Introduction to Set Theory

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 400

Subsection 1

The Axiom of Choice and its Equivalents

Well-Ordering a Set

- We investigate which sets can be well-ordered.
- Cantor thought it obvious that every set can be well-ordered.
- A fairly intuitive "proof" of this "fact":
  In order to well-order a set A, it suffices to construct a one-to-one mapping of some ordinal λ onto A. We proceed by transfinite recursion. Let a be any set not in A. Define

$$f(0) = \begin{cases} \text{some element of } A, & \text{if } A \neq \emptyset \\ a, & \text{otherwise} \end{cases}$$

$$f(1) = \begin{cases} \text{some element of } A - \{f(0)\}, & \text{if } A - \{f(0)\} \neq \emptyset \\ a, & \text{otherwise} \end{cases}$$

Generally,

$$f(\alpha) = \begin{cases} \text{some element of } A - \text{ran}(f \restriction \alpha), & \text{if } A - \text{ran}(f \restriction \alpha) \neq \emptyset \\ a, & \text{otherwise} \end{cases}$$

Intuitively, f lists the elements of A, one by one, as long as they are available. When A is exhausted, f has value a.

Finishing the "Proof"

- First notice that A does get exhausted at some stage $\lambda < h(A)$, the Hartogs number of A.

  Indeed, for $\alpha < \beta$, if $f(\beta) \neq a$, then $f(\beta) \in A - \text{ran}(f \restriction \beta)$, $f(\alpha) \in \text{ran}(f \restriction \beta)$, and thus $f(\alpha) \neq f(\beta)$. If $f(\alpha) \neq a$ were to hold for all $\alpha < h(A)$, f would be a one-to-one mapping of $h(A)$ into A, contradicting the definition of $h(A)$ as the least ordinal which cannot be mapped into A by a one-to-one function.
- Let λ be the least $\alpha < h(A)$, such that $f(\alpha) = a$. The previous considerations show that $f \restriction \lambda$ is one-to-one. The "proof" is complete if we show that $\text{ran}(f \restriction \lambda) = A$.

  Clearly $\text{ran}(f \restriction \lambda) \subseteq A$: if $\text{ran}(f \restriction \lambda) \subset A$, $A - \text{ran}(f \restriction \lambda) \neq \emptyset$ and $f(\lambda) \neq a$, contradicting our definition of λ.

Issues with the "Proof"

- If one tries to justify this transfinite recursion by the Recursion Theorem, one discovers a need for a function **G** such that f can be defined by $f(\alpha) = \mathbf{G}(f \restriction \alpha)$. Such a function **G** should satisfy:
  - $\mathbf{G}(f \restriction \alpha) \in A - \text{ran}(f \restriction \alpha)$, if $A - \text{ran}(f \restriction \alpha) \neq \emptyset$,
  - $\mathbf{G}(f \restriction \alpha) = a$, otherwise.

  If A were well-orderable, some such **G** could easily be defined:

$$\mathbf{G}(x) = \begin{cases} \text{the } \prec\text{-least element of } A - \text{ran}x, \\ \qquad\qquad \text{if } x \text{ is a function and } A - \text{ran}x \neq \emptyset \\ a, \text{ otherwise} \end{cases},$$

  where $\prec$ is some well-ordering of A.
  In the absence of well-orderings on A, no property which could be used to define such a function **G** is obvious.

Choice Functions and Zermelo's Theorem

- Let S be a system of sets. A function g defined on S is called a **choice function** for S if $g(X) \in X$, for all nonempty $X \in S$.
- If we now assume that there is a choice function g for $\mathcal{P}(A)$, we are able to fill the gap in the previous proof by defining

$$\mathbf{G}(x) = \begin{cases} g(A - \text{ran}x), & \text{if } x \text{ is a function and } A - \text{ran}x \neq \emptyset \\ a, & \text{otherwise} \end{cases}$$

- We proved the difficult half of Zermelo's theorem:

Theorem

A set A can be well-ordered if and only if the set $\mathcal{P}(A)$ of all subsets of A has a choice function.

$\Rightarrow$: If $\prec$ well-orders A, we define a choice function g for $\mathcal{P}(A)$:

$$g(x) = \begin{cases} \text{the least element of } x \text{ in } \prec, & \text{if } x \neq \emptyset \\ \emptyset, & \text{if } x = \emptyset \end{cases}$$

Finite Systems of Sets have Choice Functions

- The problem of well-ordering the set A is now reduced to an equivalent question, that of finding a choice function for $\mathcal{P}(A)$.

Theorem

Every finite system of sets has a choice function.

- Proceed by induction. Let us assume that every system with n elements has a choice function. Let $|S| = n + 1$. Fix $X \in S$. The set $S - \{X\}$ has n elements, and, consequently, a choice function g_X.
  - If $X = \emptyset$, $g = g_X \cup \{(X, \emptyset)\}$ is a choice function for S.
  - If $X \neq \emptyset$, $g^x = g_X \cup \{(X, x)\}$ is choice function for S (for any $x \in X$).
- This proof cannot be generalized to show that every countable system of sets has a choice function.
- It is easy to find a choice function for $\mathcal{P}(\mathbb{N})$ or $\mathcal{P}(\mathbb{Q})$, but no such function for $\mathcal{P}(\mathbb{R})$ is evident.

The Axiom of Choice

- Choice functions for infinite systems of sets of real numbers have been used by analysts at least since the end of the nineteenth century.

The Axiom of Choice (Zermelo 1904)

There exists a choice function for every system of sets.

