n$ and S is convex set, f is a* ***convex function*** *on S, if, for each $x^1, x^2 \in S$, we have $f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda)f(x^2)$, for $\forall\ \lambda \in [0,1]$. (This means f lies below every chord.)*

Definition 8 *Let $S \subseteq R^n$ and S is convex set, f is a* ***strict convex function*** *on S, if, for each $x^1, x^2 \in S$, we have $f(\lambda x^1 + (1-\lambda)x^2) < \lambda f(x^1) + (1-\lambda)f(x^2)$, for $\forall\ \lambda \in (0,1)$. (This means f lies below every chord.)*

Definition 9 *Let $S \subseteq R^n$, $S \neq \phi$, and S is convex; let $f : S \to R^1$, then the* ***level set*** *is defined as $S_\alpha = \{x \in S | f(x) \leq \alpha\}$*

Theorem 10 *Let $S \subseteq R^n$, $S \neq \phi$, and S is convex; let $f : s \to R^1$. If f is convex function on S, then the S_α for $\forall \alpha \in R^1$ is a convex set. (If S_α for $\forall \alpha \in R^1$ is a convex set, then f is quasi-convex)*

Definition 11 *Let $S \subseteq R^n$, $S \neq \phi$, and S is convex; let $f : s \to R^1$, then* ***epigraph of*** *f is defined as $epi(f) = \{(x,y) | x \in S, y \in R^1, y \geq f(x)\}$.*

Theorem 12 *(Equivalence of proving convex function) Let $S \subseteq R^n$, $S \neq \phi$, and S is convex; let $f : s \to R$. Then the following are equivalent:*

1. f is convex function on S;

2. for each $x^1, x^2 \in S$, we have $f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda)f(x^2)$, for $\forall\ \lambda \in [0,1]$

3. $epi(f)$ is a convex set;

4. for each $\bar{x} \in R^n$, $f(x) \geq f(\bar{x}) + (x-\bar{x})^t \nabla f(\bar{x})$ for $\forall x \in R^n$ (given f is differentiable);

5. $(\nabla f(x_2) - \nabla f(x_1))^t (x_2 - x_1) \geq 0$ for $\forall x_1, x_2 \in R^n$ (given f is differentiable);

6. for $H(x)$ is PSD $\forall x \in R^n$ (given f is twice differentiable); (For how to show H is PSD, refer to Theorem 17)

7.$-f$ is concave function on S;

8. f^{-1} is concave function (given f is invertible)

Theorem 13 *(Equivalence of proving strict convex function) Let $S \subseteq R^n$, $S \neq \phi$, and S is convex; let $f : s \to R$. Then the following are equivalent:*

1. f is strict convex function on S;

2. for each $x^1, x^2 \in S$, we have $f(\lambda x^1 + (1-\lambda)x^2) < \lambda f(x^1) + (1-\lambda)f(x^2)$, for $\forall\ \lambda \in (0,1)$.

3. for each $\bar{x} \in R^n$, $f(x) > f(\bar{x}) + (x-\bar{x})^t \nabla f(\bar{x})$ for $\forall x \in R^n$ and $x \neq \bar{x}$. (given f is differentiable);

4. $(\nabla f(x_2) - \nabla f(x_1))^t (x_2 - x_1) > 0$ for $\forall x_1, x_2 \in R^n$ and $x \neq \bar{x}$ (given f is differentiable);

5. $-f$ is strict concave function on S;

6. f^{-1} is strict concave function (given f is invertible)

Theorem 14 *If f is convex, than $f(x) + f(x+a) \geq f(x+\lambda a) + f(x + (1-\lambda)a)$ for $\lambda \in [0,1]$. (From Scarf 1959: bayes solutions of the statistical inventory problem, page 498 before equation 25)*

Theorem 15 *Let $S \subseteq R^n$, $S \neq \phi$, and S is open and convex; let $f : s \to R^1$ and f is twice differentiable, then if $H(x)$ is PD $\forall x \in R^n$, then f is strict convex function. (PS: f is strict convex function only infer $H(x)$ is PSD $\forall x \in R^n$).*

Theorem 16 *(From Wiki)The convexity property is preserved under*

1. Non-negative weighted maximum: $f = \max\{w_1 f_1, ..., w_n f_n\}$ is convex, where $f_1, ..., f_n$ are convex; $w_1, ..., w_n$ are non-negative;

2. Summation: if f and g are convex, then $f + g$ is convex;

3. Positive Linear Combination: $f = w_1 f_1 + w_2 f_2 + ... + w_n f_n$ is convex, where $f_1, ..., f_n$ are convex; $w_1, ..., w_n$ are non-negative; (so convexity is preserved under expectation and integration)

4. Composition with a non-decreasing function: let $f : R^n \to R$ is convex, $h : R \to R$ is convex and non-decreasing, then $h \circ f$ is convex[1]

5. Under affine maps: if $f(x)$ is convex with $x \in R^n$, then $g(y) = f(Ay + b)$ is convex, where $y \in R^m$, $A \in R^{n*m}$, $b \in R^n$.

6. Maximization: Let $f(x, y) : X \times Y \to R$. If Y is non-empty and X is convex set, $f(\cdot, y)$ is convex function on a convex set X for each $y \in Y$. Then $g(x) = \sup_{y \in Y} f(x, y)$ in convex on X. (Heyman and Sobel, 1984:525)

7. Minimization: Let $f(x, y) : X \times Y(x) \to R$. If $Y(x)$ is a nonempty set for every $x \in X$, X is convex set, and $(X, Y(x))$ is convex set, $f(x, y)$ is convex function on $(X, Y(x))$, $g(x) > -\infty$ for $\forall x \in X$. Then $g(x) = \inf_{y \in C} f(x, y)$ in convex on X. (Heyman and Sobel, 1984:525)

8. Under perspective function transformation: if $f(x)$ is convex, then its perspective function $g(x, t) = tf(x/t)$ ic convex.

