Game Theory with Applications to Finance and Marketing

Mathematical Review

1. (**Set and Function.**)

   A well-defined collection of objects (called elements) is a set. A set of sets is usually referred to as a *collection*, and a set of collections is usually referred to as a *family*. A set contains nothing is called an empty set and denoted $\emptyset$. A set contains exactly one element is a *singleton*. If two sets A and B are such that B contains each and every element of A, then we say that A is a subset of B and B a *superset* of A, and we write $A \subset B$, and if furthermore B contains an element y which is not contained in A, then A is a *proper subset* of B. Two sets A and B are equal, if $A \subset B$ and $B \subset A$. If one has a *universe set* X in mind (so that all sets under consideration are subsets of X), then define the *complement* of $A \subset X$ to be the set containing exactly those x contained by X but not by A. The complement of A is denoted by A^c. Denote by 2^X the collection containing all subsets of X, which will be referred to as the *power set* of X.

2. Given two sets A and B, an assigning rule f is a (single-valued) function from A into B if $\forall a \in A$, there is $b \in B$ such that $f(a) = b$. In this case, we write $f : A \to B$, and A and B are called the *domain* and *range* of f respectively. If A is itself a collection of sets, then we say that f is a set function. If instead B is a collection of sets then f is a multi-valued function (or a *correspondence*).[1] Given $f : A \to B$, and given any $C \subset A$ and $D \subset B$, the sets $f(C) = \{f(a) : a \in C\}$ and $f^{-1}(D) = \{a \in A : f(a) \in D\}$ are called respectively the *image* of C and the *pre-image* of D under f.[2] The function $f : A \to B$ is *surjective* if $B \subset f(A)$, *injective* if $x, y \in A,\ x \neq y \Rightarrow f(x) \neq f(y)$, and *bijective* if it is both surjective ane injective. A bijective function is also called a *one-to-one correspondence*.

[1] In general we would write for the correspondence $f : A \to \bigcup_{b \in B} b$ instead.

[2] Let $A = B$ be the set of real numbers. Let $f(x) = x^2 + 1$, $C = (-1, 0]$, and $D = [0, \frac{1}{2}]$. What is $f^{-1}(D)$? What is $f^{-1}(f(C))$? Is the latter a subset or a superset of C?

3. A bijective function $f : A \to B$ has an inverse function $f^{-1} : B \to A$ which assigns for each $b \in B$ "the" $a \in A$ that satisfies $f(a) = b$. (Depending on the context, the notation f^{-1} should not create confusion with the pre-image operator.)

4. Two sets A and B are of the same *cardinality* if we can find a one-to-one correspondence between them. A set A is *finite* if it is empty or there is a one-to-one correspondence between A and the set $\{1, 2, \cdots, n\}$ for some $n \in \mathbf{Z}_+$ (and we say in the latter case the cardinality of A, denoted $|A|$ or $\#(A)$, is n), where $\mathbf{Z}_+$ stands for the set of strictly positive integers.[3] A set A is *countably infinite* (*denumerable*) if there is a one-to-one correspondence between A and $\mathbf{Z}_+$. A set which is neither finite nor countably infinite is *uncountable*. A set is *countable* if it is not uncountable.[4]

5. Given an indexed family of sets $\{A_b; b \in \beta\}$, where β is an arbitrary non-empty index set, the *Cartesian product* of these sets, denoted by $\Pi_{b\in\beta} A_b$, is a set containing each and every function $f : \beta \to \bigcup_{b\in\beta} A_b$ satisfying that for all $b \in \beta$, $f(b) \in A_b$.[5] The union of sets $\{A_b; b \in \beta\}$, denoted $\bigcup_{b\in\beta} A_b$, contains exactly those x contained by some A_b. The intersection of these sets, denoted $\bigcap_{b\in\beta} A_b$, contains exactly those x contained by all A_b's. Two sets are *disjoint* if they have an empty intersection.

6. A collection $\{A_b; b \in \beta\}$ of subsets of X is called a *partition* of X if $\bigcup_{b\in\beta} A_b = X$ and $\forall a, b \in \beta,\ a \neq b \Rightarrow A_a \bigcap A_b = \emptyset$.

---

[3]The notation $\mathbf{Z}_+$ is taken from James R. Munkres, 1975, *Topology: A First Course*, New Jersey: Prentice-Hall. It represents the set of natural numbers. We shall interchangeably use the notation $\mathbf{N}$ to stand for the same set, as in Kai Lai Chung, 1974, *A Course in Probability Theory*.

[4]The following three statements are apparently equivalent: (i) A is countable; (ii) there is a surjective function $f : \mathbf{Z}_+ \to A$; (iii) there is an injective function $g : A \to \mathbf{Z}_+$. Because of (ii) a subset of a countable set is apparently countable. Because of (iii), $A = \mathbf{Z}_+^2$ is countable: simply define $g(n, m) = 2^n 3^m$. Now since the set of positive rationals has a one-to-one correspondence with $\mathbf{Z}_+^2$, the former is countable. The following three statements are also equivalent, but less apparent: (a) A is infinite; (b) there exists an injective function $f : \mathbf{Z}_+ \to A$; (c) there exists a bijective function of A with a *proper* subset of itself.

[5]For example, Suppose that $\beta = \{1, 2\}$, $A_1 = \{x, y\}$, $A_2 = \{a, b, c\}$. Then $\Pi_{b\in\beta} A_b = A_1 \times A_2 = \{(x, a), (x, b), (x, c), (y, a), (y, b), (y, c)\}$, where, for example, (x, a) is a short-hand notation for $f(x) = a$. Similarly, if instead $\beta = \{2, 1\}$, then $\Pi_{b\in\beta} A_b = A_2 \times A_1 = \{(a, x), (a, y), (b, x), (b, y), (c, x), (c, y)\}$.

7. DeMorgan's Law says that

$$[\bigcup_{b\in\beta} A_b]^c = \bigcap_{b\in\beta} A_b^c,\ [\bigcap_{b\in\beta} A_b]^c = \bigcup_{b\in\beta} A_b^c. \tag{1}$$

8. The following facts are well known. A countable union or a finite product of countable sets is countable. A infinite product of countable sets need not be countable; in particular, the power set of $\mathbf{Z}_+$ is uncountable.

9. (**Linear Space or Vector Space.**)

10. A set V is a (real) linear space on $\Re$, where $\Re$ is the set of real numbers, if it is equipped with and closed under operations of vector addition and scalar multiplication together with a number of well-defined operational (such as commutative and associative) laws. Elements in V are called *vectors*. Given a finite number of elements $x_1, x_2, \cdots, x_n$ in V, an element $y \in V$ is a *linear combination* of these elements if there exist real numbers $a_1, a_2, \cdots, a_n$ such that $y = \sum_{i=1}^n a_i x_i$. A subset B of V is said to span V if all vectors in V are linear combinations of elements of B. If no proper subset of B can span B, then B is linearly independent, and in this case B is a *basis* for V if B spans V. The cardinality of (the number of elements in) a basis (any basis) is called the linear space V's *dimension*. V may be finite-dimenional or infinite-dimensional depending on whether the cardinality of B is finite or infinite.

11. Let V_x be a real linear space with origins 0_x. A function $f : V_x \to \Re$ is *concave* (respectively *convex*, and *affine*) if for any two elements $\mathbf{x}, \mathbf{x}' \in V_x$ and for any $a \in [0, 1]$,

$$f(a\mathbf{x} + (1-a)\mathbf{x}') \geq \text{ (respectively } \leq, =) \ af(\mathbf{x}) + (1-a)f(\mathbf{x}').$$

The function f is said to be *linear*, if it is affine and $f(\mathbf{0}_x) = 0$.

12. A subset A of the linear space V is a *convex set* if for all $x, y \in A$ and for all real numbers $\lambda \in [0, 1]$, $\lambda x + (1-\lambda)y \in A$. The intersection of a collection of convex sets is itself convex.

13. (**Metric Space and Topology.**)

Given any non-empty set X a function $d : X \times X \to \mathcal{R}_+$ (where $\mathcal{R}_+$ or interchangeably $\Re_+$ denotes the set of non-negative real numbers) is a *metric* if
(i) $d(x, y) = 0 \Leftrightarrow x = y, \ \forall x, y \in X$;
(ii) $d(x, y) = d(y, x) \geq 0, \ \forall x, y \in X$;
(iii) $d(x, y) + d(y, z) \geq d(x, z)$.
The pair (X, d) is then a *metric space*. For example, let $X = \Re$, which is the set of real numbers, and define for all $x, y \in X$, $d'(x, y) = 1$ if $x \neq y$ and $d'(x, y) = 0$ if otherwise. Define $d(x, y) = |x - y|$ for all $x, y \in X$ also. Then (X, d) and (X, d') are both metric spaces, but they are different metric spaces.

14. A function $x : \mathbf{Z}_+ \to X$ is called a sequence in (X, d). It is said to *converge to* $z \in X$ if for all real $\epsilon > 0$ there exists $n(\epsilon) \in \mathbf{Z}_+$ such that $n \geq n(\epsilon) \Rightarrow d(x(n), z) < \epsilon$. A sequence is *Cauchy* if for all $e > 0$, there exists $n(e) \in \mathbf{Z}_+$ such that $m, n \geq n(e) \Rightarrow d(x(m), x(n)) < e$.

15. Given two metric spaces (X, d) and (Y, d'), a function $f : X \to Y$ is *continuous* if $\{f(x(n))\}$ is a convergent sequence in (Y, d') whenever $\{x(n)\}$ is a convergent sequence in (X, d).

16. Given a metric space (X, d), for all $z \in X$ and $e \in \mathcal{R}_{++}$,[6] the subsets $B(z, e) = \{x \in X : d(x, z) < e\}$ are called *open balls* in X. A subset G of X is *open* if it can be represented as a (perhaps empty) union of open balls. A subset F of X is *closed* if F^c is open. The *closure* of $G \subset X$ is the smallest (with respect to the order relation "inclusion" on X) closed set containing G, and the *interior* of F is the largest open set contained in F. A point $x \in X$ is a *limit point* of $A \subset X$ if for all $e > 0$, $B(x, e) \bigcap A \bigcap \{x\}^c \neq \emptyset$. The set of limit points of A, denoted A', is the *derived set* of A. A point $x \in A$ is an *isolated point* if for some $e > 0$, $B(x, e) \bigcap A \bigcap \{x\}^c = \emptyset$. A closed subset A of X is *perfect* if it contains no isolated points. The *boundary* of A is the intersection of the closure of A and the closure of A^c in (X, d). The collection τ of all the open sets on X is called a *topology* on X. The pair (X, τ) is called a *topological space*.

[6]Here $\mathcal{R}_{++}$, or interchangeably $\Re_{++}$, denotes the set of strictly positive real numbers.

17. A set A contained in the metric space (X, d) is *bounded* if for some $x \in X$ and $e \in \mathcal{R}_{++}$, $A \subset B(x, e)$.

18. Take the linear space $(\Re, \|\cdot\|)$ for example, where for all $x, y \in \Re$, the norm or the length of vector x, denoted by $\|x\|$, is such that $\|x\| = |x|$ and we can define a metric from the norm by $d(x, y) = |x - y|$, which generates the so-called *standard topology* on $\Re$. In this metric space $(\Re, d)$, an open ball is an open interval, and an open set is a countable union of pair-wise disjoint open intervals. (Why countable, and why pair-wise disjoint?)