- In 1963, Paul Cohen showed that the Axiom of Choice cannot be proved from the axioms of Zermelo-Fraenkel set theory.
- The Axiom of Choice asserts that certain sets, the choice functions, exist without describing those sets as collections of objects having a particular property.
- Because of this, and because of some of its counterintuitive consequences, some mathematicians raised objections to its use.
- We look at equivalent formulations and some consequences.
- To keep track of the use of the Axiom of Choice, we denote the theorems whose proofs depend on it by an asterisk.

Equivalent Formulations

Theorem

The following statements are equivalent:

(a) (The Axiom of Choice) There exists a choice function for every system of sets.

(b) Every partition has a set of representatives.

(c) If $\langle X_i : i \in I \rangle$ is an indexed system of nonempty sets, then there is a function f such that $f(i) \in X_i$, for all $i \in I$.

- Recall that a **partition** of a set A is a system of mutually disjoint nonempty sets whose union equals A. We call $X \subseteq A$ a **set of representatives** for a partition S of A if, for every $C \in S$, $X \cap C$ has a unique element.
- The statement (c) can be equivalently formulated as:

(d) If $X_i \neq \emptyset$, for all $i \in I$, then $\prod_{i \in I} X_i \neq \emptyset$.

Proof of the Theorem

- (a) implies (b): Let f be a choice function for the partition S. Then $X = \text{ran} f$ is a set of representatives for S: Notice that for any $C \in S$, $f(C) \in X \cap C$, but $f(D) \notin X \cap C$ for $D \neq C$ (because $f(D) \in D$ and $D \cap C = \emptyset$). So $X \cap C = \{f(C)\}$, for any $C \in S$.
- (b) implies (c): Let $C_i = \{i\} \times X_i$. Since $i \neq i'$ implies $C_i \cap C_{i'} = \emptyset$, $S = \{C_i : i \in I\}$ is a partition. If f is a set of representatives for S, f is a set of ordered pairs, and for each $i \in I$, there is a unique x such that $(i, x) \in f \cap C_i$. But this means that f is a function on I and $f(i) \in X_i$, for all $i \in I$.
- (c) implies (a): Let S be a system of sets. Set $I = S - \{\emptyset\}$, $X_C = C$, for all $C \in I$. Then $\{X_C : C \in I\}$ is an indexed system of nonempty sets.
  - If $\emptyset \notin S$, and $f \in \prod_{C \in I} X_C$, f is a choice function for S.
  - If $\emptyset \in S$, then $f \cup \{(\emptyset, \emptyset)\}$ is a choice function for S.

Two Consequences of the Axiom of Choice

- Other equivalents to the Axiom of Choice are the Well-Ordering Theorem and Zorn's Lemma, which we prove later.
- First, some consequences of the Axiom of Choice:

Theorem*

Every infinite set has a countable subset.

- Let A be an infinite set. A can be well-ordered, i.e., arranged in a transfinite one-to-one sequence $\langle a_\alpha : \alpha < \Omega \rangle$, whose length Ω is an infinite ordinal. The range $C = \{a_\alpha : \alpha < \omega\}$ of the initial segment $\langle a_\alpha : \alpha < \omega \rangle$ of this sequence is a countable subset of A.

Theorem*

For every infinite set S there exists a unique aleph $\aleph_\alpha$ such that $|S| = \aleph_\alpha$.

- As S can be well-ordered, it is equipotent to some infinite ordinal. Hence, also to a unique initial ordinal number ω_α.

Rigorous Justification of Cardinal Numbers

- In set theory with the Axiom of Choice, we can define, for any set X, its cardinal number $|X|$ as the initial ordinal equipotent to X.
- Sets X and Y are equipotent if and only if $|X|$ is the same ordinal as $|Y|$ (i.e., $|X| = |Y|$).
- Also, the ordering $<$ of cardinal numbers by size agrees with the ordering of ordinals by $\in$: $|X| < |Y|$ if and only if $|X| \in |Y|$.
- These considerations rigorously justify: There are sets called **cardinals** with the property that, for every set X, there is a unique cardinal $|X|$, and sets X and Y are equipotent if and only if $|X|$ is equal to $|Y|$.

Theorem*

For any sets A and B either $|A| \leq |B|$ or $|B| \leq |A|$.

- $\in$ is a linear ordering (a well-ordering) on any set of ordinal numbers.

Countable Collections of Countable Sets

Theorem*

The union of a countable collection of countable sets is countable.

- Let S be a countable set whose every element is countable, and let $A = \bigcup S$. We show that A is countable. As S is countable, there is a one-to-one sequence $\langle A_n : n \in \mathbb{N} \rangle$, such that $S = \{A_n : n \in \mathbb{N}\}$. For each $n \in \mathbb{N}$, the set A_n is countable. Thus, there exists a sequence whose range is A_n. By the Axiom of Choice, we can choose one such sequence for each n: For each n, let S_n be the set of all sequences whose range is A_n. Let F be a choice function on $\{S_n : n \in \mathbb{N}\}$, and let $s_n = F(S_n)$ for each n.
  Having chosen one $s_n = \langle a_n(k) : k \in \mathbb{N} \rangle$ for each n, we obtain a mapping f of $\mathbb{N} \times \mathbb{N}$ onto A by letting $f(n, k) = a_n(k)$. Since $\mathbb{N} \times \mathbb{N}$ is countable and A is its image under f, A is also countable.

$2^{\aleph_0}$ and ω_1

Corollary*

The set of all real numbers is not the union of countably many countable sets.

- The set $\mathbb{R}$ is uncountable.

