Chapter 33

Mathematical Preliminaries

In this appendix, we provide essential definitions and key results which are used at various points in the book. We also provide a list of sources where more details and proofs may be looked up. The topics covered include: probability, linear algebra, linear programming, mathematical analysis, and computational complexity. We have only presented key definitions and results here and for rigorous technical details, the reader is urged to look up the references provided at the end.

33.1 Probability Theory

A probability model is a triple $(\Omega, \mathbb{F}, \mathbb{P})$ where

- Ω is a set of outcomes of an experiment and is called the sample space.
- $\mathbb{F} \subseteq 2^{\Omega}$ satisfies closure under complement and countable union of sets, and contains Ω; $\mathbb{F}$ is called a σ-algebra over Ω and the elements of $\mathbb{F}$ are called events.
- $\mathbb{P} : \mathbb{F} \longrightarrow [0, 1]$ is a probability mapping that satisfies
  (a) $\mathbb{P}(\emptyset) = 0 \leq \mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1, \ \forall A \in \mathbb{F}$
  (b) (Countable Additivity): Given a countably infinite number of disjoint subsets A_i $(i = 1, 2, \ldots)$ of Ω,

$$\mathbb{P}(A_1 \cup A_2 \cup \ldots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots$$

The above two properties are also called axioms of probability. A random variable X is a mapping $X : \Omega \longrightarrow \mathbb{R}$ such that the probability $\mathbb{P}\{X \leq x\}$ $\forall x \in \mathbb{R}$ is well defined and can be computed.

The cumulative distribution function (CDF) of a random variable X is a mapping $F_X : \mathbb{R} \longrightarrow [0, 1]$ defined by

$$F_X(x) = \mathbb{P}\{X \leq x\}$$

The CDF is monotone non-decreasing and right continuous and satisfies

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

A random variable where range is countable is called a discrete random variable while, a random variable whose CDF is continuous is called a continuous random variable. Clearly the range of a continuous random variable is uncountably infinite.

If X is a discrete random variable with range $\{1, 2, \dots\}$ we define its probability mass function as

$$\mathbb{P}\{X = i\} \text{ for } i = 1, 2, \dots$$

It can be shown easily that $\sum_i \mathbb{P}\{X = i\} = 1$. If X is a continuous random variable, we define its probability density function, if it exists, as

$$f_X(x) = \frac{dF_X(x)}{dx}; \ \forall x \in \text{range}(X)$$

It is a simple matter to show that

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Two events E and F are said to be mutually exclusive if $E \cap F = \emptyset$ which means

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$$

The events E and F are called independent if the probability of occurrence of E does not depend on the occurrence of F. It can be shown that independence of events E and F is equivalent to:

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$$

If $E_1, E_2, \dots, E_n$ are mutually disjoint events such that $E_1 \cup E_2 \cdots \cup E_n = \Omega$ and $\mathbb{P}(E_i) > 0 \ \forall i$, then for any event $F \in \mathbb{F}$, we can write

$$\mathbb{P}(F) = \sum_{i=1}^{n} \mathbb{P}(F|E_i)\mathbb{P}(E_i)$$

The Bayes Rule is an important result on conditional probabilities which states that for any two events $E, F \in \mathbb{F}$ such that $\mathbb{P}(E) > 0$ and $\mathbb{P}(F) > 0$,

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F)}$$

$\mathbb{P}(E)$ is called the prior; $\mathbb{P}(E|F)$ is called the posterior; $\mathbb{P}(F|E)\mathbb{P}(E)$ represents the support F provides to E.

An immediate extension of Bayes' rule is when the sets $E_1, E_2, \dots, E_n$ are mutually exclusive such that $\cup_{i=1}^{n} E_i = \Omega$ and $\mathbb{P}(E_i) > 0 \ \forall i$. In such a case, we have for $i = 1, \dots, n$,

$$\mathbb{P}(E_i|F) = \frac{\mathbb{P}(F|E_i)\mathbb{P}(E_i)}{\sum_{i=1}^{n} \mathbb{P}(F|E_i)\mathbb{P}(E_i)}$$

Suppose we have a random vector $X_1, \dots, X_n$. Define the random variable X as:

$$\mathbb{P}\{X = (s_1, \dots, s_n)\} = \mathbb{P}\{X_1 = s_1; \dots; X_n = s_n\}$$