    - Let $A = \{x\}$ for some $x \in \Re$. Is x a limit point of A? Is x an isolated point of A? Is A a closed set?
    - Suppose that $\{x_n; n \in \mathbf{Z}_+\}$ is a sequence in $\Re$, which converges to $x_0 \in \Re$. Let X contains exactly all the x_n's in the convergent sequence. Is x_0 a limit point of X? If this is not always true, then offer a condition that ensures this conclusion.

19. Let (X, τ) be a topological space. A subset γ of τ is an open covering of X if $X \subset \bigcup_{U \in \gamma} U$. X is *compact* if every open covering of X contains a finite sub-covering of X; that is, if for each open covering γ of X, there exists a finite number of member sets $U_1, U_2, \cdots, U_n \in \gamma$, such that $X \subset \bigcup_{i=1}^{n} U_i$.

    Note that $\Re$ with the standard topology is not compact, for the open covering $\{(n, n+2) : n \in \mathbf{Z}\}$ does not have a finite sub-covering of $\Re$. (We shall see that this is because $\Re$ is not bounded.) Similarly, the interval $(0, 1)$ is not a compact subset of $\mathcal{R}$ under the standard topology, for the collection $\{(\frac{1}{n+2}, \frac{1}{n}) : n \in \mathbf{Z}_+\}$ is an open covering of $(0, 1)$ which does not have a finite sub-covering. (We shall see that this is because $(0, 1)$ is not closed.) It can be shown that compact subsets of $\Re$ under the standard topology are exactly those subsets which are both bounded and closed.

20. The following statements are true.

    - A closed subset of a compact space is compact.
    - A (finite or infinite) Cartesian product of compact spaces is compact.

- Consider the metric space $(\Re^n, d)$, where the Euclidean metric d is such that for $x, y \in \Re^n$, with $x = (x_1, x_2, \cdots, x_n)$ and $y = (y_1, y_2, \cdots, y_n)$, we have $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$. Then under the metric topology d, $A \subset \Re^n$ is compact if and only if it is both closed and bounded.
- Is the first orthant in $\Re^2$ closed? Is it bounded?
- Given two topological spaces (X, τ_x) and (Y, τ_y) and a continuous function $f : (X, \tau_x) \to (Y, \tau_y)$, $f(A)$ is compact in (Y, τ_y) whenever A is compact in (X, τ_x). That is, a continuous image of a compact space is compact.

21. (**Real Sequence and Real Function**.)

    If $A = \mathbf{Z}_+$ and $B = \Re$, then (the image of) $f : A \to B$ is called a real (real number) sequence, and we write $f(n)$ as f_n for all $n \in \mathbf{Z}_+$. If $h : \mathbf{Z}_+ \to \mathbf{Z}_+$ is increasing and $x : \mathbf{Z}_+ \to \Re$, then $\{x(h(n))\}$ is called a *subsequence* of $\{x(n)\}$.

22. We say that $l \in \Re$ is a *limit* of the sequence $\{x_n\}$ if for all $\epsilon > 0$ there is an $N(\epsilon) \in \mathbf{Z}_+$ such that

    $$n \in \mathbf{Z}_+,\ n \geq N(\epsilon) \Rightarrow |x_n - l| < \epsilon.$$

    In this case we say that $\{x_n\}$ *converges to* l and $\{x_n\}$ a *convergent sequence*. The limit of a sequence so defined is unique.

23. A set $A \subset \Re$ is bounded above by $u \in \Re$, if for all $x \in A, x \leq u$. In this case, u is called an upper bound of the set A. The least upper bound of A, if it exists, is denoted by $\sup A$. Lower bounds are similarly defined. The greatest lower bound of A, if it exists, is denoted by $\inf A$. The set A is *bounded* if it has an upper and a lower bound in $\Re$.

24. Given a sequence $\{x_n\}$ in $\Re$, define its limit inferior as $\overline{\lim} x_n = \inf_n \sup_{k \geq n} x_k$ and its limit superior as $\underline{\lim} x_n = \sup_n \inf_{k \geq n} x_k$, which will be shown to always exist in $\Re \bigcup \{-\infty, +\infty\}$, and we can show that they are equal if and only if the real sequence $\{x_n\}$ converges.

25. A set $I \subset \Re$ is an *interval* if it is not a singleton (a one-point set) and if $x, y \in I,\ x < y$ implies that there exists some $z \in I$ such that $x < z < y$.

26. A function $g : I \to \Re$, where I is some interval, is *continuous* if for all $x \in I$ and all $\{x_n\}$ converging to x in I, $\{g(x_n)\}$ is a sequence converging to $g(x)$. The *derivative* of $g(\cdot)$ at $x \in I$, if it exists (i.e., it is a finite number), is the (common) limit of $\frac{g(x_n)-g(x)}{x_n-x}$ for all $\{x_n\}$ converging to x. If the derivative of g is defined for all $x \in I$, written $g'(x)$, we say that g is *differentiable* on I. In this case, $g'(\cdot)$ itself is a mapping from I into $\mathcal{R}$. If g' is a continuous function, we say that g is *continuously differentiable* on I.

27. (**Axiom of Continuity**.) If the real sequence $\{x_n\}$ is increasing (i.e. $x_{n+1} \geq x_n$, for all $n \in \mathbf{Z}_+$), and if $x_n \leq M \in \Re$ for all $n \in \mathbf{Z}_+$, then $\{x_n\}$ converges to some $l \leq M$ and $x_n \leq l$ for all $n \in \mathbf{Z}_+$.

28. (**Nested Interval Theorem**.) Suppose that $\{[a_n, b_n]; \mathbf{Z}_+\}$ is a sequence of closed intervals such that for all n, $[a_{n+1}, b_{n+1}] \subset [a_n, b_n]$ and $\lim_{n\to+\infty}(b_n - a_n) = 0$. Then, $\bigcap_{n\in\mathbf{Z}_+}[a_n, b_n]$ is a singleton.

29. (**Least Upper Bound Property.** of $\Re$.) Suppose that A is a non-empty subset of $\Re$, and A is bounded above by some $u \in \Re$. Then $\sup A$ exists. Similarly, if A is non-empty and bounded below, then $\inf A$ exists.

30. Define the extended real line $\overline{\Re} \equiv \Re \bigcup \{+\infty, -\infty\}$. Define $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$. If $A \subset \Re$ is unbounded above (there exists no upper bound for A in $\Re$), then define $\sup A = +\infty$. If $A \subset \Re$ is unbounded below (there exists no lower bound for A in $\Re$), then define $\inf A = -\infty$. In this way, all subsets of $\Re$ have well-defined suprema and infima in $\overline{\Re}$.

31. Suppose that $\{x_n; n \in \mathbf{Z}_+\}$ is a real sequence. Then its limit inferior

$$\underline{\lim} x_n \equiv \sup_{n\in\mathbf{Z}_+} \inf_{k\geq n} x_k$$

and its limit superior

$$\overline{\lim} x_n \equiv \inf_{n\in\mathbf{Z}_+} \sup_{k\geq n} x_k$$

both exist in extended real line $\overline{\Re}$.

32. For any two real sequences $\{x_n\}$ and $\{y_n\}$, we have

$$\underline{\lim} x_n + \underline{\lim} y_n \leq \underline{\lim}(x_n + y_n) \leq \overline{\lim} x_n + \underline{\lim} y_n$$

$$\leq \overline{\lim}(x_n + y_n) \leq \overline{\lim} x_n + \overline{\lim} y_n.$$

33. (**Bolzano-Weierstrass Theorem**.) Every bounded sequence in $\Re$ contains a convergent subsequence.

34. A function $f : A \subset \Re \to \Re$ is *continuous* at $x \in A$ if for all $\epsilon > 0$, there exists $\delta(x, \epsilon) > 0$ such that

$$\forall y \in A,\ |x - y| < \delta(x, \epsilon) \Rightarrow |f(x) - f(y)| < \epsilon.$$

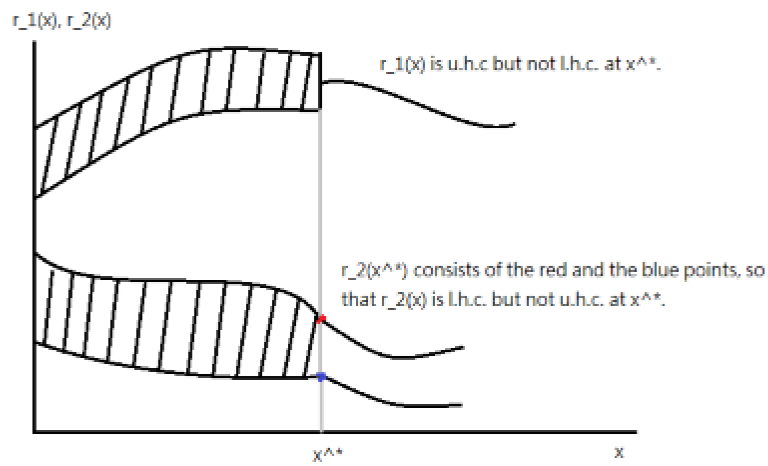
The function is *uniformly continuous* at x if δ can be chosen to be independent of x. We say f is a (uniformly) continuous function on A if it is (uniformly) continuous at x for all $x \in A$. We can show that if $f : [a, b] \to \Re$ is continuous on $[a, b]$, then it is uniformly continuous on $[a, b]$.

35. (**Intermediate Value Theorem**.) If $f : I \to \Re$ is continuous where I is an interval, then $f(I)$ is either a singleton or an interval.

36. (**Extreme Value Theorem**.) If $f : [a, b] \to \Re$ is continuous then $f([a, b])$ is either a singleton or a bounded closed interval. In particular, there exist $m, M \in I$ such that for all $x \in I$, $f(m) \leq f(x) \leq f(M)$.

37. Suppose that $r : X \to Y$ is a correspondence (a multi-valued function), where X and Y are some subsets of $\Re^n$.

    - If $r(x) \neq \emptyset$ for all $x \in X$, then we say that $r(\cdot)$ is non-empty.
    - If $r(x)$ is a convex (compact, closed) subset of $\Re^n$ for all $x \in X$, then we say that $r(\cdot)$ is convex-valued (compact-valued, closed-valued), or simply convex.



- We say that $r(\cdot)$ is upper hemi-continuous (u.h.c.) at x if for every open set V containing $r(x)$, there exists an open set U containing x such that $a \in U \Rightarrow r(a) \subset V$.[7] We say that $r(\cdot)$ is u.h.c. if it is u.h.c. at each and every $x \in X$.
- Suppose that $r(\cdot)$ is closed. If $r(\cdot)$ is u.h.c. at x, then (Θ) whenever $\{x_n\}$ converges to x and $\{y_n \in r(x_n)\}$ converges to y, we have $y \in r(x)$. If Y is compact, then $r(\cdot)$ is u.h.c. at x if (Θ) holds.
- We say that $r(\cdot)$ is lower hemi-continuous (l.h.c.) at $x \in X$ if for all open set $V \subset Y$ such that $V \bigcap r(x) \neq \emptyset$, there exists some open set $U \in X$ such that $a \in U \Rightarrow r(a) \bigcap V \neq \emptyset$. We say that $r(\cdot)$ is l.h.c. if it is l.h.c. at each and every $x \in X$.
- It can be shown that $r(\cdot)$ is lower hemi-continuous (l.h.c.) at $x \in X$ if and only if whenever $\{x_n\}$ converges to x, for each $y \in r(x)$, there exist a subsequence $\{x_{n(k)}\}$ of $\{x_n\}$ and a sequence

[7] See:
https://en.wikipedia.org/wiki/Hemicontinuity#Sequential_characterization.