Corollary*

The ordinal ω_1 is not the supremum of a countable set of countable ordinals.

- If $\{\alpha_n : n \in \mathbb{N}\}$ is a set of countable ordinals, then its supremum $\alpha = \sup\{\alpha_n : n \in \mathbb{N}\} = \bigcup_{n\in\mathbb{N}} \alpha_n$ is a countable set, $\alpha < \omega_1$.

Theorem*

$2^{\aleph_0} \geq \aleph_1$

- This follows from the theorem and the fact that $2^{\aleph_0} > \aleph_0$.
- As a result, the Continuum Hypothesis can be reformulated as the conjecture that $2^{\aleph_0} = \aleph_1$, the least uncountable cardinal number.

Cardinality of Image and Union

Theorem*

If f is a function and A is a set, then $|f[A]| \leq |A|$.

- For each $b \in f[A]$, let $X_b = f^{-1}(\{b\})$. Note that $X_b \neq \emptyset$ and $X_{b_1} \cap X_{b_2} = \emptyset$ if $b_1 \neq b_2$. Take $g \in \prod_{b \in f[A]} X_b$. Then $g : f[A] \to A$ and $b_1 \neq b_2$ implies $g(b_1) \in X_{b_1}$, $g(b_2) \in X_{b_2}$. So $g(b_1) \neq g(b_2)$, i.e., g is one-to-one mapping of $f[A]$ into A, and, thus, $|f[A]| \leq |A|$.

Theorem*

If $|S| \leq \aleph_\alpha$, and, for all $A \in S$, $|A| \leq \aleph_\alpha$, then $|\bigcup S| \leq \aleph_\alpha$.

- We assume that $S \neq \emptyset$ and all $A \in S$ are nonempty. Write $S = \{A_\nu : \nu < \aleph_\alpha\}$. For each $\nu < \aleph_\alpha$, choose a transfinite sequence $a_\nu = \langle a_\nu(\kappa) : \kappa < \aleph_\alpha \rangle$, such that $A_\nu = \{a_\nu(\kappa) : \kappa < \aleph_\alpha\}$. We define a mapping f on $\aleph_\alpha \times \aleph_\alpha$ onto $\bigcup S$ by $f(\nu, \kappa) = a_\nu(\kappa)$. By the preceding theorem, $|\bigcup S| = |\aleph_\alpha \times \aleph_\alpha| = \aleph_\alpha$.

The Well-Ordering Principle and Zorn's Lemma

Theorem

The following statements are equivalent:

(a) (The Axiom of Choice) There exists a choice function for every system of sets.

(b) (The Well-Ordering Principle) Every set can be well-ordered.

(c) (Zorn's Lemma) If every chain in a partially ordered set has an upper bound, then the partially ordered set has a maximal element.

- Recall that a **chain** is a linearly ordered subset of an ordered set.
- (a) equivalent to (b): Follows immediately from a preceding theorem.
- (a) implies (c): Let $(A, \preccurlyeq)$ be a partially ordered set in which every chain has an upper bound. Our strategy is to search for a maximal element of $(A, \preccurlyeq)$ by constructing a $\preccurlyeq$-increasing transfinite sequence of elements of A. We fix some $b \in A$ and a choice function g for $\mathcal{P}(A)$, and define $\langle a_\alpha : \alpha < h(A) \rangle$ by transfinite recursion.

Proof of (a) implies (c)

- (a) implies (c) (Cont'd): Given $\langle a_\xi : \xi < \alpha \rangle$, we consider two cases:
  - If $b \neq a_\xi$ for all $\xi < \alpha$ and $A_\alpha = \{a \in A : a_\xi \prec a \text{ holds for all } \xi < \alpha\}$, we let $a_\alpha = g(A_\alpha)$;
  - Otherwise we let $a_\alpha = b$.

  The definition is justified by recursion. We note that $a_\alpha = b$ for some $\alpha < h(A)$: Otherwise, $\langle a_\xi : \xi < h(A) \rangle$ would be a one-to-one mapping of $h(A)$ into A. Let λ be the least α for which $a_\alpha = b$. Then the set $C = \{a_\xi : \xi < \lambda\}$ is a chain in $(A, \preccurlyeq)$ and so it has an upper bound $c \in A$. If $c \prec a$ for some $a \in A$, we have $a \in A_\lambda \neq \emptyset$ and $a_\lambda = g(A_\lambda) \neq b$, a contradiction. So c is a maximal element of A. (It is easy to see that, in fact, $\lambda = \beta + 1$ and $c = a_\beta$.)

Proof of (c) implies (a)

- (c) implies (a): It suffices to show that every system of nonempty sets S has some choice function. Let F be the system of all functions f for which $\mathrm{dom} f \subseteq S$ and $f(X) \in X$ holds for any $X \in \mathrm{dom} f$.
  - The set F is ordered by inclusion $\subseteq$.
  - If F_0 is a linearly ordered subset of $(F, \subseteq)$ (i.e., either $f \subseteq g$ or $g \subseteq f$ holds for any $f, g \in F_0$), $f_0 = \bigcup F_0$ is a function.
  - We can check that $f_0 \in F$ and f_0 is an upper bound on F_0 in $(F, \subseteq)$.

  The assumptions of Zorn's Lemma being satisfied, we conclude that $(F, \subseteq)$ has a maximal element $\overline{f}$.

  Claim: $\mathrm{dom}\overline{f} = S$.