This is called a joint probability distribution. If $X_1, \ldots, X_n$ are mutually independent, then $\forall(s_1, \ldots, s_n)$, we will have:

$$\mathbb{P}\{X = (s_1, \ldots, s_n)\} = \mathbb{P}\{X_1 = s_1\} \ldots \mathbb{P}\{X_n = s_n\}$$

The probabilities $\mathbb{P}\{X_i = s_i\}$ are called the marginal probabilities. Also, given that $X_1, \ldots, X_n$ are mutually independent random variables, we can define the following joint distributions with the probability mass functions:

- $\mathbb{P}\{X_1 = x_1\}\mathbb{P}\{X_2 = x_2\}$
- $\mathbb{P}\{X_1 = x_1\}\mathbb{P}\{X_2 = x_2\}\mathbb{P}\{X_3 = x_3\}$
- ...
- $\mathbb{P}\{X_1 = x_1\} \ldots \mathbb{P}\{X_n = x_n\}$

33.2 Linear Algebra

We present here a few key concepts in linear algebra. For the sake of brevity, we avoid defining a vector space here. The books [1, 2] must be consulted for more details.

Suppose $V = \{v_1, v_2, \ldots\}$ is a set of vectors and I is the index set $\{1, 2, \ldots\}$ for V. We say a vector x can be expressed as a linear combination of vectors in V if there are real numbers λ_i $(i \in I)$ such that not all λ_i are zero and

$$x = \sum_{i \in I} \lambda_i v_i$$

The set of all vectors that can be expressed as a linear combination of vectors in V is called the span of V and denoted by span(V).

Linear Independence and Linear Dependence

A finite set of vectors $V = \{v_1, v_2, \ldots, v_n\}$ is said to be linearly dependent if there exist $\lambda_i(i \in I)$, not all zero, such that

$$\sum_{i \in I} \lambda_i v_i = 0.$$

A finite set of vectors $V = \{v_1, v_2, \ldots, v_n\}$ is said to be linearly independent if they are not linearly dependent.

Example 33.1. The set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is linearly independent. The set $\{(5, 0, 0), (0, 1, 0), (0, 0, 10)\}$ is linearly independent. The set $\{(1, 0, 0),\ (0, 1, 0),\ (1, 1, 0)\}$ is linearly dependent. The set $\{(1, 0, 0),\ (0, 1, 0),\ (5, 6, 0)\}$ is also linearly dependent. □

Rank

The rank of a set of vectors V is the cardinality of a largest subset of linearly independent vectors in V.

Basis

Let V be a set of vectors and B a finite linearly independent subset of V. The set B is said to be a maximal linearly independent set if

$$B \cup \{x\} \text{ is linearly dependent } \forall x \in V \setminus B$$

A basis of V is a maximal linearly independent subset of V. It can be shown that every vector space $V \subseteq \mathbb{R}^n$ has a basis and if B is a basis of V, then $\text{span}(V) = \text{span}(B)$. Moreover if B and B' are two bases of V, then $|B| = |B'| = \text{rank}(V)$. The cardinality of the set B is called the dimension of V.

33.3 Linear Programming and Duality

A linear program (LP) consists of

- a set of variables $x_1, x_2, \ldots, x_n \in \mathbb{R}$,
- a linear objective function

$$\sum_{i=1}^{n} c_i x_i$$

  where $c_1, c_2, \ldots, c_n \in \mathbb{R}$ are known real numbers (called weights), and
- a set of linear constraints that weighted sums of variables must satisfy.

A linear program in canonical form is described by

$$\begin{aligned} \text{minimize} \quad & cx \\ \text{subject to} \quad & Ax \geq b \\ & x \geq 0 \end{aligned}$$

where

$$c = [c_1 \ldots c_n]; \;\; x = [x_1 \ldots x_n]^T; \;\; A = [a_{ij}]_{m \times n}; \;\; b = [b_1 \ldots b_m]^T.$$

A linear program in standard form is described by

$$\begin{aligned} \text{minimize} \quad & cx \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

The following is a typical maximization version of a linear program:

$$\begin{aligned} \text{maximize} \quad & cx \\ \text{subject to} \quad & Ax \leq b \\ & x \geq 0 \end{aligned}$$

A linear programming problem without an objective function is called a feasibility problem. A vector $x = (x_1 \ldots x_n)^T$ which satisfies the constraints is called a feasible solution. A feasible solution x that optimizes (minimizes or maximizes as the case may be) the objective function is called an optimal solution and is usually denoted by the vector $x^* = (x_1^*, \ldots, x_n^*)^T$.