$\{y_{n(k)} \in r(x_{n(k)})\}$ such that the latter sequence $\{y_{n(k)}\}$ converges to y.

- If r is u.h.c. and compact-valued, then $r(A)$ is compact in Y if A is compact in X.
- The correspondence $r(\cdot)$ is said to be continuous if it is both upper and lower hemi-continuous.

38. **Concave Maximization.**

39. A (real-valued) symmetric matrix $\mathbf{A}_{n\times n}$ is **positive definite** (or PD), if for all $\mathbf{x}_{n\times 1} \neq \mathbf{0}_{n\times 1}$, we have $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$. A symmetric matrix $\mathbf{A}_{n\times n}$ is **negative definite** (or ND), if $-\mathbf{A}$ is positive definite. A symmetric matrix $\mathbf{A}_{n\times n}$ is **positive semi-definite** (or PSD), if for all $\mathbf{x}_{n\times 1} \in \mathcal{R}^n$, we have $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$. A symmetric matrix $\mathbf{A}_{n\times n}$ is **negative semi-definite** (or NSD), if $-\mathbf{A}$ is positive semi-definite.

40. Consider a twice differentiable function $f : \mathcal{R}^n \to \mathcal{R}$. Let the $Df : \mathcal{R}^n \to \mathcal{R}^n$ be the vector function

$$Df = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix},$$

which will be referred to as the *gradient* of f. Let $D^2 f : \mathcal{R}^n \to \mathcal{R}^{n^2}$ be the matrix function

$$D^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix},$$

which will be referred to as the *Hessian* of f.

41. A function $f : \mathcal{R}^n \to \mathcal{R}$ is *concave*, if for all $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$ and all $\lambda \in [0,1]$, $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$. A concave function is said to be strictly concave if the above defining inequality is always strict. A function f is (strictly) *convex* if $-f$ is (strictly) concave. A function is *affine* if it is both concave and convex. An affine function is *linear* if it passes through the origin; that is, if $f(\mathbf{0}_{n\times 1}) = 0$. Note that by definition, f is concave if f is strictly concave.

Theorem 1 *A twice differentiable function $f : U \subset \mathcal{R}^n \to \mathcal{R}$ is* concave*, where U is an open convex subset of $\mathcal{R}^n$,* **if and only if** *$D^2 f$ is a negative semi-definite matrix at each and every $\mathbf{x} \in U$; and f is* strictly concave **if** *$D^2 f$ is a negative definite matrix at each and every $\mathbf{x} \in U$.*[8]

Suppose that $n = 1$ in the above theorem. Then $f'' \leq 0$ if f is concave, and if $\tilde{x}$ is a random variable with finite expectation, then we have $f(E[\tilde{x}]) \geq E[f(\tilde{x})]$. (This is the so-called Jensen's inequality.)

42. **Execise 1** *Define the function*

$$f(x,y) = x^a y^{1-a}, \ \forall x, y \in (0, +\infty),$$

where a is a constant with $0 < a < 1$. Clearly $f : (0,+\infty) \times (0,+\infty) \to \Re$ is twice continuously differentiable. Let us verify that f is strictly concave. Note that

$$D^2 f = a(1-a)x^{a-2}y^{-1-a} \begin{bmatrix} -y^2 & xy \\ xy & -x^2 \end{bmatrix}.$$

Hence for any

$$\begin{bmatrix} k \\ h \end{bmatrix} \in \Re^2,$$

[8] If $f(x) = -x^4$, then $f''(0) = 0$ (so that the Hessian of f at zero is negative semi-definite but not negative definite), but f is strictly concave. Hence the "only if" part for strict concavity in general does not hold. However, if $f : U \subset \Re \to \Re$ is twice continuously differentible and concave (so that U becomes an open interval), and if f'' is not constantly zero over any subinterval in U, then f is strictly concave; see the following link: people.exeter.ac.uk/dgbalken/BME05/LectTwo.pdf.

we have

$$\begin{bmatrix} k & h \end{bmatrix} D^2 f \begin{bmatrix} k \\ \\ h \end{bmatrix}$$

$$= a(1-a)x^{a-2}y^{-1-a} \begin{bmatrix} k & h \end{bmatrix} \begin{bmatrix} -y^2 & xy \\ \\ xy & -x^2 \end{bmatrix} \begin{bmatrix} k \\ \\ h \end{bmatrix}$$

$$= -a(1-a)x^{a-2}y^{-1-a}(ky-hx)^2 \leq 0,$$

and the last inequality is strict unless $k = h = 0$. *This proves that* $D^2 f$ *is negative definite for all* $x, y \in (0, +\infty)$, *and hence* $f(\cdot,\cdot)$ *is strictly concave.*

Execise 2 *Define the function*

$$f(x) = \sqrt{x}, \quad \forall x > 0.$$

Clearly $f : (0, +\infty) \to \Re$ *is twice continuously differentiable. Note that*

$$D^2 f = \frac{\partial^2 f}{(dx)^2} = -\frac{1}{4x^{\frac{3}{2}}},$$

so that for all $k \in \Re$, *we have*

$$k \times D^2 f \times k = -\frac{k^2}{4x^{\frac{3}{2}}} \leq 0,$$

and the last inequality is strict unless $k = 0$. *This proves that* $D^2 f$ *is negative definite for all* $x > 0$, *and hence* $f(\cdot)$ *is strictly concave.*

Execise 3 *Define the function*

$$f(x, y) = g(x) + h(y),$$

where $g, h : \Re \to \Re$ *are two strictly concave twice-differentiable functions. Then* $f(\cdot,\cdot)$ *is a strictly concave function also. Indeed, we have in this case*

$$D^2 f = \begin{bmatrix} g''(x) & 0 \\ \\ 0 & h''(y) \end{bmatrix},$$

so that given any

$$\begin{bmatrix} a \\ b \end{bmatrix} \in \Re^2,$$

we have

$$\begin{bmatrix} a & b \end{bmatrix} D^2 f \begin{bmatrix} a \\ b \end{bmatrix}$$

$$= a^2 g''(x) + b^2 h''(y) \leq 0,$$

and the last inequality holds strictly except in the case where $a = b = 0$. *This proves that* $D^2 f$ *is negative definite, and hence* $f(\cdot, \cdot)$ *is strictly concave.*

The preceding example shows that a twice-differentiable additively separable function

$$f(x_1, x_2, \cdots, x_n) = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n)$$

is strictly concave if for all j, $f_j : \Re \to \Re$ is strictly concave and twice differentiable.

Theorem 2 *Suppose that* $f : \Re^n \to \Re$ *is twice-differentiable and concave, then for all* $\mathbf{x}, \mathbf{a} \in \Re^n$, *we have*[9]

$$f(\mathbf{x}) - f(\mathbf{a}) \leq Df(\mathbf{a})'(\mathbf{x} - \mathbf{a}).$$

The above inequality becomes strict if the Hessian of f *is everywhere negative definite and* $\mathbf{x} \neq \mathbf{a}$.

Proof. Recall the *Theorem of Taylor Expansion with a Remainder*: if $f : \Re^n \to \Re$ is twice continuously differentiable, then for each $\mathbf{x}, \mathbf{a} \in \Re^n$, there exists some $\mathbf{y}$ lying on the line segment connecting $\mathbf{x}$ and $\mathbf{a}$ such that

$$f(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})'(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})' D^2 f(\mathbf{y})(\mathbf{x} - \mathbf{a}).$$

[9]The reader should draw a graph for these inequalities for the case of $n = 1$.

Now suppose that f is concave, so that $D^2 f(\mathbf{y})$ is negative semi-definite. Then we have

$$\frac{1}{2}(\mathbf{x}-\mathbf{a})' D^2 f(\mathbf{y})(\mathbf{x}-\mathbf{a}) \leq 0,$$

so that

$$f(\mathbf{x}) - f(\mathbf{a}) \leq Df(\mathbf{a})'(\mathbf{x}-\mathbf{a}).$$

If $D^2 f(\mathbf{y})$ is negative definite, then given $\mathbf{x} \neq \mathbf{a}$, we have

$$\frac{1}{2}(\mathbf{x}-\mathbf{a})' D^2 f(\mathbf{y})(\mathbf{x}-\mathbf{a}) < 0,$$

implying that

$$f(\mathbf{x}) - f(\mathbf{a}) < Df(\mathbf{a})'(\mathbf{x}-\mathbf{a}). \qquad \text{Q.E.D.}$$

43. According to the *Theorem of Taylor Expansion with a Remainder*, when moving from $\mathbf{a}$ to $\mathbf{x} \in B(\mathbf{a}, r)$ in f's domain of definition, the functional value of f would increase if

$$Df(\mathbf{a})^T(\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T D^2 f(\mathbf{y})(\mathbf{x}-\mathbf{a}) > 0.$$

Note that when r is small enough, the preceding inequality holds if and only if[10]

$$Df(\mathbf{a})^T(\mathbf{x}-\mathbf{a}) = \|Df(\mathbf{a})\| \|\mathbf{x}-\mathbf{a}\| \cos(\theta) > 0,$$

where θ is the angle between the two vectors $Df(\mathbf{a})$ and $\mathbf{x}-\mathbf{a}$, which must be acute to ensure the above inequality.[11]

[10]Imagine that we replace $(\mathbf{x}-\mathbf{a})$ by $\epsilon(\mathbf{x}-\mathbf{a})$, and observe that when $\epsilon > 0$ is a tiny number, $\epsilon|Df(\mathbf{a})^T(\mathbf{x}-\mathbf{a})|$ is much greater than $\epsilon^2|\frac{1}{2}(\mathbf{x}-\mathbf{a})^T D^2 f(\mathbf{y})(\mathbf{x}-\mathbf{a})|$. Thus $\epsilon Df(\mathbf{a})^T(\mathbf{x}-\mathbf{a}) + \epsilon^2[\frac{1}{2}(\mathbf{x}-\mathbf{a})^T D^2 f(\mathbf{y})(\mathbf{x}-\mathbf{a})]$ is positive (respectively, negative) if and only if $Df(\mathbf{a})^T(\mathbf{x}-\mathbf{a})$ is positive (respectively, negative).