9. Limitation operation: if f^n are convex, then $\lim_{n \to \infty} f^n$ is convex;

[1] The analogous claim for concave function is odd: let $f : R^n \to R$ is concave, $h : R \to R$ is concave and non-decreasing, then $h \circ f$ is concave. If f is strict concave and h is strictly increasing and concave then $h \circ f$ is strict concave

Convexity and Hessian Matrix

Theorem 17 *Let one of the following assumptions hold for the Hessian matrix H:*

1. *All its eigenvalues are positive; (Necessary and Efficient condition)*
2. *the determinant of every principal miinor is nonnegative; (Necessary and Efficient condition)*
3. *has positive diagonal elements and is diagonally dominant; (Sufficient condition)*
4. *$H = A^T A$: Hessian matrix H can be decomposite to product of A^T and A*
5. *If $H = B^T AB$, where A is $n \times n$ and positive definite and B is $n \times m$ with rank m, and $m \leq n$;*
6. *H^{-1} is positive definite*

Then, Hessian matrix H is positive semidefinite.

Theorem 18 *(Young, 1971:14) The number λ is an eigenvalue of $A_{n\times n}$ iff λ is a root of the characteristic equation*

$$\det(A - \lambda I) = 0$$

where $\det(\cdot)$ *is the determinant and I is the identity matrix*

Theorem 19 *(Young, 1971:14) The eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ satisfy*

$$\prod_{i=1}^{n} \lambda_i = \det(A) \text{ and } \sum_{i=1}^{n} \lambda_i = trace(A)$$

where $trace(\cdot)$ is the sum of diagonal elements of the squared matrix.

(2 by 2 matirx's eigenvalues can be found by this way very efficiently)

Theorem 20 *(Ostrowski, 1960) The eigenvalues of a matrix are continuous functions of its elements.*

Theorem 21 *(Young, 1971:16) All eigenvalues of a symmetric matrix are eral.*

Theorem 22 *(Gerschgorin Bounds on Eigenvalues) Let δ_i denote the sum of the absolute values of the off-diagonal elements in row i. That is: $\delta_i = \sum_{j\neq i} |a_{ij}|$. All eigenvalues of A lie in the union of the following sets:*

$$\{\lambda| \, |\lambda - a_{ii}| \leq \delta_i\} \text{ for } 1 \leq i \leq n$$

Definition 23 *A matrix is **diagonally dominant** if the absolute value of each diagonal element exceeds the sum of the absolute values of the off-diagonal elements in its row:*

$$|a_{ii}| \geq \sum_{j\neq i} |a_{ij}| \text{ for all } i$$

*It is **strictly diagonally dominant** if the inequality holds strictly for the above equation.*

Theorem 24 *If a symmetric matrix has positive diagonal elements and is diagonally dominant / strictly diagonally dominant. Then it is positive semidefinite / definite.*

Theorem 25 *(Heyman and Sobel, 1984:537) Any matrix of the form $A^T A$ is positive semidefinite.*

Theorem 26 *(Heyman and Sobel, 1984:537) If A is positive definite, then A is nonsigular and A^{-1} is positive definite.*

Theorem 27 *(Heyman and Sobel, 1984:537) If A is $n \times n$ and positive definite and B is $n \times m$ with rank m, and $m \leq n$, then $B^T AB$ is positive definite.*

1.1.3 K-Convex Function

K-convexity is only defined for functions of a single real variable, while convexity is defined for functions of many real variables.

In general, K-convex function is defined for dynamic inventory with fixed order cost model. Also, there are quasi-K-convexity, quasi-K-convexity with changeover, and nontrivially quasi-K-convex definied needed for more general (s, S) invenotry models, for reference, please check chapter 9 of Porteus

Definition 28 *(Equivalence of $K-$convex function) Let $f : R \to R$, and $K \geq 0$. Then the following are equivalent:*

1. *f is $K-$convex;*
2. *For each $x \leq y$, $0 \leq \theta \leq 1$, $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)[K + f(y)]$.*
3. *$K + f(x+a) \geq f(x) + \frac{a}{b}[f(x) - f(x-b)]$, for all $x \in R$, $a \geq 0$, and $b > 0$.*
4. *$K + f(y) \geq f(x) + f^{'}(x)(y-x)$ for all $x \leq y$. (f is C^1)*

Theorem 29 *$K-$convex function can NOT have a positive jump at a discontinuity.*
(A negative jump cannot be too large)

Theorem 30 *If f is convex, then f is $K-$convex for any $K \geq 0$.*
(However, if f is $K-$convex, it is not necessary f is convex or quasi-convex.)