  If not, select some $X \in S - \mathrm{dom}\overline{f}$ and $x \in X$. But, then, the function $\overline{\overline{f}} = \overline{f} \cup \{(X, x)\} \in F$. Moreover, $\overline{\overline{f}} \supset \overline{f}$, contradicting the maximality of $\overline{f}$.

Subsection 2

The Use of the Axiom of Choice in Mathematics

Example I: Closure Points

- A sequence of real numbers $\langle x_n : n \in \mathbb{N} \rangle$ **converges** to $a \in \mathbb{R}$ if, for every real number $\varepsilon > 0$, there exists $n_\varepsilon \in \mathbb{N}$, such that $|x_n - a| < \varepsilon$ holds for all natural numbers $n \geq n_\varepsilon$ (note, $|x|$ denotes the absolute value of x; not the cardinality of x).
- Let A be a set of real numbers. Closure points of A can be defined in either (or both) of the following ways:
  - (a) $a \in \mathbb{R}$ is a **closure point** of A if and only if there exists a sequence $\langle x_n : n \in \mathbb{N} \rangle$ with values in A, which converges to a.
  - (b) $a \in \mathbb{R}$ is a **closure point** of A if and only if, for every positive real number ε, there exists $x \in A$, such that $|x - a| < \varepsilon$.
- It is then necessary to prove that (a) and (b) are equivalent:
  - (a) implies (b): Given $\varepsilon > 0$, there is $n_\varepsilon \in \mathbb{N}$, such that $|x_n - a| < \varepsilon$ for all $n \geq n_\epsilon$. In particular, $|x_{n_\varepsilon} - a| < \varepsilon$ and $x_{n_\varepsilon} \in A$.
  - (b) implies (a): Let $X_n = \{x \in A : |x - a| < \frac{1}{n}\}$. By (b), $X_n \neq \emptyset$, for all $n \in \mathbb{N}$. Let $\langle x_n : n \in \mathbb{N} \rangle$ be a sequence such that $x_n \in X_n$, for all $n \in \mathbb{N}$. Then each $x_n \in A$ and $\langle x_n : n \in \mathbb{N} \rangle$ converges to a.

Equivalence Requires the Axiom of Choice

- What reasons do we have to assume that the sequence $\langle x_n : n \in \mathbb{N} \rangle$ exists?
- Notice that we do not give any property $\mathbf{P}(x, y)$, such that $\mathbf{P}(n, y)$ holds if and only if $y = x_n$, for all $n \in \mathbb{N}$.
- Such a property can be exhibited in special cases, e.g., if A is open.
- However, it has been shown that the equivalence of (a) and (b) for all $A \subseteq \mathbb{R}$ cannot be proved from the axioms of Zermelo-Fraenkel set theory alone.
- Of course, if we do assume the Axiom of Choice, the fact that $X \neq \emptyset$, for all $n \in \mathbb{N}$, immediately implies that $\prod_{n \in \mathbb{N}} X_n \neq \emptyset$.

Example II: Continuity of a Function

- Continuity of a real-valued function of a real variable is defined in one of the following ways:
  (a) $f : \mathbb{R} \to \mathbb{R}$ is **continuous** at $a \in \mathbb{R}$ if and only if, for every $\varepsilon > 0$, there is $\delta > 0$, such that $|f(x) - f(a)| < \varepsilon$, for all x such that $|x - a| < \delta$.
  (b) $f : \mathbb{R} \to \mathbb{R}$ is continuous at $a \in \mathbb{R}$ if and only if for every sequence $\langle x_n : n \in \mathbb{N} \rangle$ converging to a, $\langle f(x_n) : n \in \mathbb{N} \rangle$ converges to $f(a)$.
- (a) implies (b): If $\langle x_n : n \in \mathbb{N} \rangle$ converges to a and if $\varepsilon > 0$ is given, then, first, we find $\delta > 0$ as in (a), and, because $\langle x_n : n \in \mathbb{N} \rangle$ converges, there exists n_δ, such that $|x_n - a| < \varepsilon$ whenever $n \geq n_\delta$. Clearly, $|f(x_n) - f(a)| < \varepsilon$, for all such n.
- If we assume the Axiom of Choice, then (b) also implies (a), and, hence, (a) and (b) are two equivalent definitions of continuity.

Example II: Continuity of a Function (Cont'd)

- Recall the two definitions:
  (a) $f : \mathbb{R} \to \mathbb{R}$ is **continuous** at $a \in \mathbb{R}$ if and only if, for every $\varepsilon > 0$, there is $\delta > 0$, such that $|f(x) - f(a)| < \varepsilon$, for all x such that $|x - a| < \delta$.
  (b) $f : \mathbb{R} \to \mathbb{R}$ is continuous at $a \in \mathbb{R}$ if and only if for every sequence $\langle x_n : n \in \mathbb{N} \rangle$ converging to a, $\langle f(x_n) : n \in \mathbb{N} \rangle$ converges to $f(a)$.
- Suppose that (a) fails. Then, there exists $\varepsilon > 0$, such that, for each $\delta > 0$, there exists an x such that $|x - a| < \delta$ but $|f(x) - f(a)| \geq \varepsilon$. In particular, for each $k = 1, 2, 3, \ldots$, we can choose some x_k, such that $|x_k - a| < \frac{1}{k}$ and $|f(x_k) - f(a)| \geq \varepsilon$. The sequence $\langle x_k : k \in \mathbb{N} \rangle$ converges to a, but the sequence $\langle f(x_k) : k \in \mathbb{N} \rangle$ does not converge to $f(a)$. So (b) fails as well.
- The equivalence of (a) and (b) cannot be proved from the axioms of Zermelo-Fraenkel set theory alone.