[11]Recall the following *Law of Cosines*: Let ΔABC be a triangle whose sides a, b, c are such that a is opposite A, b is opposite B and c is opposite C. Let D be such that the two line segments $\overline{BD}$ and $\overline{CD}$ are orthogonal to each other. Then, since the length of $\overline{BD}$ is $a \sin C$ and the length of $\overline{AD}$ is $|b - a\cos C|$, we have $c^2 = a^2 + b^2 - 2ab\cos C$. Now, given two vectors $\mathbf{v}, \mathbf{w}$ and the angle θ between them, if we let $a = \|\mathbf{v}\|$, $b = \|\mathbf{w}\|$ and $c = \|\mathbf{v}-\mathbf{w}\|$, then it follows from the *Law of Cosines* that

$$\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2(\mathbf{v}\cdot\mathbf{w}) = (\mathbf{v}-\mathbf{w})\cdot(\mathbf{v}-\mathbf{w}) = c^2 = a^2 + b^2 - 2ab\cos\theta = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\|\|\mathbf{w}\|\cos(\theta),$$

44. We shall from now on confine attention to twice differentiable functions.

A necessary condition for $\mathbf{x}^* \in \Re^n$ to solve the following maximization program (P1)

$$\max_{\mathbf{x}\in\Re^n} f(\mathbf{x})$$

is that $Df(\mathbf{x}^*) = 0$, which will be referred to as *the first-order conditions* for the optimal solution $\mathbf{x}^*$. This necessary condition is also sufficient, if f is concave.

45. Consider the following maximization program (P2):

$$\max_{\mathbf{x}\in\Re^n} f(\mathbf{x})$$

subject to

$$\forall i = 1, 2, \cdots, m, \quad g_i(\mathbf{x}) = 0,$$

where $m < n$.

Theorem 3 *(Lagrange Theorem) Suppose that* $\mathbf{x}^*$ *solves (P2) and* $\{Dg_j(\mathbf{x}^*); j \in \{1, 2, \cdots, m\}$ *is a set of linearly independent gradient vectors. (This is known as a* **constraint qualifications** *condition.) Then, there must exist m constants (called the* **Lagrange multipliers***)* $\pi_1, \pi_2, \cdots, \pi_m$ *such that*

(i) $\forall i = 1, 2, \cdots, m, \quad g_i(\mathbf{x}^*) = 0$*; and*

(ii)

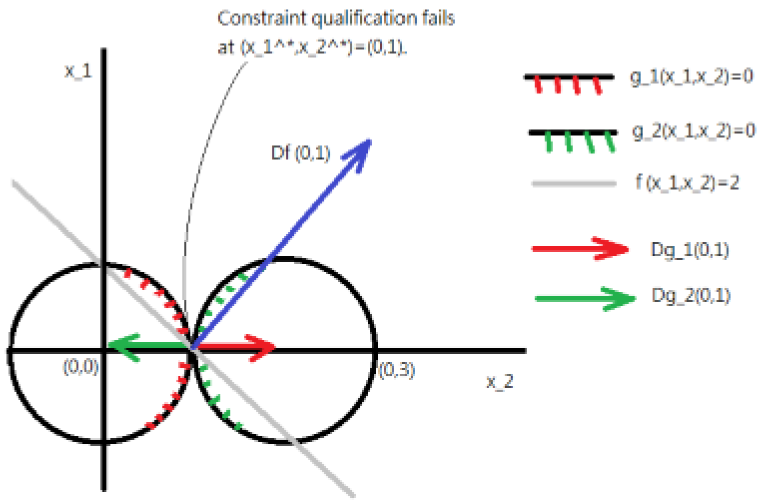
$$\sum_{i=1}^{m} \pi_i Dg_i(\mathbf{x}^*) = Df(\mathbf{x}^*).$$

Conversely, if f *is concave and all* g_i*'s are affine, then if* $\mathbf{x}^*$ *satisfies (i) and (ii), then* $\mathbf{x}^*$ *solves (P2). If* f *is strictly concave, such* $\mathbf{x}^*$ *is unique.*

---

implying that the following *Cosine Formula for Dot Product* must hold:

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos(\theta).$$



To see the role of the **constraint qualifications** condition in the above necessary condition, consider the following example with $n = m = 2$: $f(x_1, x_2) = x_1 + x_2$, $g_1(x_1, x_2) = x_1^2 + x_2^2 - 1$, $g_2(x_1, x_2) = x_1^2 + (x_2 - 2)^2 - 1$. There is only one feasible solution in this case, namely, $(x_1, x_2) = (0, 1)$, which is apparently the optimal solution. Thus $\mathbf{x}^* = (0, 1)$. However, there does not exist $\pi_1, \pi_2 \in \Re$ such that $Df(\mathbf{x}^*) = \pi_1 Dg_1(\mathbf{x}^*) + \pi_2 Dg_2(\mathbf{x}^*)$. This happens because $Dg_1(\mathbf{x}^*)$ and $Dg_2(\mathbf{x}^*)$ are linearly dependent!

46. Now, let us prove the sufficiency of Lagrange Theorem. Suppose that there exists $\mathbf{x}^*$ satisfying
(i) $\forall i = 1, 2, \cdots, m, \quad g_i(\mathbf{x}^*) = 0$; and
(ii)
$$\sum_{i=1}^{m} \pi_i Dg_i(\mathbf{x}^*) = Df(\mathbf{x}^*),$$
for some $\pi_1, \pi_2, \cdots, \pi_m \in \Re$. Consider any other $\mathbf{x}$ satisfying

$$\forall i = 1, 2, \cdots, m, \quad g_i(\mathbf{x}) = 0.$$

We must show that

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

To see that this is true, first note that for all $i = 1, 2, \cdots, m$, by the fact that g_i is affine, we have $\mathbf{a}_i^T\mathbf{x} + b_i = g_i(\mathbf{x}) = 0 = g_i(\mathbf{x}^*) = \mathbf{a}_i^T\mathbf{x}^* + b_i$, where $\mathbf{a}_i$ is the gradient of g_i, so that

$$0 = g_i(\mathbf{x}) - g_i(\mathbf{x}^*) = \mathbf{a}_i^T(\mathbf{x} - \mathbf{x}^*) = Dg_i(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*).$$

Now, observe that

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq Df(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) = \sum_{i=1}^{m} \pi_i Dg_i(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) = 0,$$

and hence $\mathbf{x}^*$ is indeed an optimal solution to (P2). Finally, if both $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are optimal solutions to (P2), then $\mathbf{x}' \equiv \frac{1}{2}(\mathbf{x}^* + \mathbf{x}^{**})$ must be feasible also (i.e., $g_i(\mathbf{x}') = 0$ for all i), and yet by Jensen's inequality we have

$$f(\mathbf{x}') > \frac{1}{2}[f(\mathbf{x}^*) + f(\mathbf{x}^{**})] = f(\mathbf{x}^*),$$

which is a contradiction. Hence the optimal solution must be unique when f is strictly concave.

47. Consider the following maximization program (P3):

$$\max_{\mathbf{x} \in \Re^n} f(\mathbf{x})$$

subject to

$$\forall i = 1, 2, \cdots, m, \quad g_i(\mathbf{x}) \leq 0.$$

Theorem 4 *(Kuhn-Tucker Theorem) Suppose that there exists some* $\hat{\mathbf{x}}$ *such that* $g_i(\hat{\mathbf{x}}) < 0$ *for all* $i = 1, 2, \cdots, m$. *(This is called the* **Slater Condition**.*)*[12] *Then if* $\mathbf{x}^*$ *is a solution to (P3), there must exist* m

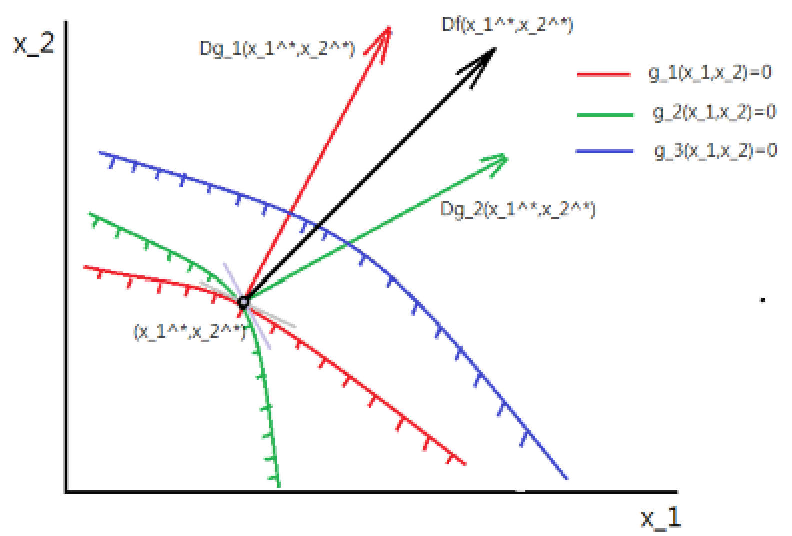
[12] Slater condition is a constraint qualification condition, and it applies mainly to the case where f is concave and g_i's are convex. There are several other constraint qualification conditions. For example, one of them requires that there exist $\hat{\mathbf{x}}$ such that, for all $i = 1, 2, \cdots, m$, $Dg_i(\mathbf{x}^*)^T\hat{\mathbf{x}} < 0$ whenever $g_i(\mathbf{x}^*) = 0$. Let us call this condition A.

non-negative constants (called the **Lagrange multipliers** *for the* m *constraints)* $\pi_1, \pi_2, \cdots, \pi_m$ *such that (i)*

$$\sum_{i=1}^{m} \pi_i Dg_i(\mathbf{x}^*) = Df(\mathbf{x}^*);$$

and (ii) (complementary slackness) for all $i = 1, 2, \cdots, m$, $\pi_i g_i(\mathbf{x}^*) = 0$. *Conversely, if* f *is concave and for all* $i = 1, 2, \cdots, m$, $g_i : \mathcal{R}^n \to \mathcal{R}$ *is convex, and if* $\mathbf{x}^*$ *satisfies the above (i) and (ii), then* $\mathbf{x}^*$ *is a solution to the above program (P3). If* f *is strictly concave, such* $\mathbf{x}^*$ *is unique.*

To see the role of the **Slater Condition** consider the following example with $m = n = 2$: $f(x_1, x_2) = x_1 + x_2$, $g_1(x_1, x_2) = x_1^2 + x_2^2 - 1$, $g_2(x_1, x_2) = x_1^2 + (x_2 - 2)^2 - 1$. There is only one feasible solution in this case, namely, $(x_1, x_2) = (0, 1)$, which is apparently the optimal solution. Thus $\mathbf{x}^* = (0, 1)$. However, there does not exist $\pi_1, \pi_2 \in \Re_+$ such that $Df(\mathbf{x}^*) = \pi_1 Dg_1(\mathbf{x}^*) + \pi_2 Dg_2(\mathbf{x}^*)$. This happens because we cannot find some $\hat{\mathbf{x}}$ such that $g_1(\hat{\mathbf{x}}) < 0$ and $g_2(\hat{\mathbf{x}}) < 0$!