Theorem 31 *(Scarf, 1960)The $K-$convexity property is preserved under*

1. *scalar multiply: if f is $K-$convex and s is a positive scalar, then sf is $k-$convex for all $k \geq sK$.*
2. *Summation: if f is $K-$convex and g is $k-$convex, then $f+g$ is $(k+K)-$convex.*

Theorem 32 *If v is $K-$convex, ϕ is the probability density of a positive random variable, and $G(y) := E[v(y-D)] = \int_0^\infty v(y-\xi)\phi(\xi)d\xi$. Then G is $K-$convex.*

Theorem 33 *If f is $K-$convex, $x < y$, and $f(x) = K + f(y)$. Then $f(z) \leq K + f(y)$ for all $z \in [x,y]$. ($K-$convex function f can cross the value $K + f(y)$ at most once on $(-\infty, y)$ for each real y.)*

1.1.4 Quasi-Convex Function

Definition 34 *Let $S \subseteq R^n$ and S is convex set, f is a **quasi-convex function** on S, if, for each $x^1, x^2 \in S$, we have $f(\lambda x^1 + (1-\lambda)x^2) \leq \max\{f(x^1), f(x^2)\}$, for $\forall\ \lambda \in [0,1]$.*

Remark 35 *For quasi-convex function, it can have sadder point / point of inflection, flat spot, and discontinuity.*

Theorem 36 *(Equivalence of proving quasi-convex function) Let $S \subseteq R^n$, $S \neq \phi$, and S is convex; let $f : s \to R$. Then the following are equivalent:*

1. *f is quasi-convex function on S;*
2. *for each $x^1, x^2 \in S$, we have $f(\lambda x^1 + (1-\lambda)x^2) \leq \max\{f(x^1), f(x^2)\}$, for $\forall\ \lambda \in [0,1]$.*
3. *the S_α for $\forall \alpha \in R^1$ is a convex set.[**?**]*
4. *for all $x^{'}, x \in S$, $\nabla f(x)(x^{'} - x) \geq 0$ whenever $f(x^{'}) \geq f(x)$ (given f is differentiable);[**?**]*
5. *for all $x^{'}, x \in S$, $f(x^{'}) > f(x)$ whenever $\nabla f(x)(x^{'} - x) > 0$ whenever (given f is differentiable);[**?**]*
6. *for each $x \in S$, the Hessian matrix $D^2 f(x)$ is negative semidefinite in the subspace $\left\{z \in R^N : \nabla f(x) \cdot z = 0\right\}$, that is, if and only if $z \cdot D^2 f(x) \cdot z \geq 0$ whenever $\nabla f(x) \cdot z = 0$ (given f is twice differentiable);[**?**]*

Theorem 37 *A bivariate function $g(x,y)$ is jointly quasiconcave in two variables if and only if every vertical slice of the function is quasiconcave, or more formally, if and only if $g(x,y)$ is quasiconcave given $mx+y = k$ for any real values m and k. (Lemma 1 of Zhao and Atkins. 2008. Nesvendors under simultaneous price and inventory competition. MSOM. 10(3).)*

Theorem 38 *If f is quasi-convex function on S. Let $g : S \to R$ be an non-decreasing function. Then $g \circ f$ is quasiconcave function.*

Theorem 39 *If f is quasi-convex function on S and $k > 0$, then kf is quasiconcave function.*

Definition 40 *Let $S \subseteq R^n$ and S is convex set, f is a* ***strictly quasi-convex function***[2] *on S, if, for each $x^1, x^2 \in S$ with $f(x^1) \neq f(x^2)$, we have $f(\lambda x^1 + (1-\lambda)x^2) < \max\{f(x^1), f(x^2)\}$, for $\forall\ \lambda \in (0,1)$.*

Remark 41 *For strictly quasi-convex function, it can have sadder point / point of inflection, flat spot at the botton, and discontinuity. (So, strictly quasi-convex eliminate flat spot except at the botton from quasi-convex function)*

Remark 42 *Strictly quasi-convex function generally NOT infer quasi-convex function unless we add continuity condition.*

Definition 43 *Let $S \subseteq R^n$ and S is convex set, f is a* ***strongly quasi-convex function***[3] *on S, if, for each $x^1, x^2 \in S$ and $x^1 \neq x^2$, we have $f(\lambda x^1 + (1-\lambda)x^2) < \max\{f(x^1), f(x^2)\}$, for $\forall\ \lambda \in (0,1)$.*

Remark 44 *For strongly quasi-convex function, it can have sadder point / point of inflection and discontinuity. (So, strong quasi-convex eliminate all flat spot from quasi-convex function)*

Theorem 45 *(Equivalence of proving strongly quasi-convex function[4]) Let $S \subseteq R^n$, $S \neq \phi$, and S is convex; let $f : s \to R$. Then the following are equivalent:*

1. f is strong quasi-convex function on S;

2. for each $x^1, x^2 \in S$ and $x^1 \neq x^2$, we have $f(\lambda x^1 + (1-\lambda)x^2) < \max\{f(x^1), f(x^2)\}$, for $\forall\ \lambda \in (0,1)$.

3. for all $x', x \in S$ and $\nabla f(x) \neq 0$, $\nabla f(x)(x' - x) > 0$ whenever $f(x') \geq f(x)$ and $x^1 \neq x^2$ (given f is differentiable);

*4. for each $x \in S$, the Hessian matrix $D^2 f(x)$ is negative definite in the subspace $\left\{z \in R^N : \nabla f(x) \cdot z = 0\right\}$, that is, if and only if $z \cdot D^2 f(x) \cdot z > 0$ whenever $\nabla f(x) \cdot z = 0$ (given f is twice differentiable);[**?**]*

Theorem 46 *If f is strongly quasi-convex function on S. Let $g : S \to R$ be an increasing function. Then $g \circ f$ is strongly quasiconcave function.*

Theorem 47 *If f is strongly quasi-convex function on S and $k > 0$, then kf is strongly quasiconcave function.*

Theorem 48 *(From Wiki)The quasiconvexity property is preserved under*

1. Non-negative weighted maximum: $f = \max\{w_1 f_1, ..., w_n f_n\}$ where $f_1, ..., f_n$ are quasi-convex; $w_1, ..., w_n$ are non-negative;

2. Composition with a non-decreasing function: let $g : R^n \to R$ is quasiconvex, $h : R \to R$ is non-decreasing, then $f = h \circ g$ is quasiconvex[5]

3. Maximization: If Y is a nenempty set and $f(\cdot, y)$ is a quasi-convex function on a convex set X for every $y \in Y$. Then $g(x) = \sup\limits_{y \in Y} f(x, y)$ is a quasi-convex function on X.