Example III: Basis of a Vector Space

- We assume familiarity with the notion of a vector space over a field (e.g., over the field of real numbers).
- A set A of vectors is **linearly independent** if no finite linear combination $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ of elements $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of A, with $a_1, \ldots, a_n$ not all zero from the field, is equal to the zero vector.
- A **basis** of a vector space V is a maximal (in the ordering by inclusion) linearly independent subset of V.

Theorem*

Every vector space has a basis.

- The theorem is a straightforward application of Zorn's Lemma: If C is a $\subseteq$-chain of independent subsets of the given vector space, then the union of C is also an independent set. Consequently, a maximal independent set, exists.
- The theorem cannot be proved in Zermelo-Fraenkel set theory without using the Axiom of Choice.

Example IV: Hamel Basis

- Consider the set of all real numbers as a vector space over the field of rational numbers.
- By the preceding theorem, this vector space has a basis, called a **Hamel basis** for $\mathbb{R}$. In other words, a set $X \subseteq \mathbb{R}$ is a Hamel basis for $\mathbb{R}$ if every $x \in \mathbb{R}$ can be expressed in a unique way as

$$x = r_1 x_1 + \cdots + r_n x_n,$$

  for some mutually distinct $x_1, \ldots, x_n \in X$ and some nonzero rational numbers $r_1, \ldots, r_n$.
- A set of real numbers X is called **dependent** if there are mutually distinct $x_1, \ldots, x_n \in X$ and $r_1, \ldots, r_n \in \mathbb{Q}$ not all zero, such that $r_1 x_1 + \cdots + r_n x_n = 0$.
- A set which is not dependent is called **independent**.

Example IV: Hamel Basis (Cont'd)

- Let A be the system of all independent sets of real numbers. We use Zorn's Lemma to show that A has a maximal element in the $\subseteq$-ordering. Finally, we show that any maximal independent set is a Hamel basis.
  Consider $A_0 \subseteq A$ linearly ordered by $\subseteq$. Let $X_0 = \bigcup A_0$. X_0 is an upper bound of A_0 in $(A, \subseteq)$, since $X_0 \in A$, i.e., X_0 is independent:

  > Suppose there were $x_1, , \ldots, x_n \in X_0$ and $r_1, \ldots, r_n \in \mathbb{Q}$, not all zero, such that $r_1 x_1 + \cdots + r_n x_n = 0$. Then there would be $X_1, \ldots, X_n \in A_0$, such that $x_1 \in X_1, \ldots, x_n \in X_n$. Since A_0 is linearly ordered by $\subseteq$, the finite subset $\{X_1, \ldots, X_n\}$ of A_0 would have a $\subseteq$-greatest element, say X_i. But then $x_1, \ldots, x_n \in X_i$, so X_i would not be independent.

  By Zorn's Lemma, $(A, \subseteq)$ has a maximal element X. It remains to be shown that X is a Hamel basis.

Example IV: X is a Hamel Basis (Existence)

- Suppose that $x \in \mathbb{R}$ cannot be expressed as $r_1x_1 + \cdots + r_nx_n$, for any $r_1, \ldots, r_n \in \mathbb{Q}$ and $x_1, \ldots, x_n \in X$. Then $x \notin X$ (otherwise, $x = 1x$), so $X \cup \{x\} \supset X$, and $X \cup \{x\}$ is dependent (X is a maximal independent set). Thus, there are $x_1, \ldots, x_n \in X \cup \{x\}$ and $s_1, \ldots, s_n \in \mathbb{Q}$, not all zero, such that $s_1x_1 + \cdots + s_nx_n = 0$. Since X is independent, $x \in \{x_1, \ldots, x_n\}$, say $x = x_i$ and the corresponding coefficient $s_i \neq 0$. But then

$$\begin{aligned} x \quad = \quad x_i = (-\frac{s_1}{s_i})x_1 + \cdots + (-\frac{s_{i-1}}{s_i})x_{i-1} \\ + (-\frac{s_{i+1}}{s_i})x_{i+1} + \cdots + (-\frac{s_n}{s_i})x_n. \end{aligned}$$

This contradicts the assumption on x.

Example IV: X is a Hamel Basis (Uniqueness)

- Suppose now that some $x \in \mathbb{R}$ can be expressed in two ways:

$$x = r_1 x_1 \cdots + r_n x_n = s_1 y_1 + \cdots + s_k y_k,$$

where $x_1, \ldots, x_n, y_1, \ldots, y_k \in X$ and $r_1, \ldots, r_n, s_1, \ldots, s_k \in \mathbb{Q} - \{0\}$.
Then $r_1 x_1 + \cdots + r_n x_n - s_1 y_1 - \cdots - s_k y_k = 0$.
If $\{x_1, \ldots, x_n\} \neq \{y_1, \ldots, y_k\}$ (say, $x_i \notin \{y_1, \ldots, y_k\}$) then the above expression contradicts the independence of X. Hence, $n = k$ and $x_1 = y_{i_1}, \ldots, x_n = y_{i_n}$, for some one-to-one mapping $\langle i_1, \ldots, i_n \rangle$ between indices $1, 2, \ldots, n$. We, thus, obtain
$(r_1 - s_{i_1})x_1 + \cdots + (r_n - s_{i_n})x_n = 0$.
Since $x_1, \ldots, x_n$ are mutually distinct elements of X, we conclude that $r_1 - s_{i_1} = 0, \ldots, r_n - s_{i_n} = 0$, i.e., that $r_1 = s_{i_1}, \ldots, r_n = s_{i_n}$.