To see why $Df(\mathbf{x}^*)$ must be contained in the polyhedral cone generated by $\{Dg_i(\mathbf{x}^*); i = 1, 2, \cdots, m\}$, note that in the opposite case we would be able to find some tiny vector $\mathbf{a} \equiv \mathbf{x} - \mathbf{x}^*$ such that $\mathbf{a}$ and each $Dg_i(\mathbf{x}^*)$ together create an obtuse angle but it and $Df(\mathbf{x}^*)$ together create an acute angle. This implies that we can find another feasible solution $\mathbf{x}$ close to $\mathbf{x}^*$ such that $f(\mathbf{x}) > f(\mathbf{x}^*)$, which is a contradiction. I shall demonstrate this fact in class using a graph.[13]

48. Let me prove the sufficiency of Kuhn-Tucker Theorem. Suppose that f and all $-g_i$ are concave. Suppose that there exist non-negative $\pi_1, \pi_2, \cdots, \pi_m$ such that at $\mathbf{x}^*$, (i)

$$\sum_{i=1}^{m} \pi_i Dg_i(\mathbf{x}^*) = Df(\mathbf{x}^*);$$

and (ii) (complementary slackness) for all $i = 1, 2, \cdots, m$, $\pi_i g_i(\mathbf{x}^*) = 0$. We must show that given any $\mathbf{x}$ such that $g_i(\mathbf{x}) \leq 0$ for all $i = 1, 2, \cdots, m$, we have

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq 0.$$

To this end, note that

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq Df(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*)$$

$$= \sum_{i=1}^{m} \pi_i Dg_i(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) \leq \sum_{i=1}^{m} \pi_i[g_i(\mathbf{x}) - g_i(\mathbf{x}^*)]$$

[13]The necessity that $Df(\mathbf{x}^*)$ must be contained in the polyhedral cone generated by $\{Dg_i(\mathbf{x}^*); i = 1, 2, \cdots, m\}$ is actually a consequence of the following Farkas Lamma: if $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m, \mathbf{b} \in \Re^n$ are such that $\mathbf{b} \cdot \mathbf{x} \leq 0$ whenever $\mathbf{a}_i \cdot \mathbf{x} \leq 0$ for all $i = 1, 2, \cdots, m$, where $\mathbf{x} \in \Re^n$, then it must be that $\mathbf{b} \in \mathcal{P} \equiv \{\sum_{i=1}^{m} t_i \mathbf{a}_i : \ t_i \in \Re_+, \forall i = 1, 2, \cdots, m\}$. We say that $\mathcal{P}$ is a polyhedral cone generated by $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m$. The idea is that $\mathcal{P}$ is non-empty, closed and convex, and if $\mathcal{P}$ does not contain $\mathbf{b}$, then there exists a hyperplane that separates strictly $\mathcal{P}$ and $\mathbf{b}$; see section 11 of the lecture note *Separating Hyperplane Theorem*. That is, there exists some $\mathbf{y} \in \Re^n$ and $\alpha \in \Re$ such that $\mathbf{p} \cdot \mathbf{y} < \alpha$ for all $\mathbf{p} \in \mathcal{P}$ and yet $\mathbf{b} \cdot \mathbf{y} > \alpha$. Note that if $\mathbf{p} \in \mathcal{P}$ is such that $\mathbf{p} \cdot \mathbf{y} > 0$, then $\alpha > 0$, but since $t\mathbf{p} \in \mathcal{P}$ for any $t \in \Re_+$, we must have $t\mathbf{p} \cdot \mathbf{y} > \alpha$ for sufficiently large t, a contradiction. Since $\mathbf{0} \in \mathcal{P}$, we actually have $\alpha > 0 = \mathbf{0} \cdot \mathbf{y}$. Thus we must have $\mathbf{p} \cdot \mathbf{y} \leq 0$ for all $\mathbf{p} \in \mathcal{P}$ and yet $\mathbf{b} \cdot \mathbf{y} > 0$. Now, if we let $\mathbf{b} = Df(\mathbf{x}^*)$ and $\mathbf{a}_i = Dg(\mathbf{x}^*)$ for all $i = 1, 2, \cdots, m$, and if we suppose that the necessity fails but condition A holds, then for $\epsilon, e > 0$ small enough, $g_i(\mathbf{x}^* + \epsilon(1-e)\mathbf{y} + \epsilon e\hat{\mathbf{x}}) \leq g_i(\mathbf{x}^*) \leq 0$ but $f(\mathbf{x}^* + \epsilon(1-e)\mathbf{y} + \epsilon e\hat{\mathbf{x}}) > f(\mathbf{x}^*)$, which is a contradiction.

$$= \sum_{i=1}^{m} \pi_i g_i(\mathbf{x}) \leq 0,$$

where the first inequality follows from the preceding footnote and the fact that f is concave, the first equality follows from the assumption that the Kuhn-Tucker condition holds at $\mathbf{x}^*$, the second inequality follows from the fact that $\pi_i \geq 0$ and g_i is convex for all $i = 1, 2, \cdots, m$, and the last equality follows from the assumption that the complementary slackness holds at $\mathbf{x}^*$, and the fact that $g_i(\mathbf{x}) \leq 0$ for all $i = 1, 2, \cdots, m$. Thus $\mathbf{x}^*$ is indeed an optimal solution to (P3). Finally, if both $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are optimal solutions to (P2), then $\mathbf{x}' \equiv \frac{1}{2}(\mathbf{x}^* + \mathbf{x}^{**})$ must be feasible also (i.e., by Jensen's inequality $g_i(\mathbf{x}') \leq 0$ for all i), and yet by Jensen's inequality again we have

$$f(\mathbf{x}') > \frac{1}{2}[f(\mathbf{x}^*) + f(\mathbf{x}^{**})] = f(\mathbf{x}^*),$$

which is a contradiction. Hence the optimal solution must be unique when f is strictly concave.

49. **Theorem 5** *Consider the following maximization problem:*

$$\max_{\mathbf{x} \in \Re^n} f(\mathbf{x})$$

subject to

$$g_i(\mathbf{x}) \leq 0, \ \forall i = 1, 2, \cdots, m;$$
$$h_j(\mathbf{x}) = 0, \ \forall j = 1, 2, \cdots, l.$$

This maximization problem involves m inequality constraints and l equality constraints. The associated Slater condition is as follows: there exists $\hat{\mathbf{x}} \in \Re^n$ such that

$$g_i(\hat{\mathbf{x}}) < 0, \ \forall i = 1, 2, \cdots, m;$$
$$h_j(\hat{\mathbf{x}}) = 0, \ \forall j = 1, 2, \cdots, l.$$

When the Slater condition holds, the following Kuhn-Tucker necessary conditions must hold: at an optimal solution $\mathbf{x}^*$, where $\{Dh_j(\mathbf{x}^*); j = 1, 2, \cdots, l\}$ consists of a set of linearly independent vectors, there must exist $(\mu_1, \mu_2, \cdots, \mu_m)^T \in \Re^m_+$ and $(\lambda_1, \lambda_2, \cdots, \lambda_l)^T \in \Re^l$ such that

$$(\textbf{Stationarity}) \ Df(\mathbf{x}^*) = \sum_{i=1}^{m} \mu_i Dg_i(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j Dh_j(\mathbf{x}^*);$$

$$h_j(\mathbf{x}^*) = 0,\ \forall j = 1, 2, \cdots, l;$$

$$\textit{(}\textbf{Complementary Slackness}\textit{)}\ \mu_i g_i(\mathbf{x}^*) = 0,\ \forall i = 1, 2, \cdots, m.$$

Conversely, when the Slater condition holds, f is concave, g_i's are convex, and h_j's are affine, if $\mathbf{x}^*$, $(\mu_1, \mu_2, \cdots, \mu_m)^T \in \Re^m_+$, and $(\lambda_1, \lambda_2, \cdots, \lambda_l)^T \in \Re^l$ satisfy the above Kuhn-Tucker necessary conditions, then $\mathbf{x}^*$ must be an optimal solution.[14]

50. The remaining sections give some exercises.

51. Consider the following maximization problem: for some constant $a \in \Re_+$,

$$\max_{x \in \Re} f(x; a) \equiv 2a - (x - a)^2$$

subject to

$$g(x) = (x - 3)^2 - 1 \leq 0.$$

Given a, let $x^*(a)$ be the unique optimal solution and $\pi(a)$ the Lagrange multiplier associated with the constraint $g \leq 0$.

(i) Show that in the above maximization problem the Slater condition is satisfied, and that f and g are respectively concave and convex.

(ii) Show that there exist $\underline{a}, \overline{a} \in \Re_+$ with $\underline{a} < \overline{a}$ such that $\pi(a) > 0$ if and only if either $a < \underline{a}$ or $a > \overline{a}$. Find $\underline{a}$ and $\overline{a}$.

(iii) Compute $x^*(\underline{a} - 1)$, $\pi(\underline{a} - 1)$, $Df(x^*(\underline{a} - 1))$ and $Dg(x^*(\underline{a} - 1))$.

[14] Visit for example https://en.wikipedia.org/wiki/Karush%E2%80%93Kuhn%E2%80%93Tucker_conditions to see a list of regularity conditions (or constraint qualification conditions) including the Slater condition. After Harold W. Kuhn and Albert W. Tucker published the Kuhn-Tucker conditions in 1951, scholars discovered that these necessary conditions had been stated by William Karush in his master's thesis in 1939. Hence nowadays people also refer to the Kuhn-Tucker conditions as Karush-Kuhn-Tucker conditions.

(iv) Compute $x^*(\overline{a}+1)$, and $\pi(\overline{a}+1)$, $Df(x^*(\overline{a}+1))$ and $Dg(\overline{a}+1))$.

52. Suppose that $f : \Re \to \Re$ is twice continuously differentiable and strictly concave. Let x^* attain the maximum value of f. Show that if $f'(x) > 0$ then $x^* > x$, and if $f'(x) < 0$ then $x^* < x$.

53. Consider the following maximization problem:

$$\max_{(x,y)\in\Re^2} x,$$

subject to

$$g(x,y) = (x+1)^2 + (y-2)^2 - 4 \leq 0;$$
$$h(x,y) = y - 3 = 0.$$

Let μ_g and μ_h denote the Lagrange multipliers associated with respectively $g \leq 0$ and $h = 0$. Let (x^*, y^*) denote *the* optimal solution.

(i) Are constraint qualification conditions stated in section 48 of Lecture 0 satisfied?

(ii) Find μ_g, μ_h, and (x^*, y^*).

54. Consider the following maximization problem:

$$\text{Problem (P):} \max_{T_1,T_2,q_1,q_2} \frac{1}{2}[T_1 - cq_1] + \frac{1}{2}[T_2 - cq_2]$$

subject to

$$\theta_1 V(q_1) - T_1 \geq 0, \tag{2}$$
$$\theta_2 V(q_2) - T_2 \geq 0, \tag{3}$$
$$\theta_1 V(q_1) - T_1 \geq \theta_1 V(q_2) - T_2, \tag{4}$$
$$\theta_2 V(q_2) - T_2 \geq \theta_2 V(q_1) - T_1. \tag{5}$$

The interpretation is as follows. A large bank is facing a small borrowing firm, which is equally likely to be of type θ_1 or type θ_2. By borrowing q dollars from the bank today (date 0), a type-θ_j borrowing firm will generate a cash flow $\theta_j V(q)$ tomorrow (date 1). The bank cannot tell the borrowing firm's type, and hence it offers a menu of choices to the borrowing firm, and asks the firm to pick one. The menu says that, for $j \in \{1, 2\}$, if the firm would borrow q_j dollars today, then it has to repay the bank T_j dollars tomorrow.

In designing the menu of bank loan contracts, the bank must make sure that the borrowing firm is willing to accept the deal, regardless of its type (so that (1) and (2) must hold); and the bank must also make sure that a type-θ_j borrowing firm would rather accept the deal (q_j, T_j) than accept (q_i, T_i), which is designed for the other type θ_i (and hence (3) and (4) must hold). (We are assuming zero interest rates, so that there is no discounting.)