4. Minimization: Let $f(x, y) : X \times Y(x) \to R$. If $Y(x)$ is a nonempty set for every $x \in X$, X is convex set, and $(X, Y(x))$ is convex set, $f(x, y)$ is quasi-convex function on $(X, Y(x))$, $g(x) > -\infty$ for $\forall x \in X$. Then $g(x) = \inf\limits_{y \in C} f(x, y)$ in convex on X.

(However, sum of quasiconvex function can be not quasiconvex.)

[2] Some textbook and notes, e.g. MWG P933 and John Nachbar's Finite Dimensional Optimization II, use strongly quasi-convex as definition for strictly quasi-convex function

[3] Some textbook and notes use strongly quasi-convex as definition for strictly quasi-convex function

[4] Some textbook and notes use strongly quasi-convex as definition for strictly quasi-convex function

[5] Different from convex case, here, we only require h be non-decreasing.

1.1.5 Psedo-convex Function

Definition 49 *Let $S \subseteq R^n$, S is convex set, and f is differentiable. f is a* ***pseudo-convex function*** *on S, if, for each $x^1, x^2 \in S$, we have if $\nabla f(x^1)(x^2 - x^1) \geq 0$, then $f(x^2) \geq f(x^1)$.*

Remark 50 *For pseudo-convex function, it can have flat spot at the botton. (So, pseudo-convex function eliminate flat spot except at the botton, eliminate sadder point / point of inflection, and assume continuity from quasi-convex function)*

Theorem 51 *If f is not quasi-convex, then f is not pseudo-convex.*

Definition 52 *Let $S \subseteq R^n$, S is convex set, and f is differentiable. f is a* ***strictly pseudo-convex function*** *on S, if, for each $x^1, x^2 \in S$ and $x^1 \neq x^2$, we have if $\nabla f(x^1)(x^2 - x^1) \geq 0$, then $f(x^2) > f(x^1)$.*

Theorem 53 *Equivalently, f is a strictly pseudo-convex function on S, if, for each $x^1, x^2 \in S$ and $x^1 \neq x^2$, we have if $f(x^2) \leq f(x^1)$, then $\nabla f(x^1)(x^2 - x^1) < 0$.*

Remark 54 *For strictly pseudo-convex function, it can have NO flat spots, points of inflection, and discontinouity. (So, strictly quasi-convex eliminate flat spot, eliminate sadder point / point of inflection, and assume continuity from quasi-convex function)*

Theorem 55 *If f is strictly pseudo-convex function, then f is strong quasi-convex function.*