The set of all feasible solutions of a linear program corresponds to a convex polyhedron in n-dimensional space. The constraints which are linear correspond to hyperplanes in the n-dimensional space.

As a consequence of the objective function being linear, any local optimum in the feasible space is also a global optimum. Furthermore, at least one optimal solution will exist at a vertex of the polyhedron.

The well known simplex algorithm for solving linear programs works as follows. The algorithm starts from a vertex and proceeds to neighboring vertices, each time improving the value of the objective function ("decreasing" in the case of minimization and "increasing" in the case of maximization) until an optimum is found. The worst case time complexity of the simplex algorithm is exponential in the number of variables and constraints.

Interior point methods solve linear programs by exploring the interior region of the polyhedron rather than vertices. Interior point methods with worst case polynomial time complexity have also been developed.

Duality in Linear Programs

Example 33.2. First we consider an example of an LP in canonical form:

$$\begin{aligned} \text{minimize } & 6x_1 + 8x_2 - 10x_3 \\ \text{subject to } & 3x_1 + x_2 - x_3 \geq 4 \\ & 5x_1 + 2x_2 - 7x_3 \geq 7 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

The dual of this LP is given by

$$\begin{aligned} \text{maximize } & 4w_1 + 7w_2 \\ \text{subject to } & 3w_1 + 5w_2 \leq 6 \\ & w_1 + 2w_2 \leq 8 \\ & -w_1 - 7w_2 \leq -10 \\ & w_1, w_2 \geq 0 \end{aligned}$$

We now generalize this example in the following discussion. □

Given

$$\begin{aligned} c &= [c_1 \ldots c_n]; \quad x = [x_1 \cdots x_n]^T \\ A &= [a_{ij}]_{m \times n}; \quad b = [b_1 \cdots b_m]^T \\ w &= [w_1 \cdots w_m] \end{aligned}$$

the primal LP in canonical form is:

$$\begin{aligned} \text{minimize} \;\; & cx \\ \text{subject to} \; & Ax \geq b \\ & x \geq 0. \end{aligned}$$

The dual of the above primal is given by

$$\begin{aligned} \text{maximize} \;\; & wb \\ \text{subject to} \; & wA \leq c \\ & w \geq 0. \end{aligned}$$

A primal LP in standard form is

$$\begin{aligned} \text{minimize} \;\; & cx \\ \text{subject to} \; & Ax = b \\ & x \geq 0. \end{aligned}$$

The dual of the above primal is:

$$\begin{aligned} \text{maximize} \;\; & wb \\ \text{subject to} \; & wA \leq c \\ w \;\; & \text{unrestricted} \end{aligned}$$

The above forms appear in the maxminimization and minmaximization problems in matrix games (Chapter 9). It is a simple matter to show that the dual of the dual of a (primal) problem is the original (primal) problem itself. We now state a few important results concerning duality, which are relevant to our requirements.

Weak Duality Theorem

If the primal is a maximization problem, then the value of any feasible primal solution is less than or equal to the value of any feasible dual solution. If the primal is a minimization problem, then the value of any feasible primal solution is greater than or equal to the value of any feasible dual solution.

Strong Duality Theorem

Given a primal and its dual, if one of them has an optimal solution then the other also has an optimal solution and the values of the optimal solutions are the same. Note that this is the key result which is used in proving the minimax theorem.

Fundamental Theorem of Duality

Given a primal and its dual, exactly one of the following statements is true.

(1) Both possess optimal solution (say x^* and w^*) with $cx^* = w^*b$.
(2) One problem has unbounded objective value in which case the other must be infeasible.
(3) Both problems are infeasible.

33.4 Mathematical Analysis

Metric Space

A metric space (V, d) consists of a set V and a mapping $d : V \times V \to \mathbb{R}$ such that $\forall x, y, z \in V$, the following holds.

(1) $d(x, y) \geq 0$
(2) $d(x, y) = 0$ iff $x = y$
(3) $d(x, y) = d(y, x)$
(4) $d(x, z) \leq d(x, y) + d(y, z)$

The mapping d is called a *metric* or *distance* function. It may be noted that the first condition above follows from the other three.