Finally, note that the objective function in (P) says that the bank considers both types of the borrowing firm equally likely, and that for each dollar lent to the firm, the bank must incur a cost $c > 0$. The bank thus seeks to maximize its expected profit from lending to the firm.

Let us assume from now on that

$$\theta_1 = 3,\ \theta_2 = 4, V(q) = \ln(1+q),\ c = \frac{1}{4}.$$

Let $(q_2^{**}, q_1^{**}, T_2^{**}, T_1^{**})$ denote the optimal solution to the above maximization problem. Show that $q_2^{**} = 15$ and $q_1^{**} = 7$. What are the associated T_2^{**} and T_1^{**}? [15]

[15] To apply the Kuhn-Tucker Theorem, first define $h(x_j) \equiv e^{x_j} - 1$, and re-write the maximization problem as

$$\text{Problem (P): } \max_{T_1, T_2, x_1, x_2} f(T_1, T_2, x_1, x_2) \equiv T_1 - ch(x_1) + T_2 - ch(x_2)$$

subject to

$$g_1 \equiv T_1 - \theta_1 x_1 \le 0;$$

$$g_2 \equiv T_2 - \theta_2 x_2 \le 0;$$

$$g_3 \equiv T_1 - T_2 + \theta_1(x_2 - x_1) \le 0;$$

$$g_4 \equiv T_2 - T_1 + \theta_2(x_1 - x_2) \le 0.$$

Let μ_j be the associated Lagrange multiplier for constraint $g_j \le 0$.

Then, show that f is strictly concave and g_j's are convex, with

$$Df = \begin{bmatrix} 1 \\ 1 \\ -ch'(x_1) \\ -ch'(x_2) \end{bmatrix}, \; Dg_1 = \begin{bmatrix} 1 \\ 0 \\ -\theta_1 \\ 0 \end{bmatrix}, \; Dg_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -\theta_2 \end{bmatrix},$$

$$Dg_3 = \begin{bmatrix} 1 \\ -1 \\ -\theta_1 \\ \theta_1 \end{bmatrix}, \; Dg_4 = \begin{bmatrix} -1 \\ 1 \\ \theta_2 \\ -\theta_2 \end{bmatrix}.$$

It is useful to gain some insights before computing.

- Note that the θ_2 type can always pretend to be the θ_1 type and take the deal (T_1, x_1), which would allow the θ_2-buyer to obtain a payoff

$$\theta_2 x_1 - T_1 \geq \theta_1 x_1 - T_1 \geq 0,$$

  and the first inequality would be strict if $x_1 > 0$ (or equivalently $q_1 > 0$). Thus $x_1 > 0$ together with $g_4 \leq 0$ would imply $g_2 \leq 0$. That is, if we conjecture that $x_1 > 0$ then removing the second constraint would not alter the optimal solution to (P).

- Following the removal of $g_2 \leq 0$, we can further conjecture that the first constraint $g_1 \leq 0$ must be binding at an optimal solution, for otherwise we could raise T_1 and T_2 by the same tiny positive amount without violating g_1, g_3,and g_4, but this would increase f!

- The removal of $g_2 \leq 0$ and the conjecture that $g_1 = 0$ at optimum now allow us to further conjecture that g_4 must be binding at an optimal solution, for otherwise we could raise T_2 alone by a tiny positive amount without violating the other constaints, but this would increase f!

- Now, following $g_1 = 0 = g_4$ and following the removal of $g_2 \leq 0$, we can re-state $g_3 \leq 0$ as

$$(\theta_1 - \theta_2)(x_2 - x_1) \leq 0,$$

  but this last inequality would not be binding so long as $x_2 > x_1$.

Thus if we conjecture that $x_2 > x_1 > 0$ at optimum then we would also conjecture that

$$\mu_2 = \mu_3 = 0, \; T_1 = \theta_1 x_1, \; T_2 = T_1 + \theta_2(x_2 - x_1) = \theta_2 x_2 - (\theta_2 - \theta_1)x_1.$$

Now, by the fact that f is concave and g_1, g_2, g_3, g_4 are all convex, the sufficiency of Kuhn-Tucker Theorem applies, and hence we only need to find $\mu_1, \mu_4 \geq 0$ such that

$$Df = \mu_1 Dg_1 + \mu_4 Dg_4 \Rightarrow \begin{bmatrix} 1 \\ 1 \\ -ch'(x_1) \\ -ch'(x_2) \end{bmatrix} = \mu_1 \begin{bmatrix} 1 \\ 0 \\ -\theta_1 \\ 0 \end{bmatrix} + \mu_4 \begin{bmatrix} -1 \\ 1 \\ \theta_2 \\ -\theta_2 \end{bmatrix},$$

55. Consider the set $X = \{1, 2, 3\}$. We shall define different metrics on X. (Notice that X is not a linear space; a metric space need not be a linear space.)

Define $d_1(x, y) \equiv 1_{x \neq y}(x, y)$ and $d_2(x, y) \equiv |x - y|$, for all $x, y \in X$. For $j \in \{1, 2\}$, let β_j contain all open balls in the metric space (X, d_j). Let τ_j contain all (arbitrary) unions of elements of β_j. We say that τ_j is the *topology* generated by β_j. If τ_i and τ_j are two topologies defined on X, we say that τ_i is stronger than τ_j is $\tau_j \subset \tau_i$.

(i) Verify that d_1 and d_2 are both metrics on X.

(ii) Write down explicitly β_1 and β_2.

(iii) Between τ_1 and τ_2, which one is stronger?

Solution. Consider part (i). Note $1_{x \neq y}(x, y)$ equals either 0 or 1, and it equals 0 when and only when $x = y$. Moreover, $1_{x \neq y}(x, y) = 1_{x \neq y}(y, x)$ obviously. Finally, given $x, y, z \in X$, $1_{x \neq y}(x, y) + 1_{z \neq y}(y, z) \geq 1_{x \neq z}(x, z)$ obviously if $x = z$; and in case $x \neq z$, then $1_{x \neq y}(x, y) + 1_{z \neq y}(y, z) \geq 1_{x \neq z}(x, z)$ also holds because it is impossible that $x = y$ and $z = y$ at the same time. Thus d_1 is a metric on X.

Next, observe that $|x - y| \geq 0$ and equality holds when and only when $x = y$. Moreover, $|x - y| = |y - x|$ obviously. Finally, given $a, b \in \Re$, $|a| + |b| \geq |a + b|$ obviously, and hence given $x, y, z \in X$, if we define $a = x - y$ and $b = y - z$, then we have $|x - y| + |y - z| \geq |(x - y) + (y - z)| = |x - z|$. Thus d_2 is also a metric on X.

Consider part (ii). Let $B^j(x, e)$ denote the open ball centering on $x \in X$ with radius $e \in \Re_{++}$ under metric d_j. By definition, we have

$$B^1(x, e) = \begin{cases} \{x\}, & \text{if } e \leq 1; \\ \\ X, & \text{if } e > 1. \end{cases}$$

and if a solution to this system of equations exists and if the solution implies that $x_2 > x_1 > 0$ and $g_2, g_3 < 0$, then we are done. Moreover, because f is strictly concave in (x_1, x_2), the solution would be unique!

Since β_1 contains all open balls under d_1, we have

$$\beta_1 = \{\{1\}, \{2\}, \{3\}, X\}.$$

Observe also that if x, y, z are three distinct elements of X, then

$$B^2(x,e) = \begin{cases} \{x\}, & \text{if } \ e \leq \min(|x-y|, |y-z|); \\ \{x,y\} \text{ or } \ \{x,z\}, & \text{if } \ \min(|x-y|, |y-z|) < e \leq \max(|x-y|, |y-z|); \\ X, & \text{if } \ e > \max(|x-y|, |y-z|). \end{cases}$$

Since β_2 contains all open balls under d_2, we have

$$\beta_2 = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, X\}.$$

Consider part (iii). Since both β_1 and β_2 contain $\{1\}, \{2\}, \{3\}$, and since τ_j contains all (including empty) unions of elements of β_j, we have

$$\tau_1 = \tau_2 = 2^X;$$

that is, τ_1 and τ_2 are the same topology on X.

Remark. This exercise confirms the following fact:

Suppose that (X, d) is a metric space and X is a finite set. Let τ denote the collection of open sets. Then $\tau = 2^X$.[16]

56. Suppose X, Y are some nonempty sets and $f : X \to Y$. For all $A \subset X$ and $B \subset Y$, define $f(A) \equiv \{f(x) : x \in A\}$, $f^{-1}(B) \equiv \{x : f(x) \in B\}$.

    Show the following:

    (i) $f(\emptyset) = \emptyset = f^{-1}(\emptyset)$;
    (ii) $f(X) \subset Y$, $f^{-1}(Y) = X$;
    (iii) $A_1 \subset A_2 \subset X$, $B_1 \subset B_2 \subset Y \Rightarrow f(A_1) \subset f(A_2)$, $f^{-1}(B_1) \subset$

[16] For any $z \in X$, pick e_z such that $0 < e_z < d(x, z)$ for all $x \neq z$, $x \in X$. Then the open ball $B(z, e_z) = \{z\}$, proving that τ contains all singleton subsets of X. Recall that τ also contains all possible unions of open balls.

$f^{-1}(B_2)$;
(iv) $\cup_{i=1}^{\infty} A_i \subset X, \cup_{i=1}^{\infty} B_i \subset Y \Rightarrow f(\cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} f(A_i),\ f^{-1}(\cup_{i=1}^{\infty} B_i) = \cup_{i=1}^{\infty} f^{-1}(B_i)$;
(v) $\cap_{i=1}^{\infty} A_i \subset X, \cap_{i=1}^{\infty} B_i \subset Y \Rightarrow f(\cap_{i=1}^{\infty} A_i) \subset \cap_{i=1}^{\infty} f(A_i),\ f^{-1}(\cap_{i=1}^{\infty} B_i) = \cap_{i=1}^{\infty} f^{-1}(B_i)$;
(vi) $[f^{-1}(B)]^c = f^{-1}(B^c)$.
(vii) For $f : \mathcal{R} \to \mathcal{R}$ given by $f(x) = 3x^2+2$, show that $f^{-1}(f([0,1])) = [-1,1]$ and $f(f^{-1}([0,5])) = [2,5]$.

Solution. Parts (i) and (ii) are definitional. In part (i), since $\emptyset$ contains no $x \in X$, $f(\emptyset)$ contains no $f(x)$ in Y; and since $\emptyset$ contains no $f(x) \in Y$, $f^{-1}(\emptyset)$ contains no x in X. In part (ii), since X contains each of its elements x, and each $f(x) \in Y$, $f(X) \subset Y$; and since Y contains each $f(x)$ with $x \in X$, $f^{-1}(Y) = X$. It is easy to find an example where $f(X) \neq Y$; recall that this is true whenever f is not surjective.