1.1.6 Relationship between Convex functions

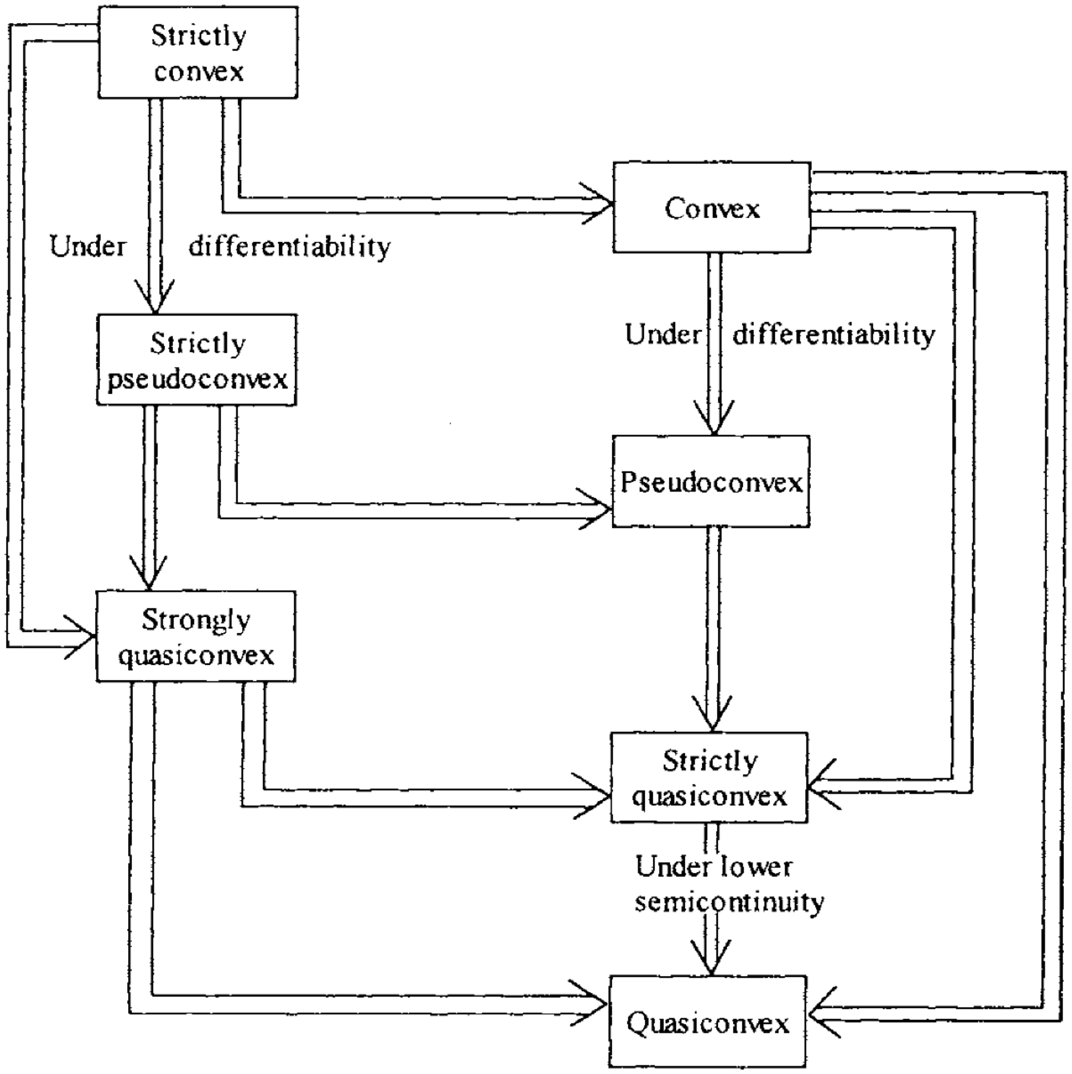

Figure 3.13 Relationship among various types of convexity.

1.1.7 Convexity of functional of convex functions

Definition 56 *A real-valued function ϕ defined on a set $T \subset R^m \times R^k$ is said to be* ***increasing-decreasing*** *on T if and only if for every $(y^1, z^1) \in T$ and $(y^2, z^2) \in T$:*

$$y^2 \geq y^1 \text{ and } z^2 \leq z^1 \text{ imply } \phi(y^2, z^2) \geq \phi(y^1, z^1)$$

Lemma 57 *Let ϕ be a real-valued differentiable function on an open convex set $T \subset R^m \times R^k$. Then ϕ is increasing-decreasing on T iff, for every $(y, z) \in T$*

$$\nabla_y \phi(y, z) \geq 0 \qquad ; \qquad \nabla_z \phi(y, z) \leq 0$$

Theorem 58 *(M. Aveiel, NLP: analysis & Methods, Theorem 6.9) Let $X \subset R^n$ be a convex set, let $f(x) = (f_1(x), ..., f_m(x))$ and $g(x) = (g_1(x), ..., g_k(x))$ be defined on X, and let ϕ be a real-valued function on $R^m \times R^k$. Define*

$$\Phi(x) = \phi(f(x), g(x))$$

and let any one of the following assumptions hold:
i). f is convex, g is concave, ϕ is increasing-decreasing;
ii). f is linear, g is linear;
iii). f is convex, g is linear, ϕ is y-increasing;
iv). f is concave, g is linear, ϕ is y-decreasing;
Then
a). If ϕ is convex, then Φ is convex.
b). If X is open, f and g are differentiable on X, and ϕ is pseudoconvex, then Φ is pseudoconvex.
c). If ϕ is quasiconvex then Φ is quasiconvex.

1.1.8 Properties under optimization

Theorem 59 *Non-negative weighted maximum: $f = \max\{w_1 f_1, ..., w_n f_n\}$ where $f_1, ..., f_n$ are convex; $w_1, ..., w_n$ are non-negative. Then f is convex.*

Theorem 60 *Non-negative weighted maximum: $f = \max\{w_1 f_1, ..., w_n f_n\}$ where $f_1, ..., f_n$ are quasi-convex; $w_1, ..., w_n$ are non-negative. Then f is quasi-convex.*

Theorem 61 *If Y is a nenempty set and $f(\cdot, y)$ is a quasi-convex function on a convex set X for every $y \in Y$. Then $g(x) = \sup\limits_{y \in Y} f(x, y)$ is a quasi-convex function on X.*

Theorem 62 *Let $f(x, y) : X \times Y(x) \to R$. If $Y(x)$ is a nonempty set for every $x \in X$, X is convex set, and $(X, Y(x))$ is convex set, $f(x, y)$ is quasi-convex function on $(X, Y(x))$, $g(x) > -\infty$ for $\forall x \in X$. Then $g(x) = \inf\limits_{y \in C} f(x, y)$ in convex on X. (In Heyman and Sobel, 1984:525, it state the same result with more strong condition by requiring $f(x, y)$ be convex)*

Theorem 63 *(Heyman and Sobel, 1984:525) Let $f(x, y) : X \times Y \to R$. If Y is non-empty and X is convex set, $f(\cdot, y)$ is convex function on a convex set X for each $y \in Y$. Then $g(x) = \sup\limits_{y \in Y} f(x, y)$ in convex on X.*

Chapter 2

Calculus

2.1 Continuity

Theorem 64 *Intermediate value theorem*

Theorem 65 *Extreme value theorem*

Remark 66 *1. The sequence of continuous function does not necessarily pointwise converge to a continuous function; if the sequence converges uniformly, then by uniformly convergence theorem, its limit function is continuous.*

Theorem 67 *1. Sum, product, difference, and quotient (if the denominator is not zero) of continuous functions is continuous.*
2. Composition of continuous functions is continuous.