Open Ball

Given a metric space (V, d), an open ball of radius $r > 0$ and center $x \in V$, is the set $B(x, r) = \{y \in V : d(x, y) < r\}$.

Open Set

An open set X in a metric space (V, d) is a subset of V such that we can find, at each $x \in X$, an open ball that is contained in X.

Bounded Set

A subset X of a metric space (V, d) is said to be bounded if X is completely contained in some open ball, around 0, with a finite radius.

Closed Set

A subset X of a metric space V is said to be a closed set iff every convergent sequence in X converges to a point which lies in X. That is, for all sequences $\{x_k\}$ in X such that $x_k \to x$ for some $x \in V$, it will happen that $x \in X$. It may be noted that a set X is closed iff the complement set $X^c = V \setminus X$ is an open set.

Compact Set

Given a subset X of a metric space (V, d), X is said to be compact if every sequence of points in X has a convergent subsequence. A key result is that if the metric space V is $\mathbb{R}^n$ (under the Euclidean metric), then a subset X is compact iff it is closed and bounded.

Example 33.3 (Compact Sets). The closed interval $[0, 1]$ is compact. None of the sets $[0, \infty)$, $(0, 1)$, $(0, 1]$, $[0, 1)$, $(-\infty, \infty)$ is compact. Observe that the sets $[0, \infty)$ and $(-\infty, \infty)$ are closed but not bounded. Any finite subset of $\mathbb{R}$ is compact. □

A Useful Result

Let $X \subset \mathbb{R}^n$ and let $f : X \to \mathbb{R}^k$ be a continuous function. Then the image of a compact set under f is also compact.

Weierstrass Theorem

Let $X \subset \mathbb{R}^n$ and let $f : X \to \mathbb{R}$ be a continuous function. If X is compact, then f has and attains a maximum and a minimum in X.

Convexity

Convex Combination

Given $x_1, \ldots, x_m \in \mathbb{R}^n$, a point $y \in \mathbb{R}^n$ is called a convex combination of $x_1, \ldots, x_m$ if there exist numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

(1) $\lambda_i \geq 0, \quad i = 1, \ldots, m$
(2) $\sum_{i=1}^{m} \lambda_i = 1$
(3) $y = \sum_{i=1}^{m} \lambda_i x_i$

Convex Set

A set $X \subset \mathbb{R}^n$ is said to be convex if the convex combination of any two points in X is also in X. The above definition immediately implies that a finite set with two or more elements cannot be convex. Intuitively, the set X is convex if the straight line segment joining any two points in X is completely contained in X.

Example 33.4 (Convex Sets). A singleton set is always convex. The intervals $(0, 1)$, $(0, 1]$, $[0, 1)$, $[0, 1]$ are all convex. The set $X = \{x \in \mathbb{R}^2 : ||x|| < 1\}$ is convex. The set $X = \{x \in \mathbb{R}^2 : ||x|| = 1\}$ is not convex. □

Concave and Convex Functions

Let $X \subset \mathbb{R}^n$ be a convex set. A function $f : X \to \mathbb{R}$ is said to be concave iff $\forall x, y \in X$ and $\forall \lambda \in (0, 1)$,

$$f[\lambda x + (1 - \lambda)y] \geq \lambda f(x) + (1 - \lambda)f(y)$$

f is said to be convex iff $\forall x, y \in X$ and $\forall \lambda \in (0, 1)$

$$f[\lambda x + (1 - \lambda)y] \leq \lambda f(x) + (1 - \lambda)f(y)$$

An alternative definition for convex and concave functions is as follows. If $X \subset \mathbb{R}^n$ is a convex set and $f : X \to \mathbb{R}$ is a function, define

$$\text{sub } f = \{(x, y) : x \in X, y \in \mathbb{R}, f(x) \geq y\}$$
$$\text{epi } f = \{(x, y) : x \in X, y \in \mathbb{R}, f(x) \leq y\}$$

f is *concave* if sub f is *convex* ; f is *convex* if epi f is *convex*.