Example V: Additive Functions

- A function $f : \mathbb{R} \to \mathbb{R}$ is called additive if $f(x+y) = f(x) + f(y)$, for all $x, y \in \mathbb{R}$.
- Example: For fixed $a \in \mathbb{R}$, f_a, with $f_a(x) = a \cdot x$, for all $x \in \mathbb{R}$, is additive.
- Any additive function looks much like f_a, for some $a \in \mathbb{R}$.

  Let f be additive, and set $f(1) = a$. Then, $f(2) = f(1) + f(1) = a \cdot 2$, $f(3) = f(2) + f(1) = a \cdot 3$, and, by induction, $f(b) = a \cdot b$, for all $b \in \mathbb{N} - \{0\}$. Since $f(0) + f(0) = f(0+0) = f(0)$, we get $f(0) = 0$. Next, $f(-b) + f(b) = f(0) = 0$, so $f(-b) = -f(b) = a \cdot (-b)$, for $b \in \mathbb{N}$. To compute $f(\frac{1}{n})$, notice that $a = f(1) = \underbrace{f(\frac{1}{n}) + \cdots + f(\frac{1}{n})}_{n \text{ times}}$.

  Consequently, $f(\frac{1}{n}) = a \cdot \frac{1}{n}$. Continuing along these lines, we can easily prove that $f(x) = a \cdot x$, for all rational numbers x. It is now natural to conjecture that $f(x) = a \cdot x$ holds for all real numbers x, i.e., that every additive function is of the form f_a for some $a \in \mathbb{R}$.

Additive Functions Not of Form f_a

- We conjectured that $f(x) = a \cdot x$ holds for all real numbers x, i.e., that every additive function is of the form f_a, for some $a \in \mathbb{R}$.
- This conjecture cannot be disproved in Zermelo-Fraenkel set theory.
- It is, however, false if we assume the Axiom of Choice.

Theorem*

There exists an additive function $f : \mathbb{R} \to \mathbb{R}$, such that $f \neq f_a$, for all $a \in \mathbb{R}$.

- Let X be a Hamel basis for $\mathbb{R}$. Choose fixed $\overline{x} \in X$. Define

$$f(x) = \begin{cases} r_i, & \text{if } x = r_1 x_1 + \cdots + r_i x_i + \cdots + r_n x_n \text{ and } x_i = \overline{x} \\ 0, & \text{otherwise} \end{cases}$$

It is easy to check that f is additive. Also that $0 \notin X$ and X is infinite (actually, $|X| = 2^{\aleph_0}$). We have $f(\overline{x}) = 1$, while $f(\overline{\overline{x}}) = 0$, for any $\overline{\overline{x}} \in X$, $\overline{\overline{x}} \neq \overline{x}$. If $f = f_a$ was true for some $a \in \mathbb{R}$, we would have $f(\overline{x}) = 1 = a \cdot \overline{x}$, showing $a \neq 0$. But $f(\overline{\overline{x}}) = 0 = a \cdot \overline{\overline{x}}$, showing $a = 0$.

Example VI: The Hahn-Banach Theorem

- A function f defined on a vector space V over the field $\mathbb{R}$ of real numbers and with values in $\mathbb{R}$ is called a **linear functional on** V if $f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v})$, for all $\mathbf{u}, \mathbf{v} \in V$ and $a, b \in \mathbb{R}$.
- A function p defined on V and with values in $\mathbb{R}$ is called a **sublinear functional on** V if $p(\mathbf{u} + \mathbf{v}) \leq p(\mathbf{u}) + p(\mathbf{v})$, for all $\mathbf{u}, \mathbf{v} \in V$, and $p(a\mathbf{u}) = ap(\mathbf{u})$, for all $\mathbf{u} \in V$ and $a \geq 0$.
- The following theorem, due to Hans Hahn and Stefan Banach, is one of the cornerstones of functional analysis:

Theorem*

Let p be a sublinear functional on the vector space V and f_0 be a linear functional defined on a subspace V_0 of V, such that $f_0(\mathbf{v}) \leq p(\mathbf{v})$, for all $\mathbf{v} \in V_0$. Then, there is a linear functional f defined on V, such that $f \supseteq f_0$ and $f(\mathbf{v}) \leq p(\mathbf{v})$, for all $\mathbf{v} \in V$.

Example VI: Proof of the Hahn-Banach Theorem I

- Let F be the set of all linear functionals g defined on some subspace W of V and such that $f_0 \subseteq g$ and $g(\mathbf{v}) \leq p(\mathbf{v})$, for all $\mathbf{v} \in W$. We obtain the desired linear functional f as a maximal element of $(F, \subseteq)$. To verify the assumptions of Zorn's Lemma, consider a nonempty $F_0 \subseteq F$ linearly ordered by $\subseteq$. If $g_0 = \bigcup F_0$, g_0 is a $\subseteq$-upper bound on F_0 provided $g_0 \in F$. Clearly, g_0 is a function with values in $\mathbb{R}$ and $g_0 \supseteq f_0$. Since the union of a set of subspaces of V linearly ordered by $\subseteq$ is a subspace of V, $\mathrm{dom} g_0 = \bigcup_{g \in F_0} \mathrm{dom} g$ is a subspace of V. To show that g_0 is linear consider $\mathbf{u}, \mathbf{v} \in \mathrm{dom} g_0$ and $a, b \in \mathbb{R}$. Then there are $g, g' \in F_0$, such that $\mathbf{u} \in \mathrm{dom} g$ and $\mathbf{v} \in \mathrm{dom} g'$. Since F_0 is linearly ordered by $\subseteq$, we have either $g \subseteq g'$ or $g' \subseteq g$. In the first case, $\mathbf{u}, \mathbf{v}, a\mathbf{u} + b\mathbf{v} \in \mathrm{dom} g'$ and $g_0(a\mathbf{u} + b\mathbf{v}) = g'(a\mathbf{u} + b\mathbf{v}) = ag'(\mathbf{u}) + bg'(\mathbf{v}) = ag_0(\mathbf{u}) + bg_0(\mathbf{v})$; the second case is analogous. Finally, $g_0(\mathbf{u}) = g(\mathbf{u}) \leq p(\mathbf{u})$, for any $\mathbf{u} \in \mathrm{dom} g_0$ and $g \in F_0$, such that $\mathbf{u} \in \mathrm{dom} g$. Thus, $g_0 \in F$.