2.2 Limits

Theorem 68 *(Algebraic limit theorem) If the limits of $f(x)$ and $g(x)$ exist, then*

1. $\lim\limits_{x\to p}(f(x)+g(x)) = \lim\limits_{x\to p} f(x) + \lim\limits_{x\to p} g(x)$
2. $\lim\limits_{x\to p}(f(x)-g(x)) = \lim\limits_{x\to p} f(x) - \lim\limits_{x\to p} g(x)$
3. $\lim\limits_{x\to p}(f(x)\cdot g(x)) = \lim\limits_{x\to p} f(x) \cdot \lim\limits_{x\to p} g(x)$
4. $\lim\limits_{x\to p}(f(x)/g(x)) = \lim\limits_{x\to p} f(x) / \lim\limits_{x\to p} g(x)$
5. $\lim\limits_{x\to p} s\cdot f(x) = s\cdot \lim\limits_{x\to p} f(x)$*, where s is scalar multiplier;*
6. $\lim\limits_{x\to p} s^{f(x)} = s^{\lim\limits_{x\to p} f(x)}$*, where s is a positive real number;*

Proposition 69 *(Limits of Extra Interest) The following results hold:*

1. $\lim\limits_{x\to 0} \frac{\sin x}{x} = 1$
2. $\lim\limits_{x\to 0} \frac{1-\cos x}{x} = 0$

Theorem 70 *(L'Hopital's Rule) If $\lim\limits_{x\to p}(\frac{f(x)}{g(x)})$ has the form of $\frac{0}{0}$ or $\frac{\pm\infty}{\pm\infty}$, then $\lim\limits_{x\to p}(\frac{f(x)}{g(x)}) = \lim\limits_{x\to p}(\frac{f'(x)}{g'(x)})$*

2.3 Derivative and Integral:

2.3.1 Mean Value Theorem

Theorem 71 *(Mean Value Theorem) Let $f : [a, b] \to R$ be a continuous function on the closed interval $[a, b]$, and differentiable on the open interval (a, b), where $a < b$. Then there exists some c in (a, b) such that $f^{'}(c) = \frac{f(b)-f(a)}{b-a}$.*

Theorem 72 *(Cauchy's Mean Value Theorem) Let $f : [a, b] \to R$ and $g : [a, b] \to R$ be a continuous function on the closed interval $[a, b]$, and differentiable on the open interval (a, b), where $a < b$. Then there exists some c in (a, b) such that $\frac{f^{'}(c)}{g^{'}(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}$.*

Theorem 73 *(The First Mean Value Theorem for Integration) Let $f : [a, b] \to R$ be a continuous function on the closed interval $[a, b]$, and let $g : [a, b] \to [0, \infty)$ be a integrable function, where $a < b$. Then there exists some c in $[a, b]$ such that $\int_a^b f(t)g(t)dt = f(c)\int_a^b g(t)dt$.*

(If $g(t) = 1$, then $\int_a^b f(t)dt = f(c)(b-a)$)

Theorem 74 *(The Second Mean Value Theorem for Integration) Let $f : [a, b] \to R$ be a positive monotonically decreasing function on the closed interval $[a, b]$, and let $g : [a, b] \to R$ be a integrable function, where $a < b$. Then there exists some c in $(a, b]$ such that $\int_a^b f(t)g(t)dt = f(a)\int_a^c g(t)dt$.*

Theorem 75 *(The Second Mean Value Theorem for Integration by Hiroshi Okamura) Let $f : [a, b] \to R$ be a monotonic function (not necessarily positive and decreasing) on the closed interval $[a, b]$, and let $g : [a, b] \to R$ be a integrable function, where $a < b$. Then there exists some c in (a, b) such that $\int_a^b f(t)g(t)dt = f(a)\int_a^c g(t)dt + f(b)\int_c^b g(t)dt$.*

2.3.2 Fundamental Theorem of Calculus

Theorem 76 *(The First Fundamental Theorem of Calculus) A real-valued function F is defined on a closed interval $[a, b]$ by setting, for $\forall x \in [a, b]$,*

$$F(x) = \int_a^x f(t)dt$$

where f is a real-valued function continuous on $[a, b]$. Then, F is

1. *continuous on $[a, b]$,*
2. *differentiable on the open interval (a, b),*
3. *$F^{'}(x) = f(x)$.*

(For more general case: if f is any Lebesgue integrable function on $[a, b]$ and x_0 is a number in $[a, b]$ such that f is continuous at x_0, then $F(x) = \int_a^x f(t)dt$ is differentiable for $x = x_0$ with $F^{'}(x_0) = f(x_0)$)

Theorem 77 *(The Second Fundamental Theorem of Calculus) Let f be a real-valued function defined on a closed interval $[a, b]$ that admits an antiderivative F on $[a, b]$. That is, f and F are functions such that for $\forall x \in [a, b]$, $f(x) = F^{'}(x)$. If f is integrable on $[a, b]$ then $\int_a^b f(t)dt = F(b) - F(a)$*

(Notice: if f is continuous, then f is integrable. However, not all integrable f are continuous)

(For more general case: if a real function F on $[a, b]$ admits a derivative $f(x)$ at every point x of $[a, b]$ and if this derivative f is Lebesgue integrable on $[a, b]$, then $\int_a^b f(t)dt = F(b) - F(a)$)