Example 33.5 (Convex and concave sets). The function $f_1(x) = x^3, x \in \mathbb{R}$ is neither convex nor concave. The function $f_2(x) = ax + b, x \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$ is both convex and concave. The function $f_3 : \mathbb{R}^+ \to \mathbb{R}$ defined by $f_3(x) = x^\alpha$ where $\mathbb{R}^+$ is the set of all positive real numbers is concave for $0 < \alpha < 1$ and convex for $\alpha > 1$. If $\alpha = 1$, then f_3 is both concave and convex. □

Some Results on Convexity

Suppose $X \subset \mathbb{R}^n$ is a convex set. Then:

(1) A function $f : X \to \mathbb{R}$ is concave iff the function $-f$ is convex.
(2) Let $f : X \to \mathbb{R}$ be concave or convex. If X is an open set, then f is continuous on X. If X is not an open set, then f is continuous on the interior of X.

Quasi-Concavity and Quasi-Convexity

Let $X \subset \mathbb{R}^n$ be a convex set and let $f : X \to \mathbb{R}$ be a function. The *upper contour set* of f at $a \in \mathbb{R}$ is defined as

$$U_f(a) = \{x \in X : f(x) \geq a\}$$

The *lower contour set* of f at $a \in \mathbb{R}$ is defined as

$$L_f(a) = \{x \in X : f(x) \leq a\}$$

A function $f : X \to \mathbb{R}$ is said to be *quasi-concave* if $U_f(a)$ is convex for all $a \in \mathbb{R}$ and is said to be *quasi-convex* if $L_f(a)$ is convex for all $a \in \mathbb{R}$.

Alternatively, $f : X \to \mathbb{R}$ is *quasi-concave* on X iff $\forall\, x, y \in X$ and $\forall\, \lambda \in (0, 1)$,

$$f[\lambda x + (1 - \lambda)y] \geq \min(f(x), f(y))$$

and *quasi-convex* on X iff $\forall x, y \in X$ and $\forall \lambda \in (0, 1)$,

$$f[\lambda x + (1 - \lambda)y] \leq \max(f(x), f(y))$$

It can be immediately noted that every convex (concave) function is quasi-convex (quasi-concave).

Example 33.6 (Quasi-concave and Quasi-convex sets). The function $f(x) = x^3$ on $\mathbb{R}$ is quasi-convex and also quasi-concave on $\mathbb{R}$. But it is neither convex nor concave on $\mathbb{R}$. Note that the upper contour set and also the lower contour set are both convex and hence the function is both quasi-convex and quasi-concave. Also, for every pair of points x_1 and x_2, the values of the function for points between x_1 and x_2 lie between $\min(f(x_1), f(x_2))$ and $\max(f(x_1), f(x_2))$ and therefore the function is both quasi-convex and quasi-concave. Any non-decreasing function $f : \mathbb{R} \to \mathbb{R}$ is quasi-convex and quasi-concave. But it need not be convex and need not be concave. □

33.5 Computational Complexity Classes

$\mathbb{P}, \mathbb{NP}$, *and* $\mathbb{NPC}$

The notion of $\mathbb{NP}$-completeness is studied in the framework of decision problems. Most problems we are typically interested in are optimization problems. In order to apply the theory of $\mathbb{NP}$-completeness, we have to recast optimization problems as decision problems.

Example 33.7. Consider the problem SPATH that finds, given an unweighted, undirected graph $G = (V, E)$ and two vertices $u, v \in V$, a shortest path between u and v. An instance of SPATH consists of a particular graph and two vertices of that graph. A given instance may have no solution, exactly one solution, or multiple solutions. A decision problem PATH related to optimization problem SPATH will be : given a graph $G = (V, E)$ and two vertices $u, v \in V$, and a non-negative integer k, does there exist a path between u and v of length at most k? The decision problem PATH is one way of transforming the original optimization problem into a decision problem. □

If an optimization problem is easy, then its related decision problem is easy as well. Similarly, if there is evidence that a decision problem is hard, then its related optimization problem is hard.

The classes $\mathbb{P}$ *and* $\mathbb{NP}$

The complexity of an algorithm is said to be polynomially bounded if its worst case complexity is bounded by a polynomial function of the input size. The common reference model used here is a deterministic Turing machine. $\mathbb{P}$ is the set of all problems which are solvable in polynomial time on a deterministic Turing machine.