Example VI: Proof of the Hahn-Banach Theorem II

- By Zorn's Lemma, $(F, \subseteq)$ has a maximal element f. It remains to be shown that $\mathrm{dom} f = V$. We prove that $\mathrm{dom} f \subset V$ implies that f is not maximal.
  Fix $\mathbf{u} \in V - \mathrm{dom} f$. Let W be the subspace of V spanned by $\mathrm{dom} f$ and $\mathbf{u}$. Since every $\mathbf{w} \in W$ can be uniquely expressed as $\mathbf{w} = \mathbf{x} + a\mathbf{u}$, for some $\mathbf{x} \in \mathrm{dom} f$ and $a \in \mathbb{R}$, the function f_c defined by $f_c(\mathbf{w}) = f(\mathbf{x}) + a \cdot c$ is a linear functional on W and $f_c \supset f$. The proof is complete if we show that $c \in \mathbb{R}$ can be chosen so that

$$f_c(\mathbf{x} + a\mathbf{u}) = f(\mathbf{x}) + ac \leq p(\mathbf{x} + a\mathbf{u}),$$

  for all $\mathbf{x} \in \mathrm{dom} f$ and $a \in \mathbb{R}$. Since the properties of f guarantee this, for $a = 0$, it suffices to choose c so as to satisfy:
  (a) For all $a > 0$ and $\mathbf{x} \in \mathrm{dom} f$, $f(\mathbf{x}) + ac \leq p(\mathbf{x} + a\mathbf{u})$.
  (b) For all $a > 0$ and $\mathbf{y} \in \mathrm{dom} f$, $f(\mathbf{y}) + (-a)c \leq p(\mathbf{y} + (-a)\mathbf{u})$.

Example VI: Proof of the Hahn-Banach Theorem III

- Equivalently,

$$f(\mathbf{y}) - p(\mathbf{y} - a\mathbf{u}) \leq ac \leq p(\mathbf{x} + a\mathbf{u}) - f(\mathbf{x})$$

and then

$$f\left(\frac{1}{a}\mathbf{y}\right) - p\left(\frac{1}{a}\mathbf{y} - \mathbf{u}\right) \leq c \leq p\left(\frac{1}{a}\mathbf{x} + \mathbf{u}\right) - f\left(\frac{1}{a}\mathbf{x}\right)$$

should hold for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom} f$ and $a > 0$. But, for all $\mathbf{v}, \mathbf{t} \in \mathrm{dom} f$,

$$f(\mathbf{v}) + f(\mathbf{t}) = f(\mathbf{v} + \mathbf{t}) \leq p(\mathbf{v} + \mathbf{t}) \leq p(\mathbf{v} - \mathbf{u}) + p(\mathbf{t} + \mathbf{u})$$

and, thus,

$$f(\mathbf{v}) - p(\mathbf{v} - \mathbf{u}) \leq p(\mathbf{t} + \mathbf{u}) - f(\mathbf{t}).$$

If

$$\begin{aligned} A &= \sup\{f(\mathbf{v}) - p(\mathbf{v} - \mathbf{u}) : \mathbf{v} \in \mathrm{dom} f\}, \\ B &= \inf\{p(\mathbf{t} + \mathbf{u}) - f(\mathbf{t}) : \mathbf{t} \in \mathrm{dom} f\}, \end{aligned}$$

we have $A \leq B$. By choosing c, such that $A \leq c \leq B$, we can make the required identity hold.

Example VII: The Measure Problem

- An important problem in analysis is to extend the notion of length of an interval to more complicated sets of real numbers.
- Ideally, one would like to have a function μ defined on $\mathcal{P}(\mathbb{R})$, with values in $[0,\infty)\cup\{\infty\}$, and having the following properties:
  - (0) $\mu([a,b]) = b-a$, for any $a,b\in\mathbb{R}, a<b$.
  - (i) $\mu(\emptyset)=0, \mu(\mathbb{R})=\infty$.
  - (ii) If $\{A_n\}_{n=0}^{\infty}$ is a collection of mutually disjoint subsets of $\mathbb{R}$, then $\mu(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty}\mu(A_n)$. (This property is called **countable additivity** or σ**-additivity** of μ.)
  - (iii) If $a\in\mathbb{R}$, $A\subseteq\mathbb{R}$, and $A+a=\{x+a : x\in A\}$, then $\mu(A+a)=\mu(A)$ (**translation invariance** of μ).
- Several additional properties of μ follow immediately from (0)-(iii):
  - (iv) If $A\cap B=\emptyset$, then $\mu(A\cup B)=\mu(A)+\mu(B)$ (**finite additivity**).
  - (v) If $A\subseteq B$, then $\mu(A)\leq\mu(B)$ (**monotonicity**).
- However, the Axiom of Choice implies that no function μ with the aforementioned properties exists.