Part (iii) is again definitional. Since A_2 contains every x contained by A_1, by definition $f(A_2)$ contains every $f(x)$ with $x \in A_1$; that is, $f(A_2)$ contains every element of $f(A_1)$. Thus $f(A_1) \subset f(A_2)$. Similarly, B_2 contains every $f(x)$ contained in B_1, and hence $f^{-1}(B_2)$ contains every element of $f^{-1}(B_1)$, so that $f^{-1}(B_1) \subset f^{-1}(B_2)$.

Consider part (iv). Note that $f(x) \in f(\cup_{i=1}^{\infty} A_i)$ if and only if $x \in \cup_{i=1}^{\infty} A_i$, which is true if and only if $x \in A_i$ for some i, which in turn is true if and only if $f(x) \in f(A_i)$ for some i, which in turn is true if and only if $f(x) \in \cup_{i=1}^{\infty} f(A_i)$. Thus we have $f(\cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} f(A_i)$.

Similarly, $x \in f^{-1}(\cap_{i=1}^{\infty} B_i)$ if and only if $f(x)$ is contained in each and every B_i, which is true if and only if x is contained in each and every $f^{-1}(B_i)$, or equivalently, if and only if $x \in \cap_{i=1}^{\infty} f^{-1}(B_i)$. Thus we have $f^{-1}(\cap_{i=1}^{\infty} B_i) = \cap_{i=1}^{\infty} f^{-1}(B_i)$.

Consider part (v). Note that $f(x) \in f(\cap_{i=1}^{\infty} A_i)$ if and only if x is contained in each and every A_i, which implies that $f(x)$ is contained in each and every $f(A_i)$, or equivalently, $f(x) \in \cap_{i=1}^{\infty} f(A_i)$. Hence we have $f(\cap_{i=1}^{\infty} A_i) \subset \cap_{i=1}^{\infty} f(A_i)$. It is easy to find an example where $f(\cap_{i=1}^{\infty} A_i) \neq \cap_{i=1}^{\infty} f(A_i)$. For example, let $X = Y = \Re$ and $f(x) = x^2$. Let $A_1 = (-1, 0)$ and $A_i = (0, 1)$ for all $i \geq 2$. Then we have

$f(\cap_{i=1}^{\infty} A_i) = f(\emptyset) = \emptyset$, but $f(A_j) = (0,1)$ for all j so that $\cap_{i=1}^{\infty} f(A_i) = (0,1)$ also.

Now, consider part (vi). Note that $x \in [f^{-1}(B)]^c$ if and only if $f^{-1}(B)$ does not contain x, which is true if and only if B does not contain $f(x)$, or equivalently, if and only if $f(x) \in B^c$, which in turn is true if and only if $x \in f^{-1}(B^c)$. Thus $[f^{-1}(B)]^c = f^{-1}(B^c)$.

Finally, consider part (vii). Note that $f([0,1]) = [2,5]$, and $f^{-1}([2,5]) = [-1,1]$ just as asserted. Note also that $f^{-1}([0,5]) = f^{-1}([2,5]) = [-1,1]$, and $f([-1,1]) = [2,5]$, just as asserted.

57. Consider the set $X = \{1,2\}$. Define $\tau_1 \equiv \{X, \emptyset\}$ and $\tau_2 \equiv 2^X$. Given j, we say that $f : (X, \tau_j) \to (\Re, \mathcal{B}(\Re))$ is a continuous function if $f^{-1}((a,b)) \in \tau_j$ whenever $a < b$, $a, b \in \Re$. We say that a sequence $\{x_n; n \in \mathbf{Z}_+\}$ in (X, τ_j) converges to $x \in X$ if for all $U \in \tau_j$ with $x \in U$, there exists some $N(U) \in \mathbf{Z}_+$ such that $x_m \in U$ whenever $m \geq N(U)$.

    (i) Find a function $f : (X, d_j) \to (\Re, \mathcal{B}(\Re))$ such that f is continuous if $j = 2$ but not if $j = 1$.

    (ii) Find a sequence $\{x_n; n \in \mathbf{Z}_+\}$ in X such that $\{x_n\}$ converges to some $x \in X$ if $j = 1$ but not if $j = 2$.

Appendix: Heine-Borel Theorem and Weierstrass Theorem

1. A collection of sets $\{F_j; j \in J\}$ is said to exhibit the *finite intersection property* if whenever we pick a finite number of elements from this collection, the intersection of those chosen elements is non-empty.

2. Suppose that (X, τ) is a topological space. Then X is compact if and only if every collection $\mathcal{F} = \{F_j; j \in J\}$ of closed sets satisfying the finite intersection property has a non-empty intersection itself.[17]

   Proof. Note that by DeMorgan's Law, $\{F_j^c; j \in J\}$ is an open covering for X if and only if $\mathcal{F} = \{F_j; j \in J\}$ is a collection of closed sets that has an empty intersection. If X is compact then $\{F_j^c; j \in J\}$ must have an open sub-covering for X, so that $\mathcal{F} = \{F_j; j \in J\}$ must have a sub-collection with an empty intersection. If instead $\mathcal{F}$ is a collection of closed sets that exhibits the *finite intersection property*, then $\{F_j^c; j \in J\}$ cannot be an open covering for X, so that $\mathcal{F}$ must have a non-empty intersection itself.

3. (**Weierstrass Theorem**.) Suppose that (X, τ) is a topological space. A function $f : X \to \Re$ is upper semi-continuous if and only if for all $r \in \Re$, the pre-image $f^{-1}((-\infty, r)) \in \tau$. If X is compact and f is upper semi-continuous then there exists $x^* \in X$ such that $f(x^*) \geq f(x)$ for all $x \in X$.

   Proof. The collection

   $$\mathcal{G} \equiv \{f^{-1}((-\infty, r)); r \in \Re\}$$

   is an open covering of X, which, by the fact that X is compact, must have a finite sub-covering. That is, for some $r_1 < r_2 < \cdots < r_n$,

   $$X \subset \bigcup_{j=1}^{n} f^{-1}((-\infty, r_j)) \Leftrightarrow f(x) < r_n, \ \forall x \in X.$$

[17] One can show that $A \subset X$ is compact if and only if every collection $\mathcal{F} = \{F_j; j \in J\}$ of closed sets of which the intersection of any finite number of elements is not disjoint from A also has an intersection that is not disjoint from A.

This implies that image set $f(X)$ is non-empty and bounded above in $\Re$. Thus $f(X)$ has a supremum, which we denote by f^*. Define for all positive integers n,

$$F_n = \{x \in X : \ f(x) \geq f^* - \frac{1}{n}\}.$$

By the upper semi-continuity of f, for each n, the complement of F_n is open and non-empty: if F_n is empty, then $f^* - \frac{1}{n}$ is a smaller upper bound than f^*, which is a contradiction. It follows that $\{F_n\}$ satisfies that the *finite intersection property*, so that by compactness of X, $\bigcap_n F_n \neq \emptyset$; that is, there exists $x^* \in X$ such that for all n,

$$f^* \geq f(x^*) \geq f^* - \frac{1}{n},$$

proving that $f(x^*) = f^*$. $\|$

4. (**Heine-Borel Theorem.**) A subset $A \subset \Re^n$ is compact if and only if A is bounded and closed.

   Proof. We shall prove the theorem in 5 steps.

   (**Step 1.**) If A is compact, then A is bounded.

   Let 0 be the origin of $\Re^n$. The collection $\{B_n\}$ defined by, for all positive integer n,

   $$B_n \equiv \{x \in \Re^n : d(x, 0) < n\},$$

   is an open covering for A, and hence must have a finite sub-covering, $\{B_{n(1)}, B_{n(2)}, \cdots, B_{n(J)}\}$, with $n(1) < n(2) < \cdots < n(J)$. This implies that $A \subset B_{n(J)}$, proving that A is bounded.

   (**Step 2.**) If A is compact, then A is closed.

   If $A = X$ then the assertion is obviously true. Thus let $x \in A^c$. For each $y \in A$, pick e_y such that $0 < e_y < \frac{1}{3}d(x, y)$, so that the collection

$\{B(y, e_y) : y \in A\}$ is an open covering for A, which has a finite sub-covering $\mathcal{G} \equiv \{B(y(j), e_{y(j)}); j = 1, 2, \cdots, m\}$. Note that $\mathcal{G}$ is disjoint from

$$\mathcal{H}_x \equiv \bigcap_{j=1}^{m} B(x, e_{y(j)}).$$

Note also that H_x is open,[18] and $A \subset \mathcal{G} \bigcap H_x = \emptyset$. Since we have arbitrarily picked $x \in A_c$, we have shown that A^c is actually the union of a collection of open sets H_x, and hence A^c is open. Thus A is closed.

(**Step 3.**) Suppose that (X, τ) is a compact topological space and $A \subset X$ is closed, then A is also compact.[19]

Let $\mathcal{G}$ be any open covering for A, so that $\mathcal{G} \bigcup A^c$ is an open covering for X, which has a finite sub-covering $\mathcal{G}_n \bigcup A^c$ for X, so that the finite collection $\mathcal{G}_n$ is a finite sub-covering (of $\mathcal{G}$) for A.

(**Step 4.**) The Cartesian product $R_0 \equiv \Pi_{j=1}^{n}[a_n, b_n]$, where $a_n < b_n$, is referred to as a closed rectangle in $\Re^n$, which is compact.

Suppose that $\mathcal{G}$ is an open covering for R_0, which has no finite sub-coverings, and we shall demonstrate a contradiction. By cutting each

[18]Recall that an open set is a union of open balls. We claim that a finite intersection of open sets, denoted by $\bigcap_{j=1}^{h} G_j$, is again an open set, where G_j's are open. Indeed, pick any $x \in \bigcap_{j=1}^{h} G_j$, then $x \in G_j$ for each $j = 1, 2, \cdots, h$, so that there exist h open balls $\{B_j; j = 1, 2, \cdots, h\}$ with $B_j = B(z_j, e_j) \subset G_j$ for each $j = 1, 2, \cdots, h$ such that $x \in \bigcap_{j=1}^{h} B_j$. Let e_x be such that $0 < e_x < \frac{1}{2}\min_{j=1}^{h}[e_j - d(x, z_j)]$, we conclude that $x \in B(x, e_x) \subset \bigcap_{j=1}^{h} B_j \subset \bigcap_{j=1}^{h} G_j$. Since we have picked $x \in \bigcap_{j=1}^{h} G_j$ arbitrarily, we have shown that $\bigcap_{j=1}^{h} G_j = \bigcup_{x \in \bigcap_{j=1}^{h} G_j} B(x, e_x)$; that is, $\bigcap_{j=1}^{h} G_j$ is a union of open balls. Hence a finite intersection of open sets is open.

[19]Suppose that (X, τ) is a topological space and $A \subset X$ is compact. Define the *relative topology* on A as

$$\tau_A \equiv \{A \bigcap G : G \in \tau\}.$$

Then (A, τ_A) is a compact topological space.

side of the rectangle R_0 in half, we can break R_0 into 2^n smaller rectangles, and one of them, say R_1, must take an infinite number of elements of $\mathcal{G}$ to cover. We can then switch attention to R_1, and cut each side of R_1 in half, and refer to as R_2 a smaller rectangle that takes an infinite number of elements of $\mathcal{G}$ to cover. Continue doing this, and we would obtain a sequence $\mathcal{R} \equiv \{R_m\}$ of smaller and smaller closed rectangles. It is easy to construct in $\mathcal{R}$ a Cauchy sequence $\{x_k\}$ with $x_k \in R_k$ for all k, which, by the fact that R_m is closed for all m, must have a limit contained in $\bigcap_m R_m$, and we denote the limit by x^*. Now, some element $G \in \mathcal{G}$ must contain x^*, and for some M sufficiently large, G contains all R_m with $m \geq M$ also; recall that the length of any side of R_m is no greater than $2^{-m} \max_n(b_n - a_n)$. This is a contradiction to the assertion that it takes an infinite number of elements of $\mathcal{G}$ to cover each R_m. We conclude that each open covering for R_0 must have a finite sub-covering for R_0, proving that R_0 is compact.