Theorem 78 *(Differentiation under Integral) Let $F(x) = \int_{a(x)}^{b(x)} f(x, t)dt$, then:*

$$\begin{aligned} \frac{d}{dx}F(x) &= \left(\frac{\partial F}{\partial b}\right)\frac{db}{dx} + \left(\frac{\partial F}{\partial a}\right)\frac{da}{dx} + \int_{a(x)}^{b(x)} \frac{\partial f(x,t)}{\partial x}dt \\ &= f(x, b(x))\frac{db(x)}{dx} - f(x, a(x))\frac{da(x)}{dx} + \int_{a(x)}^{b(x)} \frac{\partial f(x,t)}{\partial x}dt \end{aligned}$$

2.3.3 Derivative

Theorem 79 *Differentiation rules:*

- *Sum rule:* $(af+bg)^{'} = af^{'} + bg^{'}$
- *Product Rule:* $(fg)^{'} = f^{'}g + fg^{'}$
- *Quotient Rule: if* $g \neq 0$, *then* $(\frac{f}{g})^{'} = \frac{f^{'}g - fg^{'}}{g^2}$
- *Chain Rule: if* $f(x) = h(g(x))$, *then* $f(x) = h^{'}(g(x)) \cdot g^{'}(x)$,
- *Power Rule:* $(f^g)^{'} = f^g(g^{'} \ln f + \frac{g}{f} f^{'})$
- *Inverse Function Rule:* $(f^{-1})^{'} = (f^{'})^{-1}$ *(or, equivalently,* $Df^{-1}(y) = [Df(x)]^{-1}$*)*
- *Implicity Function Rule: if implicit function* $y(x)$ *is defined as* $F(x, y(x)) = 0$, *then* $y_x^{'} = -\frac{F_x^{'}}{F_y^{'}} = -\frac{\partial F}{\partial x} / \frac{\partial F}{\partial y}$ *(or, equivalently,* $D_x y = -[D_y F(x,y)]^{-1} D_x F(x,y)$*)*

Definition 80 *Let* $f : R^N \to R^M$ *be differentiable, then the* ***Jacobian of*** f ***at*** x^* *denoted by* $Jf(x^*)$, *is the* $M \times N$ *matrix of partial derivatives of* f *at* x^*

$$Jf(x^*) = \begin{bmatrix} D_1 f_1(x^*) & ... & D_N f_1(x^*) \\ ... & ... & ... \\ D_1 f_M(x^*) & ... & D_N f_M(x^*) \end{bmatrix}$$

Definition 81 *Let* $f : R^N \to R$ *be twice differentiable,then the* ***Hessian of*** f ***at*** x^*, *denoted by* $Hf(x^*)$, *is the twice differential matrix of* f *at* x^*

$$Hf(x^*) = \begin{bmatrix} D_{11}^2 f(x^*) & ... & D_{1N}^2 f(x^*) \\ ... & ... & ... \\ D_{N1}^2 f(x^*) & ... & D_{NN}^2 f(x^*) \end{bmatrix}$$

Theorem 82 *(Young's Theorem) Let* $f : R^N \to R$ *be* C^2. *THen the Hessian of* f *is symmetric:* $D_{ij}^2 f(x^*) = D_{ji}^2 f(x^*)$ *for* $\forall i, j$.

Theorem 83 *(Taylor's Theorem) If* $n \geq 0$ *is an integer and* f *is a function which is* n *times continuously differentiable on the closed interval* $[a, x]$, *and* $(n+1)$ *times differentiable on the open interval* (a, x), *then*

$$f(x) = f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x)$$

where $R_n(x)$ *is reminder term, which can be expressed by either one of the following terms:*

- *Lagrange Form:* $R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$ *where* $\xi \in [a, x]$
- *Cauchy Form:* $R_n(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x-a)$ *where* $\xi \in [a, x]$
- *Generazed Cauchy Form:* $R_n(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n \frac{G(x)-G(a)}{G'(\xi)}$ *where* $\xi \in [a, x]$ *and* $G(t)$ *is a continuous function on* $[a, x]$ *with non-vanishing derivative on* (a, x)

Definition 84 *The* ***directional derivative*** *of* f *in the direction of* v *at the point* x *is the limit*

$$D_v f(x) = \lim_{h \to 0} \frac{f(x+hv) - f(x)}{h}$$

Theorem 85 *If all the partial derivatives of* f *exist and are continuous at* x, *then they determine the directional derivative of* f *in the direction* v *by the formula:*

$$D_v f(x) = v \cdot \nabla f(x) = \sum\nolimits_{j=1}^{n} v_j \frac{\partial f}{\partial x_j} = \cos\theta \, \|\nabla f(x)\| \, \|v\|$$

(If $\|v\| = 1$, *then* $D_v f(x) = v \cdot \nabla f(x) = \cos\theta \, \|\nabla f(x)\|$, *where* θ *is the angle between* $\nabla f(x)$ *and* v.*)*

2.3.4 Integral

Theorem 86 *Integral Rules:*

- *Reversing Limits of Integration:* $\int_a^b f(x)dx = -\int_b^a f(x)dx$
- *Integrals over intervals of length zero:* $\int_a^a f(x)dx = 0$
- *Linearity:* $\int_a^b (\alpha f(x) + \beta g(x))dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx$
- *Additivity:* $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$
- *Integral by Parts:* $\int udv = u \cdot v - \int vdu$
- *Integral by substitution:* $\int_a^b f(g(x))dg(x) = \int_{g(a)}^{g(b)} f(x)dx$

Theorem 87 *In equalities for Integrals:*

- *Upper and Lower bounds: if* $m \leq f(x) \leq M$ *for* $\forall x \in [a, b]$*, then*

$$m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$$

- *Inequalities between functions: if* $f(x) \leq g(x)$ *for* $\forall x \in [a, b]$*, then*