$\mathbb{NP}$ represents the class of decision problems which can be solved in polynomial time by a non-deterministic Turing machine. A non-deterministic Turing machine makes the right guesses on every move and races towards the solution much faster than a deterministic Turing model. An equivalent definition of $\mathbb{NP}$ is : $\mathbb{NP}$ is the set of all decision problems whose solutions can be verified in polynomial time. More specifically, given a candidate solution to the problem (call it certificate), one can verify in polynomial time (on a deterministic Turing machine) whether the answer to the decision problem is YES or NO.

Clearly, $\mathbb{P} \subseteq \mathbb{NP}$. However it is unknown whether $\mathbb{NP} = \mathbb{P}$. This is currently the most celebrated open problem in computer science.

Reducibility of Problems

Suppose we have an algorithm for solving a problem Y. We are given a problem X and assume that there is a function T that takes an input x for X and produces $T(x)$ which is an input for Y, such that the correct answer for X on x is YES if and only if the correct answer for Y on $T(x)$ is YES. Then by using T and the algorithm for Y, we have an algorithm for X. If the function T can be computed in polynomially bounded time (on a deterministic Turing machine), we say X is polynomially reducible to Y and we write

$$X \leq_P Y$$

If $X \leq_P Y$, the implication is that Y is at least as hard to solve as X. That is, X is no harder to solve than Y. Clearly

$$X \leq_P Y \text{ and } Y \in \mathbb{P} \Longrightarrow X \in \mathbb{P}$$

$\mathbb{NP}$*-hard and* $\mathbb{NP}$*-complete Problems*

A decision problem Y is said to be $\mathbb{NP}$-hard if $X \leq_P Y, \quad \forall X \in \mathbb{NP}$. An $\mathbb{NP}$-hard problem Y is said to be $\mathbb{NP}$-complete if $Y \in \mathbb{NP}$. The set of all $\mathbb{NP}$-complete problems is denoted by $\mathbb{NPC}$.

Note. Informally, an $\mathbb{NP}$-hard problem is a problem that is at least as hard as any problem in $\mathbb{NP}$. If, in addition, the problem belongs to $\mathbb{NP}$, it would be called $\mathbb{NP}$-complete.

Note. In order to show that a decision problem Y is $\mathbb{NP}$-complete, it is enough we find a decision problem $X \in \mathbb{NPC}$ such that $X \leq_P Y$ and $Y \in \mathbb{NP}$.

Note. If it turns out that any single problem in $\mathbb{NPC}$ is in $\mathbb{P}$, then $\mathbb{NP} = \mathbb{P}$.

Note. An alternative way of characterizing $\mathbb{NPC}$ is that it is the set of all decision problems $Y \in \mathbb{NP}$ such that $X \leq_P Y$ where X is any $\mathbb{NP}$-complete problem.

A List of $\mathbb{NP}$*-complete Problems*

Here is a list of popular problems whose decision versions are $\mathbb{NP}$-complete.

(1) 3-SAT (Boolean satisfiability problem with three variables)
(2) Knapsack problem
(3) Traveling salesman problem
(4) Vertex cover problem
(5) Graph coloring problem
(6) Steiner tree problem
(7) Weighted set packing problem
(8) Weighted set covering problem

33.6 Summary and References

In this appendix, we have provided key definitions and results from probability theory, linear algebra, linear programming, mathematical analysis, and computational complexity. While these definitions and results serve as a ready reference, the reader is urged to look up numerous scholarly textbooks which are available in the area. We only mention a few sample texts here:

- Probability [3]
- Linear Algebra [1, 2]
- Linear Programming [4]
- Mathematical Analysis [5]
- Computational Complexity [6, 7]

The books by Vohra [8], Sundaram [9], and by Mas-Colell, Whinston, and Green [10] are excellent references as well for many of the mathematical preliminaries.

References

[1] Kenneth M. Hoffman and Ray Kunze. *Linear Algebra*. Prentice Hall, Second Edition, 1971.
[2] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Publishers, Fourth Edition, 2009.
[3] Sheldon M. Ross. *A First Course in Probability*. Pearson, Eighth Edition, 2010.
[4] Vasek Chvatal. *Linear Programming*. W.H. Freeman & Company, 1983.
[5] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill International Edition, Third Edition, 1976.
[6] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.
[7] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press and McGraw-Hill, Third Edition, 2009.
[8] Rakesh Vohra. *Advanced Mathematical Economics*. Cambridge University Press, 2009.
[9] Rangarajan K. Sundaram. *A First Course in Optimization Theory*. Cambridge University Press, 1996.
[10] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.