Example VII: Nonexistence of a Measure

Theorem*

There is no $\mu : \mathcal{P}(\mathbb{R}) \to [0, \infty) \cup \{\infty\}$, with the properties (0)-(v).

- We define an equivalence relation $\approx$ on $\mathbb{R}$ by:

$$x \approx y \quad \text{if and only if} \quad x - y \text{ is a rational number,}$$

and use the Axiom of Choice to obtain a set of representatives X for $\approx$. It is easy to see that $\mathbb{R} = \{X + r : r \text{ is rational}\}$. Moreover, if q and r are two distinct rationals, then $X + q$ and $X + r$ are disjoint. Note that $\mu(X) > 0$: If $\mu(X) = 0$, then $\mu(X + q) = 0$, for every $q \in \mathbb{Q}$, and $\mu(\mathbb{R}) = \sum_{q \in \mathbb{Q}} \mu(X + q) = 0$, a contradiction. By countable additivity, there exists $[a, b]$, such that $\mu(X \cap [a, b]) > 0$. Let $Y = X \cap [a, b]$. Then $\bigcup_{q \in \mathbb{Q} \cap [0,1]} (Y + q) \subseteq [a, b + 1]$ and the left-hand side is the union of infinitely many mutually disjoint sets $Y + q$, each of measure $\mu(Y + q) = \mu(Y) > 0$. Thus, the left-hand side has measure ∞, contrary to $\mu([a, b + 1]) = b + 1 - a$.

σ-Additive Measures on σ-Algebras

- The requirements on μ have to be relaxed.
- We give up the condition that μ be defined for all subsets of $\mathbb{R}$.

Definition (σ-Algebra)

Let S be a nonempty set. A collection $\mathfrak{S} \subseteq \mathcal{P}(S)$ is a σ-**algebra** of subsets of S if

(a) $\emptyset \in \mathfrak{S}$ and $S \in \mathfrak{S}$.

(b) If $X \in \mathfrak{S}$, then $S - X \in \mathfrak{S}$.

(c) If $X_n \in \mathfrak{S}$, for all n, then $\bigcup_{n=0}^{\infty} X_n \in \mathfrak{S}$ and $\bigcap_{n=0}^{\infty} X_n \in \mathfrak{S}$.

Definition (σ-Additive Measure)

A σ-**additive measure** on a σ-algebra $\mathfrak{S}$ of subsets of S is a function $\mu : \mathfrak{S} \to [0, \infty) \cup \{\infty\}$, such that

(i) $\mu(\emptyset) = 0$, $\mu(S) > 0$.

(ii) If $\{X_n\}_{n=0}^{\infty}$ is a collection of mutually disjoint sets from $\mathfrak{S}$, then $\mu(\bigcup_{n=0}^{\infty} X_n) = \sum_{n=0}^{\infty} \mu(X_n)$.

The elements of $\mathfrak{S}$ are called μ-**measurable sets**.

Nonmeasurable Sets

- $\mathcal{P}(S)$ is the largest σ-algebra of subsets of S; we refer to a measure defined on $\mathcal{P}(S)$ as a **measure on** S.
- The theorem we proved takes the form:

Corollary

Let μ be any σ-additive measure on a σ-algebra $\mathfrak{S}$ of subsets of $\mathbb{R}$ such that

(0) $[a, b] \in \mathfrak{S}$ and $\mu([a, b]) = b - a$, for all $a, b \in \mathbb{R}$, $a < b$.

(iii) If $A \in \mathfrak{S}$, then $A + a \in \mathfrak{S}$ and $\mu(A + a) = \mu(A)$, for all $a \in \mathbb{R}$.

Then there exist sets of real numbers which are not μ-measurable.

- In real analysis, one constructs a particular σ-algebra $\mathfrak{M}$ of **Lebesgue measurable sets**, and a σ-additive measure μ on $\mathfrak{M}$, the **Lebesgue measure**, satisfying properties (0) and (iii) of the Corollary.
- So existence of Lebesgue nonmeasurable sets is a consequence of the Axiom of Choice. Robert Solovay showed that the Axiom of Choice is necessary to prove this result.

The Axiom of Countable Choice

- There are many fundamental and intuitively acceptable results concerning countable sets and topological and measure-theoretic properties of the real line, whose proofs depend on the Axiom of Choice.
- However, closer investigation of the preceding proofs of Examples I and II reveals that only a very limited form of the Axiom is needed:

Axiom of Countable Choice

There exists a choice function for every countable system of sets.

- It might well be that the Axiom of Countable Choice is intuitively justified, but the full Axiom of Choice is not.
- Moreover, the full Axiom of Choice has some counterintuitive consequences, such as the existence of nonlinear additive functions, or the existence of Lebesgue nonmeasurable sets, none of which follows from the Axiom of Countable Choice.

Full Axiom of Choice: Striking Applications

- Many important applications require the Axiom of Choice in almost full strength:
  - The Hahn-Banach Theorem;
  - Tichonov's Theorem: A topological product of any system of compact topological spaces is compact.
  - The Maximal Ideal Theorem: Every ideal in a ring can be extended to a maximal ideal.
- Some of them are even equivalent to it.
- The irreplaceable role of the Axiom of Choice is to simplify general topological and algebraic considerations which, otherwise, would be bogged down in irrelevant set-theoretic detail.
- For this pragmatic reason, the Axiom of Choice will always keep its place in set theory.