(**Step 5.**) If $A \subset \Re^n$ is closed and bounded, then A is compact.

Being bounded, A is contained in some closed rectangle R_0, which by Step 4 is compact, so that by Step 3, A, being closed, is compact also.

Tarski's Fixed-point Theorem

1. **Definition 1**: A lattice is a pair $(A, \leq)$, where A is some non-empty set and $\leq$ is a partial order (i.e. a reflexive, transitive, and anti-symmetric binary relation)[20] on A such that

$$a, b \in A \Rightarrow \sup(a, b) \in A, \ \inf(a, b) \in A.$$

The lattice $(A, \leq)$ is called *complete*, if for all $E \subset A$, $\sup E, \inf E \in A$. In this case, $\sup A$ and $\inf A$ are respectively the largest and the smallest element of A.[21] Given a lattice $(A, \leq)$, $(B, \leq)$ is a sublattice of A if $B \subset A$ and

$$a, b \in B \Rightarrow \sup_A(a, b), \ \inf_A(a, b) \in B.$$

Note that it can happen that $(B, \leq)$ is itself a lattice but not a sublattice of A. For example, let $A = \mathcal{R}^2$ and define $\leq$ on A as such that $(a, b), (x, y) \in \mathcal{R}^2$, $(a, b) \leq (x, y)$ if, as real numbers, $a \leq x$ and $b \leq y$. (There should be no ambiguity about the notation.) Let $B = \{(0,0)(0,1),(1,0),(2,2)\}$. Then $(B, \leq)$ is a lattice. Although

$$\sup_B((0,1),(1,0)) = (2,2) \in B,$$

note that

$$\sup_A((0,1),(1,0)) = (1,1) \in B^c,$$

and hence $(B, \leq)$ is not a sublattice of A. Given any partially ordered non-empty set $(A, \leq)$, we say $a, b \in A$ are comparable, if either $a \leq b$ or $b \leq a$. For any two comparable elements $a \leq b$ of A, define the *interval*

$$[a, b] = \{z \in A : a \leq z \leq b\}.$$

---

[20]The binaray relation $\leq$ is reflexive if for all $a \in A$, $a \leq a$; it is transitive if for all $a, b, c \in A$, $a \leq b$ and $b \leq c$ imply that $a \leq c$; and it is anti-symmetric if for all $a, b \in A$, $a \leq b$ and $b \leq a$ imply that $a = b$.

[21]Given a set A partially ordered by $\leq$, $a \in A$ is a maximal element of A if $a \leq b$ for some $b \in A$, then $a = b$. Minimal elements are defined analogously. An element a of A is a largest element if $a \geq b$ for all $b \in A$ (so that comparability is valid whenever a is involved). Smallest elements are defined analogously. Maximal and Minimal points are in general non-unique. A largest element is unique whenever it exists: Suppose a and b are both largest elements of A so that $a \leq b$ and $b \leq a$, but then they are equal by anti-symmetry. Largest elements need not exist though.

Example 1: For $a, b \in \mathcal{R}$, $a < b$, let $I_{[a,b]}$ be the set of increasing functions $f : [a, b] \to \mathcal{R}$. Then $(I_{[a,b]}, \leq^*)$ is a complete lattice where $f, g \in I_{[a,b]}$ are such that $f \leq^* g$ if and only if for all $x \in [a, b]$, $f(x) \leq g(x)$ (in the usual order on the real line). Let $C_{[a,b]}$ and $D_{[a,b]}$ be the sets of functions $f : [a, b] \to \mathcal{R}$ which are respectively continuous and differentiable on $[a, b]$. Then with the partial order $\leq^*$, $C_{[a,b]}$ is a lattice but not a complete lattice, whereas $D_{[a,b]}$ is not even a lattice.

Example 2: The real open interval $(0, 2)$ endowed with the usual order on the real line is a lattice but not a complete lattice.

Example 3: Suppose $(A, \leq)$ is a lattice. Is it true that for all finite non-empty $B \subset A$, $\sup B \in A$?

We claim that $s \equiv \sup(b_1, \sup(b_2, \cdots \sup(b_{m-1}, b_m)) \cdots)$, which exists in A because $(A, \leq)$ is a lattice, is exactly $\sup B$ in A, where $(b_1, b_2, \cdots, b_m)$ is any enumeration of B. To see this, note that s is clearly an upper bound of B in A, and that the infimum of any two upper bounds of B in A is an upper bound of B in A (prove it!). Thus the assertion can fail only if there exists another upper bound x of B in A with $s > x$. By definition, $x \geq b_{m-1}, b_m$ so that $x \geq \sup(b_{m-1}, b_m)$, which together with the fact that $x \geq b_{m-2}$ implies that $x \geq \sup(b_{m-2}, \sup(b_{m-1}, b_m))$. Repeating this argument, we have $x \geq s$, a contradiction to $s > x$. By the same token, $\inf B$ exists in A if $B \subset A$ is non-empty. The requirement that B is non-empty is needed because as in example 2, $\inf \emptyset = \sup(0, 2) = 2$ which is not contained in $(0, 2)$.

Example 4: If $(A, \leq)$ is a lattice and A is finite, is $(A, \leq)$ a complete lattice?

The answer is positive. To see this, let $B \subset A$ be non-empty. Then by example 3, $\inf B$, $\sup A$ exist in A. What if $B = \emptyset$? In this case, $\inf \emptyset = \sup A$ and $\sup \emptyset = \inf A$. Note that by assumption A itself is a non-empty subset of A, and hence the same conclusion follows from example 3.

2. If $(A, \leq)$ is a complete lattice, then $\sup \emptyset,\ \inf \emptyset \in A$. What is $\sup \emptyset$? At first, $x \in A$ is an upper bound of $\emptyset$ if for all $z \in \emptyset,\ z \leq x$. Thus every $x \in A$ is an upper bound of $\emptyset$, and therefore the least upper bound of $\emptyset$ is simply $\inf A$. By the same reasoning, $\inf \emptyset = \sup A$.

Lemma 1: $([a,b], \leq)$ is a lattice and is complete if $(A, \leq)$ is complete.

Proof For all $x, y \in [a,b]$, we have $x \leq b,\ y \leq b$, so that $\sup(x,y) \leq b$. Similarly, $a \leq \inf(x,y)$. It follows that $a \leq \inf(x,y) \leq x, y \leq \sup(x,y) \leq b$, so that $\inf(x,y) \in [a,b]$ and $\sup(x,y) \in [a,b]$; i.e. $([a,b], \leq)$ is a lattice.

Next, for any $B \subset [a,b]$,

$$x \in B \Rightarrow x \leq b,$$

so that b is an upper bound of B both in $[a,b]$ and in A. Note that $\forall x \in B,\ a \leq x \leq \sup_A B \leq b$, where the last inquality follows from the fact that b is an upper bound of B, and $\sup_A(\cdot)$ represents the supremum operator on the original system $(A, \leq)$, which exists because the latter is a complete lattice. Thus $\sup_A B$ is an upper bound of B in $[a,b]$. Let z be another upper bound of B in $[a,b]$. Then apparently $\inf(z, \sup_A B)$ is again an upper bound of B both in $[a,b]$ and in A, which implies that $\inf(z, \sup_A B) = \sup_A B$. We conclude that $\sup_A B = \sup_{[a,b]} B$. Similarly we have $\inf_A B = \inf_{[a,b]} B$. As B was chosen arbitrarily, $([a,b], \leq)$ is a complete lattice.

Definition 2: Let $B, C \subset A$. A function $f : B \to C$ is called "increasing" if[22]

$$x, y \in B,\ x \leq y \Rightarrow f(x) \leq f(y).$$

[22] Let $(A, \leq_a)$ and $(B, \leq_b)$ be two partially ordered sets. If $f : A \to B$ is such that

$$\forall x, y \in A,\ x \leq_a y \Rightarrow f(x) \leq_b f(y),$$

then f is said to be order-preserving. If f is one-to-one, order-preserving, and is such that

$$\forall x, y \in A,\ f(x) \leq_b f(y) \Rightarrow x \leq_a y,$$

Suppose that $B \cap C \neq \emptyset$, then $x \in B \cap C$ is a fixpoint of f if $f(x) = x$.

Theorem 1: (Tarski's fixpoint theorem) Let $(A, \leq)$ be a complete lattice and $f : A \to A$ an increasing function. If P is the set of fixpoints of f, then P is nonempty, and moreover, $(P, \leq)$ is a complete lattice. In particular, $\sup P, \inf P \in P$.

Proof The proof proceeds in 3 steps.
Step 1 Let u be the supremum of the set $E \equiv \{z \in A : f(z) \geq z\}$. We have

$$\forall z \in E,\ z \leq f(z) \leq f(u).$$

Thus $f(u)$ is an upper bound of E. By the definition of u, we have

$$u \leq f(u),$$

which by monotonicity of f implies that $f(u) \leq f(f(u))$, so that $f(u) \in E$, and therefore by the fact that u is the supremum of E,

$$f(u) \leq u.$$

Thus u is a fixpoint of f. If p is a fixpoint of f, then $p \in E$, and therefore $p \leq u$. Thus u is the supremum of all fixpoints of f also. This proves that P is nonempty, and $\sup P = u \in P$.
Step 2 Now we can mimic Step 1 and turn the spotlight to the infimum of P. We can conclude that $\inf P = \inf\{z \in A : f(z) \leq z\} \in P$.

Step 3 It remains to show that $(P, \leq)$ is a complete lattice. Let Y be any subset of P. Lemma 1 implies that $([\sup Y, \sup A], \leq)$ is a complete lattice. For any $x \in Y$, we have $x \leq \sup Y$, and therefore

$$x = f(x) \leq f(\sup Y),$$

implying that $\sup Y \leq f(\sup Y)$. If $\sup Y \leq z$, then $\sup Y \leq f(\sup Y) \leq f(z)$, and so by restricting the domain of f to $[\sup Y, \sup A]$, we obtain an increasing function f' on $[\sup Y, \sup A]$ to $[\sup Y, \sup A]$. Applying

then f is called an *isomorphism*.

the above step (ii) to the lattice $([\sup Y, \sup A], \leq)$ and function f', we conclude that $\inf P' \in P'$, where P' is the set of fixpoints of f'. Now $\inf P'$ is certainly a fixpoint of f, and it has to be the least fixpoint of f which is an upper bound of Y. In other words, $\inf P'$ is the infimum of Y in the system $(P, \leq)$. By the same reasoning there exists a supremum of Y in the system $(P, \leq)$. As Y was chosen arbitrarily, this shows that $(P, \leq)$ is a complete lattice.