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx$$

- *Subintervals: if* $[c, d]$ *is subinterval of* $[a, b]$ *and* $f(x)$ *is non-negative for* $\forall x$*, then*

$$\int_c^d f(x)dx \leq \int_a^b f(x)dx$$

- *Cauchy-Schwarz Inequality:*

$$\left(\int_a^b f(x)g(x)dx\right)^2 \leq \left(\int_a^b (f(x))^2 dx\right)\left(\int_a^b (g(x))^2 dx\right)$$

- *Holder's Inequality: if* p *and* q *are two real numbers:* $1 \leq p, q \leq \infty$ *with* $\frac{1}{p} + \frac{1}{q} = 1$*, then*

$$\left|\int f(x)g(x)dx\right| \leq \left(\int |f(x)|^p dx\right)^{1/p} \left(\int |g(x)|^q dx\right)^{1/q}$$

- *Minkowski Inequality: If* $p \geq 1$ *is a real number, then*

$$\left(\int |f(x) + g(x)|^p dx\right)^{1/p} \leq \left(\int |f(x)|^p dx\right)^{1/p} + \left(\int |g(x)|^p dx\right)^{1/p}$$

2.3.5 Multivariate Differentiation

2.4 Inverse Function Theorem

Theorem 88 *(Inverse Function Theorem) Fix $x^* \in R^n$, let $f : R^n \to R^n$ be C^r, where r is a positive integer, let $y^* = f(x^*)$, and suppose $Df(x^*)$ is invertible. Then there are open sets $U, V \subseteq R^n$, with $x^* \in U$ and $y^* \in V$, such that $Df(x)$ has full rank for all $x \in U$, f maps U 1-1 onto V, and hence has an inverse $f^{-1} : V \to U$. futhermore, f^{-1} is C^r.*

In the result, "Then there are open sets $U, V \subseteq R^n$, with $x^* \in U$ and $y^* \in V$, such that $Df(x)$ has full rank for all $x \in U$, f maps U 1-1 onto V" are inherited from the assumption "Fix $x^* \in R^n$, let $f : R^n \to R^n$ be C^r, where r is a positive integer, let $y^* = f(x^*)$, and suppose $Df(x^*)$ is invertible". The importance of inverse function theorem is the last sentence "and hence has an inverse $f^{-1} : V \to U$. futhermore, f^{-1} is C^r.", which point out the existence of inverse and the continuous of the inverse.

Use Inverse function theorem, we can use chain rule to computer $Df^{-1}(x)$ even if we can not derive f^{-1} explicitly. For example, if $Df(x^*)$ is invertible, $f^{-1}(x)$ is well defined by inverse function theorem. So let $h(x) = f^{-1}(f(x))$, because $f^{-1}(f(x)) = x$, we have $Dh(x) = Df^{-1}(f(x)) \underset{\text{Chain Rule}}{=} Df^{-1}(y)Df(x) = I$. Hence, $Df^{-1}(y) = [Df(x)]^{-1}$.

$Df(x^*)$ of being full rank is not necessary condition for existence of an inverse function, $f^{-1}(x^*)$. However, $Df(x^*)$ of being full rank is necessary and sufficient condition for $f^{-1}(x^*)$ being differntiable.

2.5 Implicit Function Theorem

Theorem 89 *(Implicit Function Theorem) Let O be a nonempty open subset of R^{L+M}. Let $f : O \to R^N$ be C^r, where r is a positive integer. Fix $x^* \in O$ and let $f(x^*) = y^*$.If $Df(x^*)$ has full rank of M (if $M = 1$, then the condition becomes $Df(x^*) \neq 0$), then there is an open set W in R^{L+M} such that the restriction of the level set $f^{-1}(y^*)$ to W is the graph of a C^r function.*

In particular, suppose, for concreteness and simplicity of notation, that the last M columns of $Df(x^*)$ (the x_μ columns) are linearly independent, hence has full rank of M. Then there are open sets $U \subseteq R^L$ and $W \subseteq R^{L+M}$, and a C^r function $\psi : U \to R^M$ such that, $D_\mu f(x)$ has full rank for all $x \in U$, and

1. *$x^*_\lambda \in U, x^* \in W$,*
2. *$\psi(x^*_\lambda) = x^*_\lambda$,*
3. *For any $x \in W, x_\lambda \in U$,*
4. *For any $x_\lambda \in U$, $\psi(x_\lambda)$ is the unique x_μ such that, letting $x = (x_\lambda, x_\mu)$,*
   - *a). $x \in W$,*
   - *b). $f(x) = y^*$.*

The implicity function theorem established the existance of implicity function ψ and the differentiability of ψ, which is C^r. The Implicity Function Theorem thus states that if f is continuously differentiable and the last M columns of $Df(x^*)$ has full rank, then the level set of f through x^* is, near x^*, an L-dimensional suface in R^{L+M}. Hence, we can express f function in terms of L-demension instead of original L+M-demension. Also, by using ψ, we can write the last M variables of x as the a function of the first L variables of $\psi(\cdot) : x_L \to x_M$, where $\psi(\cdot)$ is C^r. Hence, the original variable $x = (x_L, x_M) = (x_L, \psi(x_L))$.

Use Implicity function theorem, we can use the chain rule to calculate the implicite function, $\psi : R^L \to R^M$, even if we can not derive the implicite function, ψ ,explicitly. $D\psi(x_L) = -[D_M f(x)]^{-1} D_L f(x